De-Shuang Huang   Kang-Hyun Jo
Hong-Hee Lee   Hee-Jun Kang
Vitoantonio Bevilacqua (Eds.)

LNCS 5754

Emerging Intelligent Computing Technology and Applications

5th International Conference on Intelligent Computing, ICIC 2009
Ulsan, South Korea, September 2009
Proceedings

Springer

Lecture Notes in Computer Science 5754

Commenced Publication in 1973
Founding and Former Series Editors:
Gerhard Goos, Juris Hartmanis, and Jan van Leeuwen

De-Shuang Huang   Kang-Hyun Jo
Hong-Hee Lee   Hee-Jun Kang
Vitoantonio Bevilacqua (Eds.)

Emerging Intelligent Computing Technology and Applications

5th International Conference
on Intelligent Computing, ICIC 2009
Ulsan, South Korea, September 16-19, 2009
Proceedings

Springer

Volume Editors

De-Shuang Huang
Institute of Intelligent Machines
Intelligent Computing Laboratory
Chinese Academy of Sciences
Hefei, Anhui, China
E-mail: dshuang@iim.ac.cn

Kang-Hyun Jo
Hong-Hee Lee
Hee-Jun Kang
University of Ulsan
School of Electrical Engineering
Ulsan, South Korea
E-mail: jokanghyun@gmail.com, {hhlee, hjkang}@ulsan.ac.kr

Vitoantonio Bevilacqua
Polytechnic of Bari
eBIS and DEE
Valenzano, Bari, Italy
E-mail: vitoantonio.bevilacqua@gmail.com

Preface

The International Conference on Intelligent Computing (ICIC) was formed to provide an annual forum dedicated to the emerging and challenging topics in artificial intelligence, machine learning, bioinformatics, and computational biology, etc. It aims to bring together researchers and practitioners from both academia and industry to share ideas, problems, and solutions related to the multifaceted aspects of intelligent computing.

ICIC 2009, held in Ulsan, Korea, September 16–19, 2009, constituted the 5th International Conference on Intelligent Computing. It built upon the success of ICIC 2008, ICIC 2007, ICIC 2006, and ICIC 2005 held in Shanghai, Qingdao, Kunming, and Hefei, China, 2008, 2007, 2006, and 2005, respectively.

This year, the conference concentrated mainly on the theories and methodologies as well as the emerging applications of intelligent computing. Its aim was to unify the picture of contemporary intelligent computing techniques as an integral concept that highlights the trends in advanced computational intelligence and bridges theoretical research with applications. Therefore, the theme for this conference was "Emerging Intelligent Computing Technology and Applications." Papers focusing on this theme were solicited, addressing theories, methodologies, and applications in science and technology.

ICIC 2009 received 1,082 submissions from 34 countries and regions. All papers went through a rigorous peer review procedure and each paper received at least three review reports. Based on the review reports, the Program Committee finally selected 257 high-quality papers for presentation at ICIC 2009, of which 214 papers have been included in two volumes of proceedings published by Springer: one volume of *Lecture Notes in Computer Science* (LNCS) and one volume of *Lecture Notes in Artificial Intelligence* (LNAI). The other 22 papers will be included in two international journals.

This volume of *Lecture Notes in Computer Science* (LNCS) includes 108 papers.

The organizers of ICIC 2009, including the University of Ulsan, Korea, Institute of Intelligent Machines of Chinese Academy of Sciences, made an enormous effort to ensure the success of ICIC 2009. We hereby would like to thank the members of the Program Committee and the referees for their collective effort in reviewing and soliciting the papers. We would like to thank Alfred Hofmann, executive editor at Springer, for his frank and helpful advice and guidance throughout and for his support in publishing the proceedings. In particular, we would like to thank all the authors for contributing their papers. Without the high-quality submissions from the authors, the success of the conference would not have been possible. Finally, we are especially grateful to the IEEE Computational Intelligence Society, the International Neural Network Society and the National Science Foundation of China for their sponsorship.

July 2009

De-Shuang Huang
Kang-Hyun Jo
Hong-Hee Lee
Hee-Jun Kang
Vitoantonio Bevilacqua

Organization

General Co-chairs	De-Shuang Huang, China
	Honghee Lee, Korea
	Frank L Lewis, USA
Program Committee Co-chairs	Kanghyun Jo, Korea
	Vitoantonio Bevilacqua, Italy
Organizing Committee Co-chairs	Kang-Hyun Jo, Korea
	In-Soo Koo, Korea
	Youngsoo Suh, Korea
	Naoyuki Tsuruta, Japan
	Chun-Hou Zheng, China
Award Committee Chair	Daniel S. Levine, USA
Publication Chair	Heejun Kang, Korea
Special Session Chair	Prashan Premaratne, Australia
	Tokuro Matsuo, Japan
	Vasily Gubarev, Russia
Tutorial Chair	Laurent Heutte, France
International Liaison Chair	Frank Neumann, Germany
Publicity Co-chairs	Kyungsook Han, Korea
	Vladimir Filaretov, Russia
	Zhongming Zhao, USA
	Maolin Tang, Australia
	Muhammad Khurram Khan, Saudi Arabia
	Valeriya Gribova, Russia
Exhibition Co-chairs	Young-Soo Suh, Korea
	In-Soo Koo, Korea
	Jin Hur, Korea

Organizing Committee Members

Myeong-Jae Yi, Korea	Sang-Bock Cho, Korea	Dong-Joong Kang, Korea
Myung-Kyun Kim, Korea	Munho Jeong, Korea	Jong-Bae Lee, Korea
Byeong-Ryong Lee, Korea	Jongeun Ha, Korea	Sang-Moo Lee, Korea
Won-Ho Choi, Korea		

Program Committee Members

Andrea Francesco Abate, Italy
Shafayat Abrar, UK
Peter Andras, UK
Sabri Arik, Turkey
Vasily Aristarkhov, Russian Federation
Costin Badica, Romania
Vitoantonio Bevilacqua, Italy
David B. Bracewell, USA
Uday K. Chakraborty, USA
Shih-Hsin Chen, Taiwan, China
Wen-Sheng Chen, China
Xiyuan Chen, China
Yang Chen, China
Yuehui Chen, China
Sang-Bock Cho, Korea
Won-Ho Choi, Korea
Michal Choras, Poland
Tommy Chow, Hong Kong, China
Jose Alfredo F. Costa, Brazil
Angelo Ciaramella, Italy
Kevin Curran, UK
Mingcong Deng, Japan
Eng. Salvatore Distefano, Italy
Karim Faez, Iran
Jianbo Fan, China
Minrui Fei, China
Wai-Keung Fung, Canada
Liang Gao, China
Qing-Wei Gao, China
Xiao-Zhi Gao, Finland
Chandan Giri, India
Dunwei Gong, China
Valeriya Gribova, Russia
Kayhan Gulez, Turkey
Ping Guo, China
Jongeun Ha, Korea

Aili Han, China
Fei Han, China
Kyungsook Han, Korea
Haibo He, USA
Laurent Heutte, France
Wei-Chiang Hong, Taiwan, China
Yuexian Hou, China
Peter Hung, Ireland
Chuleerat Jaruskulchai, Thailand
Munho Jeong, Korea
Li Jia, China
Zhenran Jiang, China
Jih-Gau Juang, Taiwan, China
Dah-Jing Jwo, Taiwan, China
Dong-Joong Kang, Korea
Sanggil Kang, Korea
Uzay Kaymak, The Netherlands
Muhammad Khurram Khan, Saudi Arabia
Myung-Kyun Kim, Korea
Sungshin Kim, Korea
In-Soo Koo, Korea
Donald H. Kraft, USA
Harshit Kumar, Ireland
Yoshinori Kuno, Japan
Takashi Kuremoto, Japan
Wen-Chung Kuo, Taiwan, China
Hak-Keung Lam, UK
Byeong-Ryong Lee, Korea
Jong-Bae Lee, Korea
Sang-Moo Lee, Korea
Vincent C.S. Lee, Australia
Guo-Zheng Li, China
Kang Li, UK

Li Li, China
Peihua Li, China
Hualou Liang, USA
Chunmei Liu, USA
Ju Liu, China
Van-Tsai Liu, Taiwan, China
Marco Loog, Denmark
Ahmad Lotfi, UK
Jinwen Ma, China
Shiwei Ma, China
Vishnu Vardhan Makkapati, India
Cheolhong Moon, Korea
Tarik Veli Mumcu, Germany
Roman Neruda, Czech Republic
Frank Neumann, Germany
Minh Nhut Nguyen, Singapore
Ben Niu, China
Sim-Heng Ong, Singapore
Francesco Pappalardo, Italy
Caroline Petitjean, France
Prashan Premaratne, Australia
Shaoqi Rao, China
Seeja K.R., India
Angel Sappa, Spain
Aamir Shahzad, China
Li Shang, China
Jiatao Song, China
Nuanwan Soonthornphisaj, Thailand
Joao Miguel Sousa, Portugal
Min Su, USA
Zhan-Li Sun, Singapore
Maolin Tang, Australia

Antonios Tsourdos, UK
Naoyuki Tsuruta, Japan
Sergio Vitulano, Italy
Anhua Wan, China
Chao-Xue Wang, China
Hong-Qiang Wang, USA
Jinlian Wang, China
Ling Wang, China
Xueqin Wang, China

Yong Wang, China
Xuesong Wang, China
Ling-Yun Wu, China
Shunren Xia, China
Yu Xue, China
Ching-Nung Yang,
   Taiwan, China
Jun-Heng Yeh, Taiwan,
   China

Myeong-Jae Yi, Korea
Zhi-Gang Zeng, China
Jun Zhang, China
Yong Zhang, China
Xing-Ming Zhao, China
Zhongming Zhao, USA
Bo-Jin Zheng, China
Fengfeng Zhou, USA
Huiyu Zhou, Italy

Reviewers

Li Ding, Alessandro Ibba, Al Savvaris, Antonio Celesti, Adam Ghandar, Adriao Duarte, Asit Das, Andreas Konstantinidis, Alaa Sagheer, Alan Ritz, Aldayr Araujo, Alessia D'Introno, Alessia Albanese, Alessio Ferone, Alexander Hogenboom, Jose Alfredo F. Costa, Rui Jorge Almeida, Andrey Logvinov, Soon-Min Hwang, Saleh Aly, Amar Khoukhi, Amar Balla, Amelia Badica, Asunción Mochón, Aimin Zhou, Anbumani Subramanian, Andreas Schmidt, Wen-Yuan Liao, Andrey Larionov, Angelo Ciaramella, Angelo Riccio, Anne Canuto, Wei Yu, Antonino Staiano, Anvita Bajpai, Alexander Ponomarenko, Xinping Xie, Aravindan Chandrabose, Joongjae Lee, Ardelio Galletti, Irene Artemieva, Arun D. Mahindrakar, Asaduzzaman, Asharaf S, Atsushi Shimada, Wee Keong Ng, Banu Diri, Bao Vo-Nguyen, Bo-Chao Cheng, Beilu Shao, Beilu Shao, Ibrahim Beklan Kucukdemiral, Bo-Hyeun Wang, Bijaya Ketan Panigrahi, Bin Qian, Bin Li, Shuhui Bi, Xiangrong Zhang, Bekir Karlik, Jiguang Wang, Bogdan Raducanu, Barbara Pizzileo, Ni Bu, Cheon Seong-Pyo, B.V. Babu, Alessia D'Introno, Galip Cansever, Jianting Cao, Karina Shakhgendyan, Carme Julia, Caroline Petitjean, Chia-Mei Chen, Guisheng Chen, Gang Chen, Kuei-Hsiang Chao, Tariq Chattha, Chungho Cho, Jianhua Che, bo chen, Chun Chen, Chengjian Wei, Yuhu Cheng, chen hui, Chenkun Qi, Yang Chen, Chen Asia, Chee-Meng Chew, Ching-Hung Lee, Chuang Ma, Cuco Cuistana, C.-H. Yang, Alessandro Cincotti, Chenn-Jung Huang, Ching-kun Chen, Chunlin Chen, Jimmy Lin, Chi-Min Li, Quang Nguyen, Carmelo Ragusa, Wenjie Li, Min-Chih Chen, Ching-Ti Liu, Chingti Liu, Chi Zhou, Chin-Chun Chang, Chang Wook Ahn, Joo Seop Yun, Chieh-yao Chang, Changyin Sun, dong yang, Louis Wu, Yu-Chen Lin, Ping-Min Hsu, Danfeng Zhu, Vincenzo Daniele Cunsolo, Peng Zhang, David Bracewell, Dario Bruneo, Dajun Du, David Geronimo, Liya Ding, Dmitry Serkin, Jiayin Zhou, Dongsheng Che, Yan Dong, Yongsheng Dong, Denis Orel, Jun Qin, WeiWu Wang, Woosung Yang, Ben Niu, derchian tsaih, Dunwei Gong, Wenyong Dong, Lipo Wang, Hong Fu, Tolga Ensari, Shaoli Wang, Eylem Yucel, Erkan Zergeroglu, Filippo Castiglione, Li-Jen Kao, Chonglun Fang, Ingo Feldmann, Fei Ge, Fengfeng Zhou, LingFeng Liu, Frederik Hogenboom, Chien-Yuan Lai, Wei-Chiang Hong, Francesco Longo, Francesco Napolitano, Francesco Camastra, Nuanwan Soonthornphisaj, Fu-Shiung Hsieh, Shaojing Fan, Francesco Tusa, Fu Yonggui, Lina Lu, Yen Ming Chiu, Zhaohui Gan, Xiao-Zhi Gao, Dingfei Ge, Gerrit K. Janssens, Gwang-Hyun Kim, Ginny Wong, Giuseppe Agrillo, Yaroslava Katueva, Giuseppe MAngioni, Fahad Muhaya, Guang-Ming Wu, Xiujun Gong, Gouhei Tanaka, Muhammad

Khurram Khan, Ge Lei, Zhongsheng Wang, Guo Weidong, Jie Gui, Guilherme Barreto, Tiantai Guo, Gurumurthy Swaminathan, Guangwei Zhang, Gwo-Ruey Yu, Moussa Haddad, Haibing Gao, H.K. Lam, Hanif Ullah, Hanlin He, Haini Qu, Chiung-Hua Huang, Houshang Darabi, Tomohiro Henmi, Herbert Iu, Tiefang He, Han-min Chien, Honorius Galmeanu, Hassan Taheri, Huang Ping, Wei Huang, Weitong Huang, Huifang Li, Huiyu Zhou, Junhao Hu, Hameed Ullah Khan, Rong-xiang Hu, Shahid Hussain, Bo Chen, Jaehyung Park, Hsiang-Yi Lee, Hoang-Yang Lu, Hyun-Sik Kim, Zhongkun He, Ibrahim Aliskan, Irene Artemieva, Indrajit Banerjee, Ing-Chyuan Wu, Ikhyeon Jang, Jianli Li, Seong-Joe Lim, Francesco Iorio, yaou zhao, Jin Zhou, Insoo Koo, Jian Xun Peng, John Economou, Jackson Souza, Jose Alvarez, James Cai, James Walton, James Yeh, Hasan Jamil, Janset Dasdemir, Jawid Azizi, Jayasudha John Suseela, Jianbo Fan, Jiande Sun, Jih-Gau Juang, Javad Haddadnia, Hongjun Jia, Jiajun Yan, Peilin Jiang, Changan Jiang, Jiang jl, Kai Jiang, Lihua Jiang, Wei Jia, Jindong Liu, Guang Jin, Jinsoo Kim, Jungkyu Rho, Josep M. Mirats Tur, Jun Liu, John Klein, Jong Min Lee, Ji-Hun Bae, Joydeb Mukherjee, Jianping Qiao, Jinn-Shing Cheng, Joaquin Torres-Sospedra, Joaquin Torres-Sospedra, Jyh-Ching Juang, Juan Jose Gonzalez de la Rosa, Junaid Ahmed, Jun Du, Junlin Chang, Kang Li, Kanghee Kim, Wei Jing, Kaushik Roy, Iroshi Awasaki, Tsung-Yi Chen, Ke Tang, Hyun-Deok Kang, Alexander Kleschev, Kunikazu Kobayashi, Krishna Chandramouli, Krishnanand Kaipa Narasimha, Seeja K.R.H. K, Lance C. Fung, Laks Raghupathi, Lalit Gupta, Chin-Feng Lin, Le Dong, Sungon Lee, Hong-Bo Lei, Jie Lei, Yingke Lei, Kok-Leong Ong, Lin Gao, Sun Cheol Bae, Laurent Heutte, Hualiang Li, Lijuan Xiao, Lin Li, Guohui Zhang, Lin Wang, Yuxi Liu, Bo Liu, Huiran Liu, Lei Liu, Wenyun Li, Xinyu Li, Ling-po Li, Linlin Shen, Leh Luoh, Lingling Wang, Peixing Li, Milan Lovric, Li Qingfeng, Liqing Zhang, Tian-Yu Liu, Liangxu Liu, Yixiang Lu, Marco Cortellino, Maciej Hrebien, Yasushi Mae, Sakashi Maeda, Sakashi Maeda, Margaret Knyaseva, Margarita Knyazeva, Manish Srivastava, Maqsood Mahmud, M. Loog, JeongHyun Kim, Mario Marinelli, Mario Marinelli, Markus Koskela, Kazuyuki Matsumoto, Maqsood Mahmud, Max Power, Maysam Abbod, Zhongqiang Wu, Mark Halling-Brown, Aizhong Mi, Mika Sulkava, Min Jiang, Min Wu, Mine Tsunenori, hai min, Meiling Hou, Hamid Abrishami Moghaddam, Mohammad Narimani, Monalisa Mazumdar, Lucia Moreno, Santo Motta, Marzio Pennisi, Minh-Tri Pham, Mutsumi Watanabe, Mingyu You, Naeem Ramzan, Naiara Aginako, Nestor Arana, Beijing Chen, Nelson Mascarenhas, Seref Naci Engin, Neyir Ozcan, Mingxiao Li, Li Nie, Xiushan Nie, Nataliya Nikiforova, Nataliya Nikifirova, Nitthinun Suphasetthawit, Nikolay Mikhaylov, Qun Niu, Nhan Nguyen-Thanh, Evgeni Nurminski, Bunyarit Uyyanonvara, Masaru Okumura, Olesya Kazakova, Won-Kyu Kim, Kazunori Onoguchi, Ajiboye Osunleke, Ertan Ouml Znergiz, Ping Zhang, Pallavi Vajinepalli, Pandu Devarakota, Yehu Shen, Chen Peng, Alessandro Perfetto, Hyun-Ju Park, Ping Wang, Peilin Jia, Litt Teen Hiew, Elvira Popescu, Roy Power, Roy Power, Pradip Ghanty, Pramod NC, Pramuditha Suraweera, Prashan Premaratne, Prashan Premaratne, Qi Yu, Qiao Wang, Qi Liu, Qingwei Gao, Quande Qin, Jinpeng Qi, Peng Qiu, Quanke Pan, Thanh Tho Quan, Quang Nguyen, Hai-Tao Zheng, Qi Wang, Ruhul Sarker, Rafal Kozik, Raffaele Montella, M. Rafiq Swash, M.K.M. Rahman, Randeep Singh, Peng Ren, Xianwen Ren, Romina Oliva, Rong Jin, Rosa de Duonni, Lijun Xu, Nidhi Arora, Ryuzo Okada, Shaomin Zhang, Chin-yuan Fan, Saad Bedros, Xin Hao, Sarif Naik, Mihnea Scafes, Sheng Chen, Chen Shao, Jong

Hyun Park, Sanggil Kang, Changho Yun, Shafayat Abrar, Elena Shalfeeva, Li Shang, Shao jj, Xiaojian shao, Sherif Sherif, Chuan Shi, Shaohui Liu, Shripad Kondra, S. Jamal H Zaidi, Shi-Jay Chen, Jiping SHI, Seokjoo Shin, Shiuh-Jeng Wang, Sawomir Lasota, Zhijun Tan, Mingguang Shi, Vitaliy Snytyuk, Xiaojing Song, Shengping Zhang, Sriparna Saha, Sibel Senan, Seokjin Sung, Eung Nam Ko, Sungshin Kim, S Kim, Xue-qiang Zeng, Lei Zhang, Steve Ling, Steven Guan, Shih-Ting Yang, Zhang Li, Cheng Sun, Jie Sun, Tingxu Yan, You Ouyang, Supriya Rao, Susana Vieira, Suwon Lee, Yang Shi, Syed Ismail Shah, Peixing Li, Tiong Goh, Shin-ya Takahashi, Shinya Takahashi, Toshihisa Tanaka, Atsushi Yamashita, Weidong Xu, Zhi Teng, Zhu Teng, Thomas Tawiah, Thuc Kieu Xuan, Timo Honkela, Toshiaki Kondo, Tsang-Long Pao, ThanhVu Nguyen, Thomas O'Daniel, Tomasz Andrysiak, Tomasz Rutkowski, Toni Zgaljic, Gyung-Jin Hong, Tomoaki Tsuruoka, Naoyuki Tsuruta, Mengru Tu, U. Kaymak, Uttam Roy, Youngbae Hwang, Mario Rossi, Vanta Dimitrova, Vasily Aristarkhov, Venugopal Chakravarthy, Vinod Pathangay, Bae-guen Kwon, Vito Santarcangelo, Victor Jin, Vladimir Brusic, Wan-Jui Lee, Chih-Hung Wang, Chao Wang, Furong Wang, Wang Haili, Ling Wang, Xiaojuan Wang, Yongcui Wang, Zhengyou Wang, Wen-Chung Chang, Woochang Shin, Wuchuan Yang, Wudai Liao, Wei-Chih Yang, Weidong Li, Weifeng Li, Wenkai Li, Wen Shengjun, Yu-Chen Lin, Wangheon Lee, Wing-Kuen Ling, Shanwen Zhang, Wai-keung Fung, Worasait Suwannik, Takashi Kuremoto, Chao Wu, Yu Wu, Zikai Wu, Jun Zhang, Wei Xiong, Xin Zou, Xiaochun Cao, Chungui Xu, XiaoFeng Wang, Junfeng Xia, Xian-xia Zhang, Xiaomin Liu, Xianjun Shen, Xuemei Ren, De Xu, Bing Xue, Yu Xue, Huan Xu, Lu Xu, Ye Xu, Yun Xu, Xiaolei Xia, Xiaoyan Sun, Xiaoying Wang, Yang Song, Yago Saez, Yan Li, Banghua Yang, Yan Yang, Zhixia Yang, Yanmin Liu, Akira Yanou, Yasuhiro Taniguchi, Yuan-Chang Chang, Yu-Chiun Chiou, Ye Bin, Yeonsik Kang, Y.F. Xu, Yifeng Zhang, Zhao Yinggang, Yinglei Song, Lei Yang, Yangmin Li, Mi-ran Yun, Yoshinori Kobayashi, Yu-Qing Qiu, Yoon-Seok Nam, Yuanling Hao, Ming Yu, Yong Wang, Yue Wang, Yen-Wen Wang, Zhigang Wang, Zanchao Zhang, Zhenbing Zeng, Guowei Zhang, Hehua Zhang, Jun Zhang, Liang Zhao, Zhaohui Sun, Chunhou Zheng, Min Zheng, Zhigang Yan, Zhijun Yang, Lin Zhu, Zhong Jin, Zujun Hou, Dao Zhou, Sulan Zhang, Xiangbin Zhu, Shuanghe Zhu, Xuefen Zhu, Yihai Zhu, Zhang Liangsheng, Liu Zhiping, Guoyu Zuo, Zhongming Zhao.

Table of Contents

Intelligent Computing in Computational Biology and Drug Design

Computational Genomics and Proteomics

Intelligent Computing in Signal Processing

Intelligent Computing in Pattern Recognition

Intelligent Computing in Image Processing

Intelligent Computing in Communication and Computer Networks

Intelligent Computing in Robotics

Intelligent Computing in Computer Vision

Intelligent Agent and Web Applications

Intelligent Sensor Networks

Intelligent Fault Diagnosis and Financial Engineering

Intelligent Control and Automation

Intelligent Data Fusion and Security

Intelligent Prediction and Time Series Analysis

Natural Language Processing and Expert Systems

Intelligent Image/Document Retrievals

Computational Analysis and Data Mining in Biological Systems

Knowledge-Based Systems and Intelligent Computing in Medical Imaging

Applications of Intelligent Computing in Information Assurance and Security

Computational Analysis and Applications in Biomedical System

Intelligent Computing Algorithms in Banking and Finance

Network-Based Intelligent Technologies

Erratum

Supervised Locally Linear Embedding for Plant Leaf Image Feature Extraction

Youqian Feng[1] and Shanwen Zhang[2]

[1] The Science Institute, Air-Force Engineering University, Xi' an, 710051, China
[2] Intelligent Computation Lab, Hefei Institute of Intelligent Machines,
Chinese Academy of Science, P.O. Box 1130, Hefei, Anhui 230031, P.R. China
`fengyouqian123@163.com, zhangshanwen1965@163.com`

Abstract. The objects of traditional plant identification were too broad and the classification features of it were usually not synthetic and the recognition rate was always slightly low. This paper gives one recognition approach based on supervised locally linear embedding (LLE) and K-nearest neighbors. The recognition results for thirty kinds of broad-leaved trees were realized and the average correct recognition rate reached 98.3%. Comparison with other recognition method demonstrated the proposed method is effective in advancing the recognition rate.

Keywords: Plant leaf image, Manifold learning, Supervised locally linear embedding, Plant classification.

1 Introduction

To obtain an idea of the variety of existing plants and to get to know special species, it is necessary to identify them. Identifying the plants may be a complex business. So far, this task that has mainly been carried out by taxonomists and/or botanists is time-consuming process. With the rapid development of information technology, computer vision techniques and various data management techniques, a computer-aided plant identification system is becoming more and more feasible. The availability of a rapid and accurate method to recognize and classify leaf images also becomes necessary. Although a lot of leaf image feature extraction methods for plant recognition have been presented [1-8], it is necessary to study the plant classification method. *Saitoh et al.* combined flower and leaf information to classify wild flowers [9]. *Heymans et al.* proposed an application of ANN to classify opuntia species [10]. *Du et al.* introduced a shape recognition approach based on radial basis probabilistic neural network which is trained by orthogonal least square algorithm (OLSA) and optimized by recursive OLSA [11]. It performs plant recognition through modified Fourier descriptors of leaf shape. Previous works have some disadvantages. Some are only applicable to certain species. As expert system, some methods compare the similarity between features [12]. It requires pre-process work of human to enter keys manually. This problem also happens on methods extracting features used by botanists [13].

LLE is a classical local nonlinear approach in manifold learning [14], and it has many applications such as image classification, image recognition, spectra reconstruction and

data visualization because of its few parameters, rapid computation, and global optimization. However, LLE may map faraway inputs to nearby out in the low dimensional space which lead to a visible distortion in the embedding results. One of the curses that make LLE fail is that only one set of reconstruction weights can not reflect the local geometry of the manifold. Furthermore, the constrained least squares problem for finding the weights may do not have a unique solution. Using regularization method to solve the constrained LS problem is involved in the selection of a regularization term, and we can not verify that whether the computed solution is optimal. These questions remain to be answered.

Another shortage is that the classical LLE neglects the class information, which can impair the recognition accuracy. Recently, many modified LLE algorithms are put forward to make use of the class labels. Supervised LLE (SLLE) [14] was introduced to deal with data sets labeled with class information. This method defines a distance for neighborhood selection, taking into account the class information such that the distance of two points in different classes is relatively larger than their Euclidean distance. In this method, the Euclidean distance is simply enlarged by adding a constant for the pairs of points belonging different classes, keeping others unchanged.

SVM is powerful classification systems based on a variation of regularization techniques for regression [16]. It provides state-of-the-art performance in many practical binary classification problems. SVM has also shown promise in a variety of biological classification tasks, including some involving plant leaf images. The algorithm is a particular instantiation of regularization for binary classification. Linear SVM can be viewed as a regularized version of a much older machine learning algorithm, the perceptron. The goal of a perceptron is to find a separating hyper-plane that separates positive from negative examples. In general, there may be many separating hyper-planes. In our problem, this separating hyper-plane is the boundary that separates a given tumor class from the rest (OVA) or that separates two different tumor classes (AP). The hyper-plane computed by the SVM is the maximal margin hyper-plane, that is, the hyper-plane with maximal distance to the nearest data point. Finding the SVM solution requires training an SVM, which entails solving a convex quadratic program with as many variables as training points.

The rest of the paper is organized as follows. Section 2 and Section 3 introduce a supervised locally linear embedding and a SVM classifier, respectively. Numerical experimental are illustrated in Section 4. The conclusion are given in Section 5.

2   Supervised Locally Linear Embedding

Let $X = [X_1, X_2, ..., X_n] \in R^{D \times n}$ be a set of n points in a high dimensional input data space. The data points are well sampled from a nonlinear manifold, of which the intrinsic dimensionality is d ($d \ll D$). The goal of LLE is to map the high dimensional data into a low dimensional embedding. Let us denote the corresponding set of n points in the embedding space by $Y = [Y_1, Y_2, ..., Y_n] \in R^{d \times n}$. The outline of LLE can be summarized as follows:

The supervised LLE (SLLE) proposed in [14] is an LLE-like method for dimensionality of the training data. The distance introduced in SLLE is defined by

$$dist(x_i, x_j) = \|x_i - x_j\| + \alpha \bullet M \bullet (1 - \delta(x_i, x_j))$$

where $M = \max_{ij} \|x_i - x_j\|$ is the data diameter in Euclidean distance, $\alpha \in [0,1]$ is a tuning parameter.. ... $\delta(x_i, x_j)$ ... . character function defined below,

$$\delta(x_i, x_j) = \begin{cases} 1 & x_i \text{ and } x_j \text{ belong to same class} \\ 0 & \text{otherwise} \end{cases} ;$$

In SLLE, k nearest neighbors of each training point x_i are determined using the distance defined above at first. As soon as neighbor sets are selected for each point, SLLE finds low-dimensional representation $\{Y_i\}$ of the points $\{X_i\}$ by the following steps that are the same as in LLE.

Step 1: For each data point X_i, identify its k nearest neighbors by kNN algorithm or $\varepsilon -$ ball algorithm;

Step 2: Compute the reconstruction weights that minimize the error of linearly reconstructing X_i by its k nearest neighbors;

Step 3: Construct the optimal local combination weights $\{\Omega_{\iota\varphi}\}$ of training point xi in a linear combination of its selected neighbors;

Step 4: Compute the low-dimensional embedding Y for X that best preserves the local geometry represented by the reconstruction weights;

Step 1 is typically done by using Euclidean distance to define neighborhood, although more sophisticated criteria may also be used, such as Euclidean distance in kernel space or cosine distance.

After identifying the k nearest neighbors of points X_i, *Step 2* seeks to find the best reconstruction weights. Optimality is achieved by minimizing the local reconstruction error for X_i,

$$\varepsilon_i(W) = \arg\min \left\| X_i - \sum_{j=1}^{k} W_{ij} X_j \right\|^2 \tag{2}$$

which is the squared distance between X_i and its reconstruction. And the weights are subject to following constraints.

$$\begin{cases} \sum_{j=1}^{k} W_{ij} = 1 & \text{if } X_j \in N_i(X_i) \\ W_{ij} = 0 & \text{if } X_j \notin N_i(X_i) \end{cases} \tag{3}$$

Minimizing ε_i subjected to the constraints is a constrained least squares problem. After repeating

Step 3 of the LLE algorithm is to compute the best low dimensional embedding Y based on the weight matrix W obtained from *Step 2*. This corresponds to minimizing the following function subject to the constraints $Y_{d \times n} e_n = 0_d$ and $Y_{d \times n} Y_{d \times n}^T = n I_{d \times d}$.

$$\varepsilon(Y) = \arg\min \left\| Y_i - \sum_{j=1}^{k} W_{ij} Y_j \right\|^2 \tag{4}$$

Based on the weighted matrix W, we can define a sparse, symmetric, and positive semi-definite matrix M as follows:

$$M = (I - W)^T (I - W) \tag{5}$$

Note that Eq.(5) can be expressed in a quadratic form $\varepsilon(Y) = \sum_{ij} M_{ij} Y_i Y_j^T$, where $M = \left[M_{ij} \right]_{n \times n}$. By the Rayleigh-Ritz theorem, minimizing Eq.(3) can be done by finding the eigenvectors with the smallest (nonzero) eigenvalues of the sparse matrix M.

3   SVM Classifier

Support vector machine (SVM) learning is an area of statistical learning, subject to extensive research [16-19]. The name of SVM comes from the fact that the final classification function of SVM usually only depends on part of the training samples, which are called support vectors, because they "support" the boundary. The SVM has been used extensively for a wide range of applications in science, medicine and engineering, and successively extended by a number of other researchers. The SVM is a relatively new type of machine learning model, which is of remarkably robust performance with respect to sparse and noisy data.

A MATLAB toolbox implementing SVM is freely available for academic purposes, it can be download from: http://www.isis.ecs.soton.ac.uk/resources/svminfo/.

4   Numerical Experiments

The classification method we propose is based on the SLLE, useful to deal with data sets containing multiple manifolds, corresponding to different classes. For each x_i from a class g ($1 \le g \le G$) a set of k neighbors is defined by selecting the closest points (in terms of Euclidean distance) to x_i belonging to the same class g. The first step consists in mapping the n samples of a training set into a low-dimensional embedding space, following the SLLE procedure described in Section 2, with local neighborhoods made of observations belonging to the same class. So, the G classes are mapped in a space in which they are more separated. A new $d \times n$ data matrix Y is

derived, which contains the training set data projections. In order to derive the embedding projection $y*$ of $x*$, a simpl solution is to use the linear approximation, which requires the derivation of the least square projection matrix A of the data on the embedding. Thus the embedding coordinates $y*$ of the observation $x*$ can be obtained by the pseudo-inverse of such matrix A:

$$y^* = (A^T A)^{-1} A^T x^* \tag{6}$$

Then the new observation data are assigned to the class having the nearest centroid in the reduced d-dimensional space.

To verify the classification abilities of our algorithm, we perform experiments on high dimensional real-world data on leaf image dataset. Three kinds of plant leaf image sets were experimented. Each image in the database has been size-normalized and centered to 28×28 gray-level images, so the dimensionality of the digit space is 784. We compare our method with the method PCA. We set $k=8$, $d=14$. After SLLE, SVM is applied to classifying plants. Table 1 is shown the recognition results.

Table 1. Classification rates versus dimensionality

No.	Methods	Accuracy (%)	Dimension
1	PCA	80.29±7.34	9
2	SLLE	92.65±4.44	6

To find how the adjustable parameter α affects the recognition performance, we change α from 0.0 to 1.0 with step 0.1. Fig.2 displays the average recognition rates with varied parameter α by carrying out ODP. From Fig. 3, it can be found that ODP obtains the best recognition rate when $\alpha =0.7$.
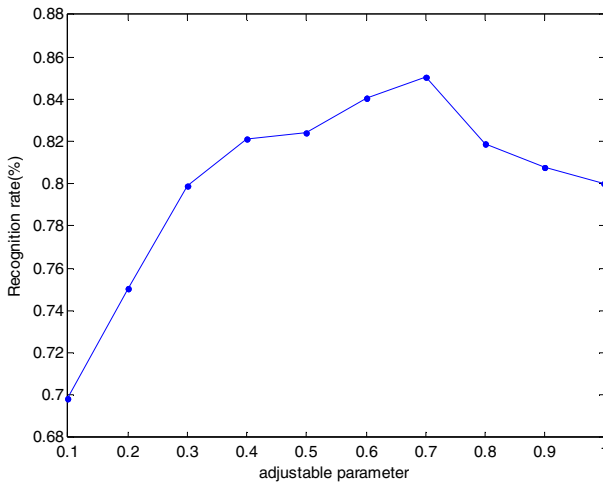
Fig. 1. Recognition rates versus α on leaf images

5   Conclusions

In this paper, we have proposed a supervised plant classification method, which performs version of supervised LLE on the data points in the manifold using a SLLE algorithm. The experiment results on a number of plant leaf image datasets demonstrated that SLLE is a powerful feature extraction method, which when coupled with simple classifiers can yield very promising recognition results. SLLE seems to be mainly applicable to high-dimensional data sets which clearly exhibit manifold structure.

Acknowledgment

References

1. Yonekawa, S., Sakai, N., Kitani, O.: Identification of Idealized Leaf Types Using Simple Dimensionless Shape Factors by Image Analysis. Transaction of the ASAE 39(4), 1525–1533 (1996)
2. Abbasi, S., Mokhtarian, F., Kittler, J.: Reliable Classification of Chrysanthemum Leaves through Curvature Scale Space. In: Proceeding of International Conference on Scale-Space Theory in Computer Vision, pp. 284–295 (1997)
3. Mokhtarian, F., Abbasi, S.: Matching Shapes with Self-Intersection: Application to Leaf Classification. IEEE Transaction on Image Processing 13(5), 653–661 (2004)
4. Li, Y.F., Zhu, Q.S., Cao, Y.K., Wang, C.L.: A Leaf Vein Extraction Method Based On Snakes Technique. In: Proceedings of IEEE International Conference on Neural Networks and Brain, pp. 885–888 (2005)
5. Du, J.X., Huang, D.S., Wang, X.F., Gu, X.: Shape Recognition Based on Neural Networks Trained by Differential Evolution Algorithm. Neurocomputing 70(4), 896–903 (2007)
6. Du, J.X., Huang, D.S., Gu, X.: A Novel Full Structure Optimization Algorithm for Radial Basis Probabilistic Neural Networks. Neurocomputing 70(1), 592–596 (2006)
7. Du, J.X., Huang, D.S., Wang, X.F., Gu, X.: Shape Recognition Based on Radial Basis Probabilistic Neural Network and Application to Plant Species Identification. In: Wang, J., Liao, X.-F., Yi, Z. (eds.) ISNN 2005. LNCS, vol. 3497, pp. 281–285. Springer, Heidelberg (2005)
8. Wan, Y.Y., Du, J.X., Huang, D.S.: Bark Texture Feature Extraction Based on Statistical Texture Analysis. In: Proceedings of The 2004 International Symposium on Intelligent Multimedia, Video & Speech Processing (ISIMP 2004), Hong Kong, China, pp. 482–485 (2004)
9. Du, J.X., Huang, D.S., Wang, X.F., Gu, X.: Computer-aided plant species identification (capsi) based on leaf shape matching technique. Transactions of the Institute of Measurement and Control 28 (2006)
10. Saitoh, T.K.: Takeshi.: Automatic recognition of wild flowers. In: Proceedings of 15th International Conference on Pattern Recognition (ICPR 2000), vol. 2 (2000)
11. Heymans, B.C., Onema, J.P., Kuti, J.O.: A neural network for opuntia leaf-form recognition. In: Proceedings of IEEE International Joint Conference on Neural Networks (1991)

12. Brendel, T., Schwanke, J., Jensch, P., Megnet, R.: Knowledge based object recognition for different morphological classes of plants. In: Proceedings of SPIE, vol. 2345 (1995)
13. Gouveia, F., Filipe, V., Reis, M., Couto, C., Bulas-Cruz, J.: Biometry: the characterization of chestnut-tree leaves using computer vision. In: Proceedings of IEEE International Symposium on Industrial Electronics, Guimaraes, Portugal (1997)
14. de Ridder, D., Duin, R.P.W.: Locally linear embedding for classification. Technical Report PH-2002-01, Pattern Recognition Group, Dept. of Imaging Science & Technology, Delft University of Technology, Delft, The Netherlands (2002)
15. Ridder, D., De, K.O., Okun, O., Pietikainen, M., Duin, R.P.W.: Supervised locally linear embedding. In: Kaynak, O., Alpaydın, E., Oja, E., Xu, L. (eds.) ICANN 2003 and ICONIP 2003. LNCS, vol. 2714, pp. 333–344. Springer, Heidelberg (2003)
16. Wang, S.L., Wang, J., Chen, H.W., et al.: SVM-based tumor classification with gene expression data. In: Li, X., Zaïane, O.R., Li, Z.-h. (eds.) ADMA 2006. LNCS (LNAI), vol. 4093, pp. 864–870. Springer, Heidelberg (2006)
17. Guyon, I., Weston, J., Barnhill, S., Vapnik, V.: Gene selection for cancer classification using support vector machines. Machine Learning 46(1-3), 389–422 (2002)
18. Vapnik, V.N.: Statistical learning theory. Wiley Inter science, New York (1998)
19. Cristianini, N., Shawe-Taylor, J.: An Introduction to Support Vector Machines. Cambridge University Press, Cambridge (2000)

A Semi-Automated Dynamic Approach to Threat Evaluation and Optimal Defensive Resource Allocation

Huma Naeem[1], Asif Masood[1], Mukhtar Hussain[1], and Shoab A. Khan[2]

[1] Department of Computer Science,
Military College of Signals
`jinnny@gmail.com, amasood@mcs.edu.pk, mukhtar.dr@gmail.com`
[2] Department of Computer Science,
College of Electrical and Mechanical Engineering
National university of Science and Tehcnology, Pakistan
`kshoab@yahoo.com`

Abstract. This paper presents a decision support based, dynamic approach to optimal threat evaluation and defensive resource scheduling. The algorithm provides flexibility and optimality by swapping between two objective functions, based on preferential and subtractive defense strategies, as and when required. Analysis part of this paper presents the strengths and weaknesses of the proposed algorithm over an alternative greedy algorithm as applied to different offline scenarios.

Keywords: Automated Threat Evaluation (TE) and Weapon Assignment (WA) - (TEWA), Optimization Algorithm, Decision Support System, Artificial intelligence – Cybernetics application, preferential/subtractive defense strategies, DA based weapon allocation.

1 Introduction

TEWA is a complex system that maintains a closed-loop interaction with a non-deterministic and ever-changing environment. TEWA is generally divided into two sub-processes i.e. threat Evaluation (TW) and Weapon Assignment (WA). According to [1], and as such no exact methods exist for the solution of even relatively small sized problems [2]. For an efficient TEWA system there is a need to create a balance between effectiveness and efficiency of weapon systems [4], [5]. Manual TEWA systems cannot provide optimality because of different limitations like human vision range constraint and established situational picture (limited awareness and information). It is also dependent on an operator's experience, observation, understanding of the situation and mental condition. Humans are prone to errors especially in stressful conditions. As a consequence in military domain, when facing a real attack scenario, an operator may fail to come up with an optimal strategy to neutralize targets. This may cause a lot of ammunition loss with an increased probability of expensive asset damage. The situation is more complex when there are multiple potential threats. Most semi-automated TEWA systems usually work on target-by-target basis using

some type of greedy algorithm thus affecting the optimality of the solution and failing in multi-target scenario [6], [7]. This paper relates to the design, simulation and analysis of a two-stage flexible dynamic decision support based optimal threat evaluation and defensive resource scheduling algorithm for multi-target air-borne threats. The algorithm provides flexibility and optimality by swapping between preferential and subtractive defense strategies as and when required. It provides a near optimal solution to the defense resource allocation problem while maintaining constraint satisfaction. Proposed algorithm uses a variant of many-to-many Stable Marriage Algorithm (SMA) to solve Threat Evaluation (TE) and Weapon Assignment (WA) problem. TE is a two stage process where first stage encompasses threat ranking while in second stage Threat-Asset pairing is done. For WA, we use a flexible dynamic approach, allowing multiple engagements using shoot-look-shoot strategy, to compute near-optimal solution for a range of scenarios. Analysis part of this paper presents the strengths and weaknesses of the proposed algorithm over an alternative greedy algorithm as applied to different offline scenarios.

    Section 2 of this paper presents an overview of TE and WA processes along with dynamic closed loop weapon allocation and defense strategies discussion. Section 3 presents our approach to TEWA. The performance of proposed model is compared with an alternative greedy algorithm and strengths and weaknesses of proposed algorithm are briefly discussed in section 4. Section 5 concludes this report.

2 Overview of TEWA Approaches

Main purpose of TEWA can be traced back to two critical concerns related to multi-target scenarios i.e. deciding the order in which threats should be neutralized and deciding which WS to be used to neutralize a particular target. First concern relates to TE while weapon selection decision is the task of WA process. The subsequent part of this section discusses these two processes in detail. Tin G. and P. Cutler classified parameters related to TEWA into critical and sorting parameters, in [7].

2.1 Threat Evaluation (TE)

TE consists of all the analytical and operational activities designed to identify, evaluate, and manage anything which might pose a threat to identifiable targets [8]. TE is highly dependent on established situational picture i.e. target's current state estimates along with available contextual information (location of the defended assets (DAs) and attributes of Weapon Systems (WSs)) [9]. It can be seen as an ongoing two step process, where in first step, we determine if an entity intends to inflict damage to the defending forces and its interests, and in second stage we rank targets according to the level of threat they pose [10]. Typically threat evaluation has been seen from capability based TE model (models that rank targets according to their capability index, where capability index defines the ability of the target to inflict damage to a DA) and Intent based TE model (models that estimate intent of a target to calculate threat index. Intent refers to the will of a target to inflict damage to a DA). Combining intent and capability we have opportunity. This introduces a hybrid approach to TE i.e. opportunity based TE model that uses intent and capability parameters together in the form of weighted parametric equations to calculate threat index.

2.2   Weapon Allocation (WA)

The process of reactive assignment of available WSs to the identified threats is exactly what comes under WA umbrella. Real world assignment problems are complex to solve as they have an inherent element of uncertainty and hard set of external constraints. [11] – [16] discuss various approaches to WA solution. The main complexities in WA process arise due to dynamic nature of the environment that may demand the assignment problem to be re-solved. Therefore, while designing an algorithm for such a real time system, there is a need to cater for the changes in input data [10]. Different AI techniques have been proved useful for finding solutions for such optimization problems, sometimes in a more flexible and efficient way [17].

In a real scenario, there can be N number of DAs having M number of WSs of different types. Each DA has its own priority and hardness i.e. vulnerability index. The defense may either want to maximize the total expected value of the DAs that survive after all weapon target allocations and target attacks, or the defense may want to minimize the total expected survival value of the survived targets. The first approach is known as DA-based weapon target allocation – preferential defense strategy while the later one is termed as Target based weapon target allocation– a subtractive defense strategy [18]. Subtractive defense strategy can be seen as a special case of preferential defense strategy when numbers of threats per DA are small. In a multi-target scenario, if number of threats directed to each DA increases, subtractive defense fails badly. On the other hand, if information available has higher element of uncertainty, preferential strategy fails [18]. Thus for an optimal solution, defense may choose to swap between these two modes of operation.

Defense strategies define the objective function of a TEWA system. But the execution model for WA may use static approach (without the notion of time) or dynamic approach (taking notion of time into account). Models that support single engagement options fall into the category of static resource allocation problem. Static open loop defense resource allocation solutions are numerous. These solutions include processes defined over a single time horizon but they fail to address smaller scale problems optimally [19]. On the contrary, a dynamic approach to weapon allocation allows multiple engagements by observing the outcomes of previous engagement before making any further engagements – shoot-look-shoot strategy [20]. In a dynamic model, weapon scheduling is done in stages where each stage iterates through Boyd's Observe-Orient-Decide-Act (OODA) loop [9]. This feedback based approach is also known as a closed loop TEWA.

3   Proposed Approach

This section proposes, a two staged model for Threat Evaluation and Weapon Assignment using a variant of many-to-many Stable Marriage Algorithm (SMA) based on weighted proposals. First, assignment problem is formulated as an optimization problem with constraints and then solved using proposed scheme.

TEWA, being a DS based system, consists of correlated threat and weapon libraries. In order to formulate the optimization problem, we need the information about the kill probabilities and lethality index for each weapon-target pair. This information comes from threat and weapon libraries. Our solution is based on two way correlation

between weapon and threat library. Like any real time system, TEWA is subject to uncertainties. To cater for this incomplete information, we chose to switch between subtractive and preferential defense strategies as and when required. The solution considers one or more type of WSs possessed by a set of DAs against a set of threats k, and allows for multiple target assignment per DA. The model is kept flexile to handle unknown type of threats.

3.1   Proposed Threat Evaluation Model

For the formulation of TE problem, we consider that the defense has i numbers of DAs to be guarded against potential threats. Where each DA has its own set of one or more type of WSs. Suppose total number of WSs is j. Assume DAs are attacked by k number of threats of any kind and capability. The main objective of TE process is to rank the targets according to their opportunity index and assign each target to the most suitable DA. The most suitable is used as a flexible term here; it means a DA that has sufficient resources in terms of capability and load on WSs. The output of this process is a ranked list (AssignedThreats) per DA, holding threat entities ranked according to opportunity index of each threat at time t. If AssignedThreatsi (T1,t) < AssignedThreatsi (T5,t) it means T5 is more threatening to Assigned DAi than T1 at time t. The following notations will be used.

I	$\overset{\text{def}}{=}$	Number of DAs
J	$\overset{\text{def}}{=}$	Total number of WSs
K	$\overset{\text{def}}{=}$	Number of Threats
DA	$\overset{\text{def}}{=}$	Set of all DAs
WS	$\overset{\text{def}}{=}$	Set of all WSs
Threats	$\overset{\text{def}}{=}$	Set of all identified threats
T_i	$\overset{\text{def}}{=}$	Set of threats aimed for DA_i where i= 1,2,3,… I and T_i € DA
N_i	$\overset{\text{def}}{=}$	Number of targets aimed at DA_i i.e. $N_i = |T_i|$ where i= 1, 2, 3,.. I
P_i	$\overset{\text{def}}{=}$	Priority / vulnerability index of DA_i, i= 1,2,.. I
TCT	$\overset{\text{def}}{=}$	Threshold on maximum number of threats assigned, this threshold is used for load balancing.
TMS	$\overset{\text{def}}{=}$	Threshold to swap the defensive strategy used – this threshold, in a way shows total number of threats that can be handled optimally by subtractive strategy.
LI_j	$\overset{\text{def}}{=}$	Lethality index of WS_j where j= 1,2,3…. J
$Load_j$	$\overset{\text{def}}{=}$	Load on WS_j where j= 1,2,3…. J
Status	$\overset{\text{def}}{=}$	Shows status of DA_i where i= 1,2,…. I
$Cond_j$	$\overset{\text{def}}{=}$	Condition of WS_j where j= 1,2,3…. J. This is used in shoot-look-shoot strategy, i.e. during damage assessment, condition of WS is checked. If it is destroyed, it is removed from WS set.

Proposal$_{i,k}$ $\overset{\text{def}}{=}$ Proposal Object holding threat$_k$ and DA$_i$ object. Threat k proposes to the DA$_p$ only. Where DA$_p \in$ DA and DA$_p$ hold instance of DAs that have K.C high enough to neutralize k.

K.C$_{i,k}$ $\overset{\text{def}}{=}$ kill capability of DF$_i$ for target$_k$, where i= 1,2,3,… I and k= 1,2,, K. This capability index shows the probability that threat k will be neutralized if assigned to DA k.

Weight$_{param}$ $\overset{\text{def}}{=}$ Weight assigned to parameter param. Where param $\in$ {Lethality, Capability and Intent index, Load, Arrival Time, Proposal}

OI$_k$ $\overset{\text{def}}{=}$ Opportunity index of threat k, where k=1,2….K. it is calculated using capability, intent   parameters along with default lethality index of threat k. the opportunity index of threat k is given by:

$$OI_k = \left((W_{CapIndex} * CIndex_k) + (W_{IntentIndex} * IIndex_k) + ILindex_k \right) \quad (1)$$

Where,

$$CIndex_k = \text{ Capability index of threat k.}$$
$$IIndex_k = \text{ Intent index of threat k}$$
$$I\,Lidex_k = \text{ Initial lethality index of threat k}$$
$$W_{xxx} = \text{ Weight for each parameter}$$

K.C$_{i,k}$ $\overset{\text{def}}{=}$ Shows probability that a threat k will be neutralized if assigned to some WS $\in$ DA$_k$ is given by:

$$\left(\prod_{k=1}^{K} \left(1 - \left((W_{OppIndex} * OI_k + W_{Load} * Load_j) * C_{j,k} \right)^{B_{i,j}} \right) \right) \quad (2)$$

Where,

$$C_{i,j} = \begin{cases} 1 & if \text{ LIndex}_j > \text{Threshold}_{capability} \\ 0 & Otherwise \end{cases} \quad (3)$$

$$B_{i,j} = \begin{cases} 1 & if \text{ j} \in \text{WS}_i \; i.e \; j \; blongs \; to \; WS \; of \; DA \; i \\ 0 & Otherwise \end{cases} \quad (4)$$

The probability that a DA i $\in$ I, survives without any loss is given by:

$$\prod_{k=1}^{K} \left(1 - \left((W_{OppIndex} * OppIndex_k \;) * P_{i,k} \right) \right) \quad (5)$$

Where, **P$_{i,k}$** = Probability that threat k will be neutralized by Asset k and k $\in$ T$_i$. Here, mutiple targets can be assigned to DA$_i$ where i={1,2…I}, subject to the number of WSs avaialbility, effectiveness and load on each WS.

Decision variable depends on a complex parametric equation and will be denoted by Assign$_{i,k}$. As stated earlier, the main task of TE is to rank targets and perform threat-Asset pairing. This pairing is based on a **Assign$_i$** as given by equation 6:

$$Assign_{i,k} = \begin{cases} 1 & if \text{ i} \in \text{I and DA}_i \; accepts \; \text{Porposal}_{i,k} \\ 0 & Otherwise \end{cases} \quad (6)$$

This is subject to following constraint:

$$\sum_{i=1}^{I} Assigned_{i,k} = 1 \tag{7}$$

After assigning initial threat index, TE processes all identified threats to calculate refined threat index based on opportunity parameters. This refined threat index specifies the order in which threats should be processed for WA. Proposed solution uses one-to-one stable marriage algorithm (SMA) with weighted proposals between threats and DAs. For each threat, list of matching capable DAs is searched on the basis of kill capabilities (index), status of DA (Free to Fire, On Hold, and Freeze) and priorities of DAs. So, a threat k, proposes to only those DAs that have capability high enough to neutralize it. Each proposed DA, looks for proposal acceptance feasibility and responds accordingly.

Weights are assigned to each DA on the basis of DA priority, vulnerability index and threat related parameters like Time to DA (TDA), heading and velocity etc. For each threat k, the proposal with the maximum weight is processed first. If load on DAi allows accepting this proposal, threat k is assigned to proposed DAi, else the next highest weighted proposal is processes. Mostly TDA calculations are based on constant speed that is quite unrealistic. Our proposed algorithm caters for this deficiency by calculating speed and velocity along different axis at different time stamps, making it scale up its efficiency.

3.2 Proposed Weapon Allocation Model

The problem to be solved here is to assign members of a set of threats to the members of a set of Weapon Systems (WSs) present in assigned DA i.e. creating matches between the elements of two disjoint sets. This matching is subject to following constraints:

Just like TE phase, we use weighted proposals for WA process. The outcome of each proposal can be 1 (accepted) or 0(rejected).

$$Assign_{j,k} = \begin{cases} 1 & if\ j \in WS_i\ and\ proposal_{i,k} = 1\ and\ Porposal_{j,k} = 1 \\ 0 & Otherwise \end{cases} \tag{8}$$

For each DA-threat pair, proposals are only sent to those WSs that belong to corresponding DA and have capability high enough to neutralize assigned target.

$$proposal_{j,k} = \begin{cases} 1\ if\ j \in J\ and\ WS_i\ shows\ WSs\ of\ DA_i \\ \quad that\ has\ accepted\ the\ Porposal_{i,k}\ Capability \\ \quad Index\ of\ WS_i > Required\ Minimum\ Value \\ \\ 0\ \ Otherwise\ (no\ proposal\ is\ sent) \end{cases} \tag{9}$$

This is subject to following constraints: At most two threats can be assigned on a WS_j. where J=1,2…I. One threat can be locked and one can wait in the queue of a WS. For a WS_j this can be expressed as:

$$\sum_{k=1}^{K} Assigned_{j,k} \leq 2\ where\ j \in WS_i \tag{10}$$

$$\sum_{j=1}^{J} Locked_{j,k} = 1 \tag{11}$$

Where,

$$Locked_{j,k} = \begin{cases} 1 & if\ Assign_{j,k} = 1\ and \\ & k\ is\ not\ enqueued\ at\ j \\ 0 & Otherwise \end{cases} \qquad (12)$$

The solution monitors total number of threats and uncertainty level. When number of threats exceeds TMS, system shifts to preferential mode, focusing on to maximizing the total expected value of the DAs that survive after all allocations and attacks.

Using threat-weapon correlation, WA creates a preference list of WS in general for the assigned threat. Algorithm then searches for matching WSs from the set of WSs in hand. Let this set be denoted by CWS (Candidate Weapon Systems) where each WS belonging to CWS has a high enough capability to neutralize the assigned threat. Using other parameters of threat like heading, course, direction, algorithm finds a subset of CWS that have or are expected to have this threat in their range at some timestamp t. let this set be represented by CWS'. For each WS' belonging to CWS', a temporary pairing of threat and WS' is made. For each pair, algorithm calculates time to WS', distance from WS', tome of flight (TOF for WS'), required elevation angle for WS', entry/exit points along with lead calculations and launch points based on velocity of threat and TOF of WS'. For each temporary pair, algorithm calculates the weight of pair using a parametric weighted equation using all significant parameters calculated in WA process. The pair that has the maximum weight is expected to be a real pair. A proposal is sent to the weapon system WS' of selected temporary pseudo pair. If it is accepted, threat is assigned to WS' else a new proposal is sent to the weapon system WS' belonging to next pseudo pair.

4   Testing and Analysis

Proposed algorithm was tested and analyzed for a range of scenarios. Before running actual scenarios, we did the typical battlefield pre-processing simulation tasks i.e.

- **DA definition:** We defined I number of DAs {DA_1, DA_2… DA_i}, each having its own set of characteristics like priority, status (free to fire, on hold, freeze) and number of firing units etcetera).
- **Weapon Description:** We defined one or more type of WS for each member of DA set.
- **Communication setup:** Threat scenario generator and TEWA application were implemented and tested on two different machines to observe the optimality. Communication server was responsible for all the communication (socket based) between test scenario generator and TEWA system.

After this battlefield pre-processing, we generated different scenarios to test the optimality of implemented system. These scenarios ranged from relaxed K target scenario against J WSs where J>K or J< K or J=K to extremely stressful scenarios where time for scheduling was less or partial information was available to schedule targets. In each scenario we added some amount of uncertainty and threat maneuvering. Percentage of threat maneuvering implicitly shows the difficulty of a scenario. We increased the amount of complexities in terms of greater number of threats per

Table 1. Simulation Summary

Algorithm	I DA Count	J WS Count	K Threat Count	% OA	% UA	% MA	SR	M	% AL	Asset Loss %	% ON
	5	30 (E) (SC)	10 (Eq)	98.02	0%	10%	Masked	S	15%	0%	15%
Proposed Algorithm	5	3o (E) (SC)	10 (Eq)	98.66	0%	10%	Catered	P	0%	0%	0%
Greedy	5	30 (U) (DC)	10 (R)	80.14	5%	10%	Partially Masked	S	21%	5%	~21%
Proposed Algorithm	5	30 (U) (DC)	10 (R)	98.82	5%	10%	Catered	P	0%	0%	0%
Greedy	5	30 (E) (SC)	30 (Eq)	50.00	10%	20%	Un Masked	S	30%	19%	30%
Proposed Algorithm	5	3o (E) (SC)	30 (Eq)	92.14	10%	20%	Catered	P	0%	0%	0%
Greedy	5	30 (U) (DC)	30 (R)	50.00	12%	20%	Un Masked	S	30%	25%	30%
Proposed Algorithm	5	3o (U) (DC)	30 (R)	92.14	12%	20%	Catered	M	0%	0%	0%
Greedy	5	30 (E) (SC)	50 (Eq)	23.43	15%	40%	Un Masked	S	54%	32%	54%
Proposed Algorithm	5	3o (E) (SC)	50 (Eq)	89.75	15%	40%	Catered	M	0%	~5%	0%
Greedy	5	30 (E) (SC)	50 (R)	20.31	20%	40%	Un Masked	S	66%	45%	66%
Proposed Algorithm	5	3o (E) (SC)	50 (R)	82.98	20%	40%	Catered	M	0%	~9%	0%

Column headers: OA= Optimality Achieved, UA= Uncertainty Added, MA= Maneuvering Added, SR= System response, M= Mode, AL= Asset Loss, ON= Over Neutralization
Table Values: (E) = Even Distribution of WS; **(U)** = Uneven Distribution of WS for each DA; **(SC)** =all WSs have same capability; **(DC)** = different WSs have different capabilities. **(Eq)=** Threat-DA distribution is equal; **(R)** = random distribution of threats. **System Response =** Catered Means system successfully responded to the changing environment and threat maneuvering. Mode: **S**= Subtractive, **P**=Preferential, **M**= Mixed

DA, fewer numbers of WS, increased amount of uncertainty and threat maneuvering. We monitored the overall system behavior and categorized system response as Unmasked, Partially Masked, Masked and Catered.

As we increased the complexities, we monitored whether the implemented system was able to handle added amount of complexities and threat maneuvering. We took average of each simulation and divided the response into four classes where Unmasked corresponds to a value of 1-25%, Partially Masked maps to 26-50%, Masked corresponds to 51-75% and Handled shows 76-100%. We considered ammunition loss as the amount of ammunition wasted in an attempt to over-neutralize a threat. We calculated asset loss for each threat that managed to cross a DA; it was assigned to, without being neutralized. Damage calculation was based on capability index of a threat times the vulnerability index of a DA combined with number of WSs deployed in that DA.

Table 1 shows summary of few simulations. The greedy algorithm used for comparison, looks for minimum arrival time without looking at the capability and intent values. It always operates under subtractive defense strategy objective Due to the lack of synchronization between WSs; greedy algorithm often results in over-neutralization whereas, proposed algorithm makes sure that at a time one and only one WS neutralizes a particular target. In case it survives previous engagement a new engagement is done using the same algorithm. Although this algorithm is computation intensive but, even under stressful conditions, it succeeds in coming up with near optimal solution where an alternative greedy algorithm based on arrival time fails.

5   Conclusion

A Two-Staged dynamic decision support based optimal threat evaluation and defensive resource scheduling algorithm for multi air-borne threats is presented that correlates threat and weapon libraries to improve the solution quality. This paper explains the main optimization steps required to react to changing complex situations and provide near optimal weapon assignment for a range of scenarios.

Proposed algorithm uses a variant of many-to-many Stable Marriage Algorithm (SMA) to solve Threat Evaluation (TE) and Weapon Assignment (WA) problem. In TE stage, Threat ranking and Threat-Asset pairing is done. Stage two is based on a new flexible dynamic weapon scheduling algorithm, allowing multiple engagements using shoot-look-shoot strategy, to compute near-optimal solution for a range of scenarios. The algorithm provides flexibility and optimality by swapping between two objective functions, based on preferential and subtractive defense strategies as and when required. Analysis part of this paper shows that this new approach to TEWA provides an optimal solution for a range of scenarios.

References

1. Lloyd, S.P., Witsenhausen, H.S.: Weapons Allocation Is NP-complete. In: Proceedings of the 1986 summer computer simulation conference, pp. 1054–1058 (1986)
2. Ahuja, R.K., Kumar, A., Jha, K.C., Orlin, J.B.: Exact and Heuristic Methods for the Weapon Target Assignment Problem 55, 1136–1146 (2007)

3. Hosein, P.A., Athans, M.: An Asymptotic Result for the Multi-stage Weapon Target Allocation Problem. In: Proceedings of the 29th IEEE Conference on Decision and Control, vol. 1, pp. 240–245 (1990)
4. Allouche, M.K.: A Pattern Recognition Approach to Threat Stabilization, tech. rep., DRDC Valcartier (2006)
5. Allouche, M.K.: Real-time Use of Kohonen's Self-organizing Maps for Threat Stabilization. In: Proceedings of the 8th International Conference on Information Fusion, vol. 6, pp. 153–163 (2005)
6. Gregory, R.B., Palma, L.: Calif.: Knowledge Based Automatic Threat Evaluation and Weapon Assignment, US Patent no – 5992288 (1999)
7. Tin, G., Cutler, P.: Accomodating Obstacle Avoidance in the Weapon Allocation Problem for Tactical Air Defense. In: 9th International Conference on Information Fusion, pp. 1–8 (2006)
8. Borum, R., Fein, R., Vossekuil, B., Berglund, J.: Threat Assessment: Defining an Approach for Evaluating Risk of Targeted Violence. Behavioral Sciences and the Law 17, 323–405 (1999)
9. Roux, R.N., Vuuren, J.H.: Threat Evaluation and Weapon Assignment Decision Support: A Review of the State of the Art. ORiON 23, 151–186 (2007)
10. Paradis, S., Benaskeur. A., Oxenham, M.G., Cutler, P.: Threat Evaluation and Weapon Allocation in Network-centric Warfare. In: Proceedings of the Seventh International Conference Fusion, "Usion 2004", Stockholm (2005)
11. Pugh, G.E.: Lagrange Multipliers and the Optimal Allocation of Defense Resources. Operations Research 12, 543–567 (1964)
12. Bellman, D., Gross, J.: On the Computational Solution of Dynamics Programming Processes: XIV: Missile Allocation Problems, RM 2282, RAND, Santa Monica California (1959)
13. Cohen, N.D.: An Attack-defense Game with Matrix Strategies, The Rand Corporation, Memorandum RM-4274-1-PR (1966)
14. Lee, Z.J., Lee, C.Y., Su, S.F.: An Immunity Based Ant Colony Optimization Algorithm for Solving Weapon-target Assignment. Applied Soft Computing 2, 39–47 (2002)
15. Lee, Z.J., Su, S.F., Lee, C.Y.: A Genetic Algorithm with Domain Knowledge for Weapon-Target Assignment Problems. Journal of the Chinese Institute of Engineers 25(3), 287–295 (2002)
16. Johansson, F., Falkman, G.: A Bayesian Network Approach to Threat Evaluation with Application to an Air Defense Scenario. In: Proceedings of 11th International Conference on Information Fusion, vol. 10, pp. 1–7 (2008)
17. Deepak, S.H.: A Comparison of Agent based Optimization Approaches Applied to the Weapon to Targets Assignment Problem (2006)
18. Hossein, P.A., Athans, M.: Preferential Defense Strategies. Part 1: The Static Case (1990)
19. Berger, J., Leong, D.: The Resolution of an Open Loop Resource Allocation Problem using a Neural Network Approach. In: 27th Annual Simulation Symposium, pp. 51–58 (1994)
20. Hosein, P.A., Walton, J.T., Athans, M.: Dynamic Weapon Target Assignment Problem with Vulnerable C2 Nodes, by Massachusetts Inst. of Tech. Cambridge Lab for Information and Decision Systems (1988)

Retracted: Quantum Quasi-Cyclic Low-Density Parity-Check Codes

Dazu Huang[1,2], Zhigang Chen[1], Xin Li[1,2], and Ying Guo[1]

[1] School of Information Science and Engineering, Central South University,
Changsha, 410083 China
[2] Department of Information Management, Hunan College of Finance and
Economics, Changsha, 410205, China

Abstract. In this paper, how to construct quantum quasi-cyclic (QC) low-density parity-check (LDPC) codes is proposed. Using the proposed approach, some new quantum codes with various lengths and rates of no cycles-length 4 in their Tanner graph are designed. In addition, the presented quantum codes can be efficiently constructed with large codeword length. Finally, we show the decoding of the proposed quantum QC LDPC.

Keywords: Quantum code, Quasi-cyclic Low-density Parity-check Codes, encoding and decoding, CSS code.

1 Introduction

The first quantum code, introduced by Shor in 1995 [1], encoded one qubit into nine qubits and could correct both bit-flip errors and phase-flip errors. Shortly after, it was shown that QECC can be constructed based on classical block codes. This led to the development of an important class of QECC by Calderbank, Shor and Steane [2, 3], known as CSS codes. Afterwards, a more general quantum codes, i.e., stabilizer quantum code, has been advanced by Gottesman [4]. Currently, almost all of the advanced code constructions may be categorized as these two classes according to the different construction methods and principles.

A class of classical code called LDPC codes was firstly proposed by Gallager [5] and the comeback of more research many years latter [6, 7]. The advantage of LDPC codes are lower decoding complexity, but its disadvantage is higher encoding complexity to get the optimal LDPC codes. LDPC codes are classified into regular and irregular one according to the weight of rows and columns in parity-check matrix. With the development of quantum information, it is necessary to generalize these notions to quantum codes, and has been proposed recently [8]. In 2005, Thomas Camarain et al. [9] proposed a general construction of quantum LDPC codes within the stabilizer formalism. In addition, the decoding of this kind of quantum code evolved pure quantum nature so that it has some novel properties. Furthermore, the quantum rate can be easily adjusted with choosing proper block matrix of generator matrix.

Since CSS code constructed by two classical linear cods, if based on LDPC codes, then it is called quantum LDPC codes. In this paper, the main idea is that: by constructing circulant permutation matrix, we can obtain quantum quasi-cyclic LDPC codes using CSS's method. Therefore, in Section 2, we propose a method using classical QC LDPC codes for constructing quantum LDPC respectively and analyze the cycle of this kind construction. Finally, the encoding and decoding of constructed quantum QC LDPC codes are studied.

2   Constructions of Classical Codes for QECC

The characteristic of QC LDPC code is that, the parity-check matrix consists of small square blocks which are the zero matrix or circulant permutation matrices. With sum-product decoding, the QC LDPC codes has good performance for short code length. Therefore, the corresponding quantum QC LDPC codes[10, 11] have encoding advantage over the other types of quantum codes[12,13].

To construct the parity check matrices H_1 of classical QC LDPC code C_1 and H_2 of $C_2^{\perp}$, we should show how to obtain the circulant permutation matrixes which H_1 and H_2 obtained from.

First, we introduce some notions. A $L \times L$ square matrix $P = (p_{ij})$ is called basic-matrix, which can be denoted as:

$$p_{ij} = \begin{cases} 1 \text{ if } i = (j+1)\text{mod}L, \\ 0 \text{ otherwise} \end{cases} \tag{1}$$

where L is a prime. Obviously, the row vectors of basic-matrix form a finite cyclic group. The $mL \times nL$ matrix G called circulant permutation matrix, which can be denoted by $G = (P^{a_{ij}})$, i.e.,

$$G = \begin{pmatrix} P^{a_{11}} & P^{a_{12}} & \dots & P^{a_{1n}} \\ P^{a_{21}} & P^{a_{22}} & \dots & P^{a_{2n}} \\ \vdots & \vdots & \ddots & \vdots \\ P^{a_{m1}} & P^{a_{m2}} & \dots & P^{a_{mn}} \end{pmatrix} \tag{2}$$

where a_{ij} is the exponent of matrix P. If $P^{\infty} = 0_{L \times L}$, $a_{ij} \in \{0, 1, \cdots, L-1, \infty\}$, then $E(G) = (a_{ij})$ is called exponent matrix of G, i.e.,

$$E(G) = \begin{pmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \dots & a_{mn} \end{pmatrix}, \tag{3}$$

When G has full rank, its code rate is given by

$$R = \frac{Ln - Lm}{Ln} = \frac{n - m}{n} = 1 - \frac{m}{n}, \tag{4}$$

regardless of its code length $N = nL$. If the parity check matrix of LDPC codes C is obtained by the circulant permutation matrix described above, then C is called (m, n) QC LDPC codes. Now, we first consider how to design two circulant permutation matrixes G_1 and G_2.

According to the theory of group, if L is prime, then Z_L is finite field. The set Z_L^* constructed by non-zero element in Z_L is a cyclic group, and its order should be even. Therefore, Z_L^* may be divided into two subsets averately, i.e., Z_{L1}^* and Z_{L2}^*. And Z_{L1}^* and Z_{L2}^* satisfy following conditions:

(i) $|Z_{L1}^*| = |Z_{L2}^*| = l, l = (L-1)/2$.
(ii) $\forall a \in Z_{L1}^*, b \in Z_{L2}^*, a \neq b$.
(iii) Let $a_i = \sigma^i \mathrm{mod} L, a_i \in Z_{L1}^*$, and $b_i = \tau \sigma^i \mathrm{mod} L, b_i \in Z_{L2}^*$, where $\sigma \in Z_L^*, \tau \in Z_L^*/\{\sigma, \sigma^2, \cdots\}, 1 \leq i \leq l$.

Now, we will use these elements in the two subsets to obtain exponent matrixes $E(G_1) = (a_{ij})$ and $E(G_2) = (b_{ij})$. Assume vectors $u_1 = (a_{11}, a_{12}, \cdots, a_{1l})$, $a_{1i} \in Z_{L1}^*, v_1 = (b_{11}, b_{12}, \cdots, b_{1l}), b_{1j} \in Z_{L2}^*, 1 \leq i, j \leq l$. We get the exponent matrixes of circulant permutation matrixes $E(G_1)$ and $E(G_2)$ by the following procedure:

Step1: one divides $E(G_1)$ into two square matrixes $E(G_{A1})$ and $E(G_{A2})$, i.e., $E(G_1) = (E(G_{A1}), E(G_{A2}))$

Step2: Let u_1 be first row of $E(G_{A1})$, and have clockwise (or anticlockwise) permutation for a position to get its second row.

Step3: As same principle, after l times later, $E(G_{A1})$ can be obtained.

Step4: Suppose v_1' is a some permutation of u_1 and $v_1' \neq u_1$, we make v_1' as the first row of $E(G_{A2})$.

Step5: The second row of $E(G_{A2})$ is defined as anticlockwise (or clockwise) permutation for a position to v_1'. Like previous, we can get $E(G_{A2})$.

Step6: The exponent matrix $E(G_1)$ corresponding $(l, 2l)$ QC LDPC codes with rate $1/2$ is achieved by combining the two sections.

Step7: According to mapping: $a_{ij} \to b_{ij}$, we can also get another exponent matrix $E(G_2)$.

From above method, and with basic-matrix the circulant permutation matrixes G_1, G_2 can be gained.

Example 1. Let $L = 7, (Z_1^* = \{1, 2, \cdots, 6\}), \sigma = 2$, and $\tau = 3$, then $l = (L-1)/2 = 3, Z_{A1}^* = \{1, 2, 4\}$ and $Z_{A2}^* = \{3, 5, 6\}$. Therefore, $u_1 = (2, 4, 1)$, $u_1' = (1, 2, 4)$, the exponent matrixes $E(G_1)$ is:

$$E(G_1) = E(G_{A1}, G_{A2}) = \begin{pmatrix} 2\,4\,1\,1\,2\,4 \\ 1\,2\,4\,2\,4\,1 \\ 4\,1\,2\,4\,1\,2 \end{pmatrix}. \tag{5}$$

According to mapping: $3a_{ij} = b_{ij}$, the exponent matrixes G_2 is:

$$E(G_2) = \begin{pmatrix} 6\,5\,3\,3\,6\,5 \\ 3\,6\,5\,6\,5\,3 \\ 5\,3\,6\,5\,3\,6 \end{pmatrix}. \tag{6}$$

Based on the two exponent matrixes and the 7 by 7 basic-matrix P, circulant permutation matrixes G_1 and G_2 are gained.

In the following, we will explain that, two $N_{k_i} \times N, (i = 1, 2)$ check matrixes can be generated from the two circulant permutation matrixes.

Lemma 1. If circulant permutation matrixes G_1 and G_2 are described previous, any row of G_1 is orthogonal to any row of G_2.

Proofs. Denote rows of $E(G_1)$ and $E(G_2)$ are $x = (x_{a1}, x_{a2})$ and $y = (y_{b1}, y_{b2})$ respectively, where $|x_{a1}| = |y_{b1}| = l$. According to the design described previous, the elements in vector Z_{L1}^* and Z_{L2}^* are in one-to-one correspondence. The inner product between any vector in x_{a1} and y_{b1} is either 0 or 1, and same to that in x_{a2} and y_{b2}. Because of the correspondence, 0 and 1 appear contemporarily in two parties. So, the inner product of x and y is 0. Therefore, any row of G_1 is orthogonal to any row of G_2.

Lemma 2. Suppose H_1 is the parity check matrix of classical code C_1 and H_2 is that of $C_2^{\perp}$. If any row of H_1 is orthogonal to any row of H_2, the linear codes C_1 and C_2 satisfy the condition(a).

Proofs. If any rows of H_1 and H_2 are orthogonal mutually, i.e., $\langle a, b \rangle = 0, \forall a \in H_1, \forall b \in H_2$, one can get that $H_1 \cdot d^T = 0, \forall d \in C_2$. Therefore, the cause $C_2 \subset C_1$ is satisfied. And then, according to the Hamming bound (3) and $k_2 < k_1$ in condition (b), we can get the following cause.

Theorem 1. Based on the previous principle of the circulant permutation matrixes construction, one gets quantum codes C with parameter $[[N, k, d]]$.

Proofs. According to the Hamming bound (3) and let $k_2 < k_1$, one chooses properly $N - k_i (i = 1, 2)$ rows of G_i as parity check matrix H_i. From Lemma 1, we also have that, any row of H_1 is orthogonal to any row of H_2. With $k_2 < k_1$, the condition (a), (b) are all ensured. Therefore, with Lemma 2 we can gain quantum codes C with parameter $[[N, k, d]]$.

As designed previously, if there is ∞ in exponent matrix, QC LDPC also can be obtained by the non-zero matrix.

Example 2. If the exponent matrixes of parity matrixes is in example 1, and make the mapping is $\infty \to \infty$ we can get the following exponent matrixes $E(H_1)$ and $E(H_2)$

$$E(G_{A1}) = E(G_{A1}, G_{A2}) = \begin{pmatrix} 2 & 4 & \infty & \infty & 2 & 4 \\ \infty & 2 & 4 & 2 & 4 & \infty \\ 4 & \infty & 2 & 4 & \infty & 2 \end{pmatrix} \tag{7}$$

according to mapping: $3a_{ij} \rightarrow b_{ij}$, the exponent matrixes G_2 is:

$$E(G_{A2}) = \begin{pmatrix} 6 & 5 & \infty & \infty & 6 & 5 \\ \infty & 6 & 5 & 6 & 5 & \infty \\ 5 & \infty & 6 & 5 & \infty & 6 \end{pmatrix}, \tag{8}$$

the parity check matrixes H_1 and H_2 also can be taken from G_1 and G_2 which based on the basic-marx and exponent matrixes designed above.

Theorem 2. Based on the described principles about the non-zero parity check matrixes, to $x \in Z_1^*, y \in Z_2^*$, there may be the following mapping:

$$x \rightarrow \infty, \infty \rightarrow y, \infty \rightarrow \infty \tag{9}$$

and take the proper parameters k_1 and k_2, the quantum codes $C[[N, k, d]]$ can be obtained.

Proof. Denote $x = (x_{a1}, x_{a2})$ and $y = (y_{b1}, y_{b2})$. In terms of the features of basic-matrix P, to this mapping, same value of inner product between $x_{ai}(i = 1, 2)$ and y_{bi} appears always contemporarily. Therefore, it induces $\langle x, y \rangle = 0$, and ensures the condition (a) is satisfied. According to the Hamming bound (3), two classical codes C_1 and C_2 with parameters $[[N, k_1, d_1]]$ and $[[N, k_2, d_2]]$ can be gained. If we take proper parameters k_1 and k_2, then the condition (b) can be satisfied. Therefore, the quantum QC LDPC codes $C[[N, k, d]]$ based on codes C_1, C_2 can be obtained, where $k = k_1 - k_2$ and $d = \min\{d_1, d_2\}$.

Furthermore, the design can provide codes with various lengths and rates with no cycles of length 4 in their Tanner graph. Denote $L - 1 = 2l$, the corresponding exponent chain of two times of cycle is represented as following:

$$a_1 \rightarrow a_2 \rightarrow \cdots a_{2l} \rightarrow a_1 \text{or}(a_1, a_2, \cdots, a_{2l}, a_1) \tag{10}$$

According to following Theorem, we can adjust the proper code rate.

Theorem 3. Let $(a_1, a_2, \cdots, a_{2l}, a_1)$ be the exponent chain corresponding to a 2-block-cycle. If r is the least positive integer such that

$$r \sum_{i=1}^{2l} (-1)^{i-1} a_i \equiv 0 \bmod L \tag{11}$$

then the block-cycle leads to a cycle of length $2lr$.

Proofs. First we consider columns from 1 to l. According to the relation of P^{a_i} and P^{a_i+1}, the "1" at the jth row of P^{a_i} in the row block is located at $j + \sum_{i=2}^{l} (-1)^k a_k$th row of in a row block $(i = 2, 3, \cdots, l)$ and the last is $(j + a_2 - a_3 \cdots a_{2l} - a_1)$. Similarly, we can also analyze the remaining columns from $l + 1$ to $2l$. Repeating this procedure r times, the "1" at the jth row of in the row block is connected to the "1" at the $j + r \sum_{i=1}^{2l} (-1)^i a_i$th row of P_1^a in the row block. This procedure may be represented by the following chain:

$$\underbrace{P^{a_1} \rightarrow P^{a_2} \rightarrow \cdots P^{a_{2l}}}_{r\text{times}} \rightarrow P^{a_1} \tag{12}$$

By the assumption of (12), we also have

$$j \equiv j + r \sum_{i=1}^{2l} (-1)^i a_i \bmod L \tag{13}$$

It is easily checked that the girth of the Tanner graph of a parity-check matrix we provided is greater than or equal to 6.

Finally, we consider quantum self-dual containing codes construction based on quasi-cyclic LDPC Codes. There is the following result for it.

Theorem 4. There is no self-dual containing quantum QC LDPC code which based on the construction described above.

Proof. Suppose that C_1, C_2 are a pair of linear codes which are used to construct self-dual containing CSS code. It implies $C_1 = C_2$. According to the condition (a), it implied that the parameters k_1, K_2 should be $0 \le k_1 < k_2 \le m$. It means that condition $C_1 = C_2$ cannot be obtained.

3   Encoding of Quantum LDPC Codes

In this section, we show the detail of how to encode the quantum codes via G_1 and G_2. Firstly, one randomly choices $N - k_1$ and $N - k_2$ rows in the two circulant permutation matrices. Furthermore, the distance d of the code can be calculated by the number of the independent lines of the check matrix H_{N-k_1} and H_{N-k_2}. Subsequently two classical codes C_1 and C_2 with the perimeters $[[N, k_1, d_1]]$ and $[[N, k_2, d_2]]$ are obtained.

To codes set C_1, C_2 the any quantum code $x \in C_1$, its quantum state can be expressed as:

$$|x + C_2\rangle \equiv \frac{1}{2^{k_2/2}} \sum_{y \in C_2} |x + y\rangle \tag{14}$$

If we take all of the quantum code word in C_1, the number of orthogonal state $|x + C_2\rangle$ is $2^{k_1 - k_2}$. Therefore, we can get a set including the different state:

$$\{|x_0 + C_2\rangle, |x_1 + C_2\rangle, \cdots, |x_{2^{k_1 - k_2} - 1} + C_2\rangle\}, \tag{15}$$

CSS code is to transform $k_1 - k_2$ qubits overlap state

$$\alpha_0 |0 \cdots 00\rangle + \alpha_1 |0 \cdots 01\rangle + \cdots + \alpha_{2^{k_1 - k_2} - 1} |1 \cdots 11\rangle, \tag{16}$$

as a code system of qubits overlap state:

$$\alpha_0 |x_0 + C_2\rangle + \alpha_1 |x_1 + C_2\rangle + \cdots + \alpha_{2^{k_1 - k_2} - 1} |x_{2^{k_1 - k_2} - 1} + C_2\rangle. \tag{17}$$

Therefore, by the two classical liner codes C_1 and C_2, we can get a quantum codes with parameter $[[N, k, d]]$, where $k = k_1 - k_2$ and $d = \min\{d_1, d_2\}$. For instant, the check matrixes H_1, H_1 are obtained from Example 1 and the corresponding codes can correct one error. If we choice k_1 and k_2 satisfy the Hamming bound

such as $k_1 = 17$, $k_2 = 15(< 21)$, and the distance $d = 12$. We can get the quantum codes $[[42, 2, 12]]$, and there will be 2^{17} code vectors in C_2, and 2^{15} in C_1. Therefore, we can get $2^{k_1-k_2} = 2^2$ states $|s_0\rangle \cdots |s_{2^2}\rangle$, which can be denoted as $|s_i\rangle = |x + C_2\rangle, (0 \leq i \leq 2^2 - 1)$, regardless of $x \notin C_2$, or $x \in C_2$. Finally, the quantum error-corrected coding set can be translated as following encoded code

$$\alpha_0|s_0\rangle + \alpha_1|s_1\rangle + \alpha_2|s_2\rangle + \alpha_3|s_3\rangle, \tag{18}$$

and then one sends the already encoded codes into the depolarizing channel. After translated the environment in which errors may happen, the receiver will decode his receipted quantum codes. In the following Section, we will show how to decode the presented quantum codes.

4   Decoding of the Presented Codes

Suppose the liner coding sets C_1 and C_2 can correct t qubits error, we will show the CSS codes constructed by C_1 and C_2 are how to correct most t bit- flip and phase- flip error.

4.1   Error Detection

As to the decoding of quantum CSS codes, we first consider the state in Eq.(14). Let the parity check matrix of liner codeword set C_1 be H_1, because $x \in C_1$, and $C_2 \subset C_1$, we have:

$$xH_1^T = 0, yH_1^T = 0, \forall y \in C_2 \tag{19}$$

On the other hand, if e_i denotes the vector of the flipped $i$qubit which happened on the ith qubits, after the encoded state in (17) sent, the state we accepted is:

$$|\varphi\rangle = \frac{1}{2^{k_2/2}} \sum_{y \in C_2} |x + y + e_i\rangle \tag{20}$$

According to the syndrome of vector $x + y + e_i$ corresponding to the state in Eq.(10), even $y \notin C_2$, the following equation can also be gotten:

$$(x + y + e_i)H_1^T = xH_1^T + yH_1^T + e_iH_1^T = e_iH_1^T \tag{21}$$

Because the liner codes set C_1 can correct t errors, the syndromes $e_1H_1^T, e_2H_1^T$, $\cdots, e_NH_1^T$ responding to the only one error are all different. Clearly, according to classical method of error correcting coding, we will know that the error happened in the ith qubit by computing the syndrome $e_iH_1^T$ in the accepted information state $|\varphi\rangle$.

4.2   Error Correction

From the phase of error detection, we can know that which position has happened flipped qubit. The error of flipped qubit include: bit-flip error(X error), phase-flip error(Z error), and the case of two kinds of error happened contemporarily. Now, we talk about how to correct the error.

To bit-flip error, we only impose X-Gate on the qubit which error happened on, the error will be corrected.

Example 3. let the prime $L = 3$, we can get the most simple form of exponent matrixes $E(G_1)$ and $E(G_2)$ as following:

$$E(G_1) = \begin{pmatrix} 1 & 1 \end{pmatrix} E(G_2) = \begin{pmatrix} 2 & 2 \end{pmatrix} \tag{22}$$

If suppose the parameters $k_1 = 4$ and $k_2 = 3$, the two classical codes are $C_1[[6, 4, 6]]$ and $C_2[[6, 3, 6]]$ and they all can correct one error. The constructed quantum code based on these classical codes should be $C[[6, 1, 4]]$ and can also correct one error. To correct the phase-flip, we impose Hadamard transition on every qubits i.e., $|x\rangle$, so, it is expressed as following:

$$|x\rangle \xrightarrow{\text{Hadamard}} \frac{1}{2^N} \sum_{y \in \{0,1\}^N} (-1)^{x \cdot y} |y\rangle. \tag{23}$$

And then applying coding, to impose Hadamard transition on every N qubits in the state in Eq.(14), according to in Eq.(12), we can get the following forma:

$$\frac{1}{2^{k_2/2} 2^{N/2}} \sum_{x \in \{0,1\}^N} \sum_{y \in C_2} (-1)^{(x+y)z} |z\rangle. \tag{24}$$

If use the lineal natural of liner coding sets C_2, according to

$$\sum_{y \in C_2} (-1)^{y \cdot z} = \begin{cases} 2^{k_2} & \text{if } z \in C_2^\perp \\ 0 & \text{if } z \notin C_2^\perp \end{cases} \tag{25}$$

then, form in Eq.(14) becomes:

$$\frac{2^{k_2/2}}{2^{N/2}} \sum_{z \in C_2^\perp} (-1)^{x \cdot z} |z\rangle. \tag{26}$$

Therefore, if the phase-flip happened in jth qubit in the state $|x + C_2\rangle$, there is the following state:

$$\frac{1}{2^{k_2/2}} \sum_{y \in C_2} (-1)^{(x+y)e_i} |x + y\rangle. \tag{27}$$

To impose Hadamard transition on every qubits in above equation, we can get:

$$\frac{1}{2^{k_2/2} 2^{N/2}} \sum_{z \in \{0,1\}^N} \sum_{y \in C_2} (-1)^{(x+y)(e_i+z)} |z\rangle =$$
$$\frac{1}{2^{k_2/2} 2^{N/2}} \sum_{z' \in \{0,1\}^N} \sum_{y \in C_2} (-1)^{(x+y)z'} |e_i + z'\rangle, \tag{28}$$

where $z' = e_i + z$. Applying Eq.(4)again, we also get:

$$\frac{2^{k_2/2}}{2^{N/2}} \sum_{z' \in C_2^\perp} (-1)^{(x \cdot z')} |z' + e_i\rangle. \tag{29}$$

From the above cause, the phase-flip error can be turned as bit-inverted error through the Hadamard transition. Therefore, to the state in in Eq.(18), we only correct the bit- flip error, then the phase-flip error can be corrected. In fact, we apply the parity check matrix H_2 of coding set $C_2^\perp$ to solve the syndrome of the vector $z' + e_i$ corresponding to state $|z' + e_i\rangle$. Because $C_2^\perp$ can correct t errors, from the syndrome, we can detect the error e_i and correct it. At last, one gets the state:

$$\frac{2^{k_2/2}}{2^{N/2}} \sum_{z' \in C_2^\perp} (-1)^{(x+y)z'} |z'\rangle. \tag{30}$$

And then impose Hadamard transition on every qubits in Eq.(19)again, one will get the primitive information. Finally, if the two kind flipped errors happened contemporarily, we only correct the bit-flip error because this procedure includes the correction to phase-flip error.

5   Simple Analysis

Because of its structure, quantum LDPC codes will reduce the encoding complexity. And, it can make the quantum circuit more simpler than other classes of CSS quantum codes .To Example 3, The circuit of measuring on the X error syndrome can be expressed as Figure 1. Therefore, it has more practical significance. Quasi-cyclic LDPC codes is a class of low-density parity-check codes, and its parity check matrix is circulant permutation matrix. It has more advantage to reduce the complex of both quantum encoding and decoding.
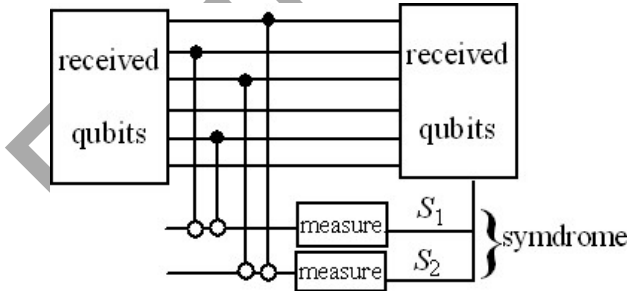
Fig. 1. Circuit correcting X errors of qubit encoded in [[6,4,6]] code

Furthermore, the design can provide codes with various lengths and rates with no cycles of length 4 in their Tanner graph. Denote $L-1 = 2l$, the corresponding exponent chain of two times of cycle is represented as Eq.(10). According to the Theorem 4, we can adjust the proper code rate.

Let $(a_1, a_2, \cdots, a_{2l}, a_1)$ be the exponent chain corresponding to a 2-block-cycle. If r is the least positive integer such that according to Eq.(11) the block-cycle leads to a cycle of length $2lr$.

6   Conclusion

In this Paper, we proposed a method to construct efficiently the quantum QC LDPC codes with the large codeword length. We showed that the constructions provide codes with various lengths and rates with no cycles of length 4 in their Tanner graph. This approach can be directly extended to other quantum QC LDPC codes. According to the features of LDPC codes, the presented codes have the advantage of simple implementation and low complexity encoding.

Acknowledgements

References

1. Shor, P.W.: Scheme for reducing decoherence in quantum computer memory. Phys. Rev. A 52, 2493 (1995)
2. Calderbank, A.R., Shor, P.W.: Good quantum error-correcting codes exist. Phys. Rev. A 54, 1098–1105 (1996)
3. Steane, A.M.: Error correcting codes in quantum theory. Phys. Rev. Lett. 77, 793 (1996)
4. Gottesman, D.: Stabilizer codes and quantum error correction. PhD thesis, California Institute of Technology, Pasadena, CA (1997)
5. Gallager, R.G.: Low density parity check codes. IRE Trans. Inform.Theory IT-8, 21–28 (1962)
6. MacKay, D.J.C., Neal, R.M.: Near Shannon limit performance of low density parity check codes. Electron. Lett. 32, 1645–1646 (1996)
7. Wiberg, N.: Codes and decoding on general graphs, Dissertation no. 440, Dept. Elect. Eng. Link?ping Univ., Link?ping, Sweden (1996)
8. MacKay, D.J.C., Mitchison, G., McFadden, P.L.: Sparse graph codes for quantum error-correction. IEEE Trans. Info. The Theory 50(10), 2315–2330 (2004)
9. Camara, J., Ollivier, H., Tillich, J.P.: (2), 0502086 (2005)
10. Li, Z.W., Chen, L., Lin, S., Fong, W., Yeh, P.S.: Efficientencoding of quasi-cyclic low-density parity-check codes. IEEE Trans. Commun. (2005)
11. Richardson, T.J., Shokrollahi, A., Urbanke, R.: Design of capacityapproaching low desnsity parity-check codes. IEEE Trans. Inform. Theory 47, 619–637 (2001)
12. Feng, K., Ma, Z.: A finite Gilbert-Varshamov bound for pure stabilizer quantum codes. IEEE Trans. Inf. Theory 50, 3323–3325 (2004)
13. Hamada, M.: Information Rates Achievable with Algebraic Codes on Quantum Discrete Memoryless Channels. IEEE Transactions on Information Theory 51(12), 4263–4277 (2005)

Contribution Degree's Application in the Research of Elements of TCM Syndromes

Rongyao Zheng, Guangcheng Xi, and Jing Chen

Key Laboratory of Complex Systems and Intelligences Science, Institute of Automation,
Chinese Academy of Sciences, 100190, Beijing, P.R. China
{rongyao.zheng,guangcheng.xi,jing.chen}@ia.ac.cn

Abstract. Using unsupervised algorithms to cluster for diagnosis information data is a mainstream and difficult area of TCM clinical research, and the optimal symptoms' number of the syndrome is even more difficult to gain. However, there is no relevant and effective research on it yet. An unsupervised clustering algorithm is proposed based on the concepts of complex system entropy and contribution degree in this work. The algorithm is based on the familiar unsupervised complex system entropy cluster algorithm, simultaneously, it introduces contribution degree to self-adaptively select the symptoms' number. This work carried out three clinical epidemiology surveys about depression, chronic renal failure and chronic hepatitis b, and obtained 1787 cases, each of which has measurements for 76 symptoms. The algorithm discovers 9 patterns, and 6 of them fit the syndrome in clinic. Therefore, we conclude that the algorithm provides an effective solution to discover syndrome from symptoms.

Keywords: contribution degree; mutual information; unsupervised cluster; syndrome; self-adaptive.

1 Introduction

Traditional Chinese Medicine (TCM) is the treasure of China with about 3000 years' history. Differentiation of syndromes is one of the most important methods of recognizing and diagnosing diseases in the theories of TCM. However, many differentiation systems of syndromes have been produced and varied during thousands of years, which brings increasing difficulties to unify these conceptions. The raise of entia of syndromes' differentiation systems is the approach to solve this problem [1]. In the past, differentiation of syndromes mostly depends on doctor's experience, which lacks of objective basis. Hence many researchers have begun to study TCM with modern techniques. There are too many available symptoms that doctors could not take all the symptoms into consideration during differentiation of a syndrome. So how to find out the most informative symptoms combination and diagnose a syndrome with them is the most important task in TCM. Hereunto, a new question comes into being. How many symptoms are the most accurately and rationally for one syndrome?

Many approaches have been introduced to determine the number of clusters [2-4]. But there is almost no research on how to determine the cluster size (the number of

elements of cluster). Because of the TCM information's specific characteristic, e.g. fuzzy symptoms, non-full symptoms, complex syndromes, personality therapy information and complex processing of data mining [5], many common statistical methods could not be applied to TCM data mining. An unsupervised clustering algorithm is proposed based on complex system entropy at the paper [6], which provides an effective solution to chronic renal failure (CRF) problem contained in traditional Chinese medicine. Based on this algorithm, a technique of self-adaptively selecting symptoms' number of TCM syndromes is proposed, which is based on contribution degree.

In this paper, we carry out three clinical epidemiology surveys in one data set and propose an unsupervised data mining model, in which we treat mutual information (MI) as an association measure of two variables and treat contribution degree as the threshold of self-adaptively selected symptoms.

This paper is organized as follows. Section 2 is devoted to patterns discovery algorithm based on MI. An unsupervised complex system entropy cluster algorithm based on contribution degree is presented in Sec.3. In Sec.4, the algorithm is applied in the data which contains three clinical epidemiology surveys. In Sec.5, the conclusions are given.

2 Patterns (Syndromes) Discovery Algorithm Based on MI

In this part, some basic information will be introduced about entropy cluster algorithm based on mutual information (MI).

2.1 Definition of Association Based on Revised Version of MI

Mutual information between two variables is formally defined as:

$$MI(X,Y) = H(X) + H(Y) - H(X \cup Y), \tag{1}$$

where $H(X)$ denotes the Shannon entropy of variable X, $H(X \cup Y)$ represents the joint entropy between variables X and Y. Formally, suppose that X and Y are both categorical variables, $H(X)$ and $H(X \cup Y)$ are denoted as:

$$H(X) = -\sum_{i=1}^{m} \frac{n_i}{N} \ln \frac{n_i}{N}, \tag{2}$$

where n_i denotes the number of occurrence of the ith category of X with m categories, N is the total number of sample X.

$$H(X \cup Y) = -\sum_{i=1}^{m} \sum_{j=1}^{l} \frac{n_{ij}}{N} \ln \frac{n_{ij}}{N}, \tag{3}$$

where n_{ij} represents the number of simultaneous occurrence of the ith category of X with m categories and the jth counterpart of Y with l categories.

Mutual information is universally used to measure the similarity between two variables' distributions and is taken here as an association measure of two variables. MI-based association between two variables is symmetric in information theory, but the relation between two symptoms is usually asymmetric. This problem was solve by correlation coefficient, which was proposed in the paper [7] and denoted as:

$$\mu(X,Y) = \frac{MI(X,Y)}{H(Y)}. \tag{4}$$

By this definition, the association between two variables is asymmetric because two variables' Shannon entropies ($H(X)$ and $H(Y)$) are usually different. Furthermore, by information theory, the form of MI can be recast as:

$$MI(X,Y) = H(X) - H(X \mid Y). \tag{5}$$

That is to say, $MI(X,Y)$ represents the information content with regard to knowing X under the condition of knowing Y. Therefore, associations of two mostly close symptoms and completely opposite counterparts are both very large, making the association defined by MI composed of both positive and negative. Thus, we'll come up against some problem in practice. For example, the correlation coefficient of two correlative symptoms (e.g. Spiritlessness and Hypodynamia) is large, but the correlation coefficient of two opposite symptoms (e.g. Floating pulse and Sunken pulse) is even larger than it. We present a revised version of MI to identify both positive and negative associations. The frequency that X and Y are both of nonzero categories is denoted as $Pofr(X,Y)$, it is this positive frequency of X and Y that separates positive and negative associations. We revised the form of correlation coefficient as:

$$\mu(X,Y) = \begin{cases} \dfrac{H(X)+H(Y)-H(X \cup Y)}{H(Y)} & Pofr(X,Y) \geq \theta \\[3mm] \dfrac{H(X)+H(Y)-b*H(X \cup Y)}{H(Y)} & Pofr(X,Y) < \theta, \end{cases} \tag{6}$$

where b is a real number and is greater than 1, it can be seen as penalty coefficient, θ is pre-assigned positive quantity, we call it threshold in this paper. Proper setting of the two parameters will not only distinguish the positively associated symptoms from the negatively associated counterparts, but also can avoid disturbances by the wrong data.

2.2 Pattern Discovery Algorithm

2.2.1 "Relative" Set
For a specific variable X, a set, which is collected by means of gathering $Z(1 \leq Z \leq N-1)$ (where N is the total number of the variables) variables whose associations with X are larger than others with regard to X, is attached to it and is denoted as $R(X)$. Each variable in the set can be regarded as a "Relative" of X while other variables that do not belong to the set are consider as irrelative to X. So we name $R(X)$ "Relative" set of X.

2.2.2   The Pattern Discovery Algorithm Based on "Relative" Set

A pair of variables (X and Y) is defined to be significantly associated with each other if and only if X belongs to the "Relative" set of Y ($X \in R(Y)$) and *vice versa* ($Y \in R(X)$). It is available to extend this definition to a set with multiple variables. If and only if each pair of these variables is significantly associated, then we can call that the set is significant associated. In order to find the patterns hidden in the data, an unsupervised algorithm is proposed. The specific steps are as follow:

Step 1. Based on the $N \times Z$ matrix, which is composed by the N variables' "Relative" set, we can find all the significantly associated pairs (a $M_2 \times 2$ matrix), where M_2 represents the number of significantly associated pairs.

Step 2. Based on the $M_2 \times 2$ matrix, significantly associated set concluding 3 variables can be obtained, denoted by a $M_3 \times 3$ matrix, where M_3 represents the number of sets containing 3 significantly associated variables. Similarly, if there exist significantly associated multiple variables, the corresponding result is denoted by $M_m \times m$, where M_m represents the number of sets containing m significantly associated variables and m stands for the number of variables. Since Z is bounded, the algorithm is bound to converge.

Step 3. Finding the maximal m. Matrix $M_m \times m$ has M_m patterns with m variables. A set that contains $m-1$ variables is certainly not a pattern since it does not fulfill the third criterion of a pattern. All these kinds of sets are removed from the matrix $M_{m-1} \times (m-1)$, and the rest are certainly patterns with $m-1$ variables. Similarly, all the patterns can be discovered.

Compared with others, the unsupervised pattern discovery algorithm mentioned above has its own advantages and plays a major role in selecting TCM syndromes, which also leads to some satisfied results. But it is too rigid in dealing with the cluster size (i.e., the symptoms' number of syndromes) and lacks of foundation in theory. In order to solve these problems, we propose a novel algorithm to self-adaptively determine the number of symptoms, which is based on the contribution degree.

3   Patterns (Syndromes) Discovery Algorithm Based on Contribution Degree

There is a long time concerning among clinics about how to effectively and accurately determine the number of syndromes and select suitable syndromes in TCM differentiation. However, there is still no convincing research in this area. In this condition, we try to do some tentative work to fill the current blank.

3.1   Contribution Degree

How to select symptoms and estimate symptoms' contribution degree is always a basic problem in TCM data mining process. The commonly used methods in TCM are as follows: selecting symptoms by the logistic regression and estimating the

contribution degree by normalized regression coefficient [8]. In the paper, contribution degree is used to select the syndrome factors of outer pathogenies. Non-conditional Logistic multivariate regression is employed to screen the symptoms to all outer pathogenies, and the contribution and peculiarity of every symptom to the diagnosis of these syndrome factors are discussed; Estimating each symptom's contribution degree in the syndrome by chi-square and likelihood ratio [9], and the method based on the aim of minimum classification error-rate is considered as the best one; Scoring the contribution degree by "double-frequency power scissors"[10]. In this paper, the contribution degree based on variables' (symptoms') MI will be used.

Every syndrome is a set of several special symptoms, so the correlation coefficient of each symptom with the other symptoms of the syndrome is defined as the symptom's contribution for the syndrome. The contribution degree reflects the symptom's importance degree in the syndrome. Based on the Eq. (6), assuming that to random $i, j (i \neq j)$:

$$Con_i = \sum_{j=1, j \neq i}^{Z} \frac{\mu(X_i, X_j) + \mu(X_j, X_i)}{2} \quad (i \neq j), \quad (7)$$

where $\mu(X_i, X_j)$ and $\mu(X_j, X_i)$ are the ith variable's correlation coefficient with the other variables of "Relative" set, Z is the total symptoms of "Relative" set.

However, the contribution degree, defined by Eq.(7), can't be used to compare among different "Relative" sets. Thus, we adopt standardization steps and gain relative contribution degree to denote the symptoms' contribution. Specifically:

$$Con_i^* = \frac{Con_i}{\sum_{i=1}^{Z} Con_i}. \quad (8)$$

The Con_i^* is defined as the ith symptom's contribution degree for the "Relative" set. The "Relative" set is composed by several symptoms which are significantly associated with each other, but not all of these symptoms can be composed by the optimal combination of the syndromes. In order to select the optimal syndrome from the "Relative" set, the symptoms are aligned in descending order according to contribution degree. If the sum of first C symptoms' contribution degree ($\sum_{i=1}^{C} con_i^*$) reaches the threshold, the C symptoms can explain the optimal combination of the syndrome.

3.2   The Algorithm Based on Contribution Degree

Combined with Eq. (8) and the unsupervised algorithm of 2nd chapter, the self-adaptive unsupervised algorithm based on contribution degree is described in the following:

Step 1. Initialize parameters α and β, where α controls the convergence process of the self-adaptive algorithm, β controls the circle number of the algorithm and we set $\beta = 2$ here. As the step 1 of the algorithm of 2nd chapter, based on the $N \times Z$ matrix, which is all of the N variables' "Relative" set, we can find all the significantly associated pairs (a $M_\beta \times \beta (\beta = 2)$ matrix), where M_β represents the number

of significantly associated pairs. Simultaneously, we could calculate significantly associated pairs' relative contribution degree $C_i = Con_{i1}{}^* + Con_{i2}{}^* (i = 1, 2, \cdots M_2)$.

Step 2. If $C_i \geq \alpha$ jump to step 4. Otherwise, jump to step 3.

Step 3. Use the $M_\beta{}^* \times \beta$ matrix which is obtained in step 2 to find all the significantly associated $\beta + 1$ variables, denoted by $M_{\beta+1} \times (\beta + 1)$ matrix. And calculate significantly associated $\beta + 1$ variables' relative contribution degree $C_i = \sum_{j=1}^{\beta+1} Con_{ij}{}^* (i = 1, 2, \cdots M_{\beta+1})$. If $C_i \geq \alpha$ jump to step 4, otherwise let $\beta = \beta + 1$ and repeat step 3 until the limit of β.

Step 4. Save all the significantly associated variables which is satisfy $C_i \geq \alpha$.

The flow chart of the algorithm is shown in Fig.1.
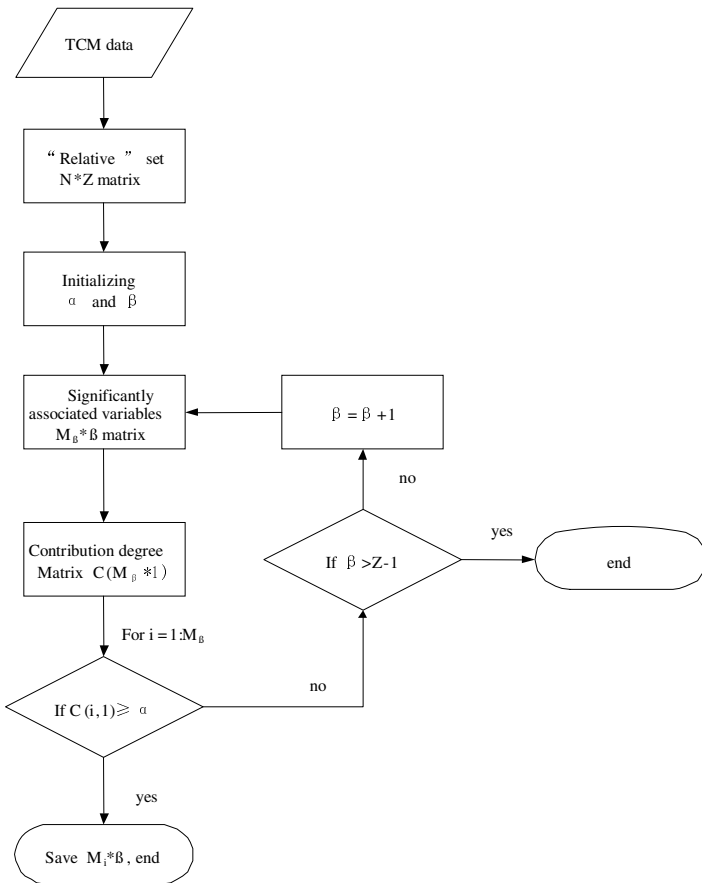
Fig. 1. The flow chart of the algorithm

The convergence of the algorithm mentioned above is decided by whether the significantly associated variables' relative contribution degree in "relative" set reaches the threshold α. In differentiation of syndromes if α reaches 0.8, we can say that the importance of symptoms of significantly associated variables in the symptoms of "relative" set reaches 80 percents. So we can believe in statistics that the symptoms of significantly associated variables can be regarded as the main components of syndrome. For this reason, we can also answer the question proposed in the beginning.

Because of the TCM system's complexity and the differences between syndromes, it is difficult to select the threshold α. According to many experiments and the TCM knowledge, $\alpha = 0.8$ is optimal for this work. Of course, we'll do some advanced research for the future.

4   Application in TCM

We carry out three clinical epidemiology surveys about depression, chronic renal failure (CRF) and chronic hepatitis b, and obtain 1787 cases (604 cases of depression, 582 cases of chronic hepatitis b and 601 cases of CRF), which measure up standards and have good compliance. And these cases were collected from 6 TCM hospitals of 6 different areas including Beijing, Liaoning, Shanxi, Shanxi', Shandong and Hubei from 2005.10.15 to 2006.1.15. Each report includes 76 symptoms, which grade to four variables (0,1,2,3) by severity degree of the symptom. Name of symptoms are shown in table 1.

Table. 1. Name of symptoms

No.	Name	No.	Name	No.	Name
1	Deprementia	25	Abdominal distention	50	Thin tongue
2	Sporotlessness	26	Abdominal pain	51	Tooth-marked tongue
3	Amnesia	27	Soreness and weakness of	52	Fissured tongue
4	Fear of cold		waist and knees	53	Collaterals under tongue
5	Feverishness in palms	28	Dull pain of waist	54	White tongue coating
	and soles	29	Soreness of waist	55	Yellow-white tongue
6	Spontaneous perspiration	30	Lumbago and desire to		coating
	tion		warm	56	Yellow tongue coating
7	Night sweat	31	Lumbago and desire to	57	Less tongue coating
8	Hypodynamia		press	58	Thin tongue coating
9	Sallow complexion	32	Anorexia	59	Thick tongue coating
10	Ochriasis	33	Dry mouth	60	Moist tongue coating
11	White complexion	34	Thirst without desire to	61	Smooth tongue coating
12	Red complexion		drink	62	Dry tongue coating
13	Black complexion	35	Thirst and desire to drink	63	Greasy tongue coating
14	Dim complexion	36	Loose stool	64	Exfoliative tongue
15	Dry eye	37	Constipation		coating
16	Headache	38	Deep-colored urine	65	Ecchymosis on tongue
17	Bombus	39	Drowsiness	66	Floating pulse
18	Cardiopalmus	40	Dreaminess	67	Sunken pulse
19	Chest distress	41	Frequent micturition	68	Slow pulse
20	Hypochondriac	42	Pale tongue	69	Rapid pulse
	distension	43	Light red tongue	70	Thready pulse
21	Hypochondriac	44	Red tongue	71	Slippery pulse
	distending pain	45	Lavender tongue	72	Uneven pulse
22	Hypochondriac dull	46	Purple tongue	73	Powerful pulse
	pain	47	Old tongue	74	Asthenia pulse
23	Gastric cavity	48	Tender tongue	75	Taut pulse
24	Nausea	49	Puffy tongue	76	Tense pulse

The result of the self-adaptive algorithm is shown in table 2.

Table. 2. The result of the self-adaptive algorithm

significantly associated variables	Contribution degree
1. Feverishness in palms and soles, Spontaneous perspiration, Night sweat	0.8272
2. Ochriasis, Dim complexion, Deep-colored urine,	0.8307
3. Dry mouth, Thirst and desire to drink, Frequent micturition	0.9130
4. Sunken pulse, Asthenia pulse, Taut pulse	0.8222
5. Hypochondriac distension, Gastric cavity, Abdominal distention	0.8173
6. Soreness and weakness of waist and knees, Soreness of waist, Lumbago and desire to warm, Lumbago and desire to press	0.9087
7. Gastric cavity, Nausea, Abdominal distention, Anorexia	0.8000
8. Dull pain of waist, Soreness of waist, Lumbago and desire to warm, Lumbago and desire to press	0.8616
9. Deprementia, Amnesia, Drowsiness, Dreaminess	0.8273

In these 9 patterns, we can make sure that 6 patterns can be diagnosed as syndrome in clinic. And the patterns and the syndromes diagnosed by TCM physicians are shown in table 3. The other 3 patterns can't be diagnosed by now.

Table. 3. Patterns and syndromes

Patterns	syndrome	Disease location
1. Feverishness in palms and soles, Spontaneous perspiration, Night sweat	deficiency of Qi and Yin	unsure
2. Deprementia, Amnesia, Drowsiness, Dreaminess	deficiency of Qi	heart
3. Soreness and weakness of waist and knees, Soreness of waist, Lumbago and desire to warm, Lumbago and desire to press	deficiency of Yang	kidney
4. Gastric cavity, Nausea, Abdominal distention, Anorexia	Qi stagnation or deficiency of Qi	spleen and stomach
5. Hypochondriac distension, Gastric cavity, Abdominal distention	Qi stagnation	liver, spleen and stomach
6. Dry mouth, Thirst and desire to drink, Frequent micturition	deficiency of Ying	kidney

5   Conclusion

In this paper, we carry out three clinical epidemiology surveys about depression, chronic renal failure (CRF) and chronic hepatitis b, and obtained 1787 cases, each of which has measurements for 76 symptoms. By studying the symptoms' contribution degree in syndrome and the optimal symptoms' number of syndrome, an unsupervised complex system entropy cluster algorithm based on contribution degree is proposed to allocate significantly associated symptoms into several patterns and self-adaptively select the number of symptoms. The algorithm successfully discovered 9 patterns, whose contribution degree reach 0.8. And 6 of them can be diagnosed as syndrome, which is completely in accordance with the corresponding results diagnosed by TCM physicians. The algorithm is viable in the TCM practice.

Acknowledgments

References

1. Wang, Y.Y., Zhang, Q.M., Zhang, Z.B.: The induction of syndrome factor and affecting target. Journal of Shandong University of Traditional Chinese Medicine 30(1), 6–7 (2006)
2. Xiao, L., Wu, Q.: A cluster number specification-free algorithm in networks intrusion detection. In: 7th World Congress on Intelligent Control and Automation, pp. 6272–6276 (2008)
3. Borgelt, C., Girimonte, D., Acciani, G.: Learning vector quantization: cluster size and cluster number. Circuits and systems 5, 808–811 (2004)
4. Xu, L.: How many clusters? A ying-yang machine based theory for a classical open problem in pattern recognition. Neural networks 3, 1546–1551 (1996)
5. Jiang, Z.S., Ni, Q., Lin, L.: Data mining's application in the research of TCM. Journal of Chinese Modem Medicine 3(11), 1240–1242 (2006)
6. Chen, J., Xi, G., Chen, J., Zhen, Y., Xing, Y., Wang, J., Wang, W.: An unsupervised pattern (Syndrome in traditional Chinese medicine) discovery algorithm based on association delineated by revised mutual information in chronic renal failure data. Journal of biological systems 15(4), 435–451 (2007)
7. Xi, G.: Entropy method of complex system partition. ACTA AUTOMATICA SINICA 13(3), 216–220 (1987)
8. Zhang, Q.M., Wang, Y.Y., Zhang, Z.B., Song, G.L.: Selection on the syndrome factors of outer pathogenies. Journal of shandong university of TCM 29(5), 339–341 (2005)
9. Zhang, Y.L., Liu, Q., Gao, Y., Ji, L.: Contribution level and feature selection for discriminating Zheng. Journal of Liaoning TCM 35(3), 354–355 (2008)
10. Huang, B.Q., Zhu, Z.H., Zhu, W.F.: The importance of common manifestation of the pulse in differentiating syndromes. Chinese journal of the practical Chinese with modem medicine 21(13), 1164–1166 (2008)

Normalizing Human Ear in Proportion to Size and Rotation

Ali Pour Yazdanpanah[1] and Karim Faez[2]

[1] Member of Young Researchers Club, Islamic Azad University of Najaf Abad, Isfahan, Iran
[2] Department of Electrical Engineering, Amirkabir University of Technology, Tehran, Iran
ali.p@iaun.ac.ir, kfaez@aut.ac.ir

Abstract. There are always two main problems in identification of human beings through their ear images: 1- If distances of the individual from camera changes, the sizes of ears in the saved images are varied in proportion to this distance. 2- If head of people in taken images is tilted upwards or downwards, this causes ear images of these people rotate in proportion to saved ear images in database. In both of these cases, all identification systems do not work properly. In this article, we proposed a new method for normalizing human ear images by detecting the rotation and scaling variation, and normalizing the ear images accordingly. Our proposed method works well on all ear databases and all ear images (either left or right) which have been taken from front side of the ears. Our method provides high performance to the biometric identification systems to identify human being, even when the images of human ears are taken from long distance with small scale.

Keywords: Ear image normalization, Scale invariance, Rotation invariance.

1 Introduction

With the rapid technological advances in electronics and Internet commerce as well as with the increased emphasis on security, there will be a growing need to authenticate individuals. Biometric recognition is emerging as a reliable method which can overcome some of the limitations of the traditional automatic personal identification technologies. Ears have long been considered as a potential means of personal identification, yet it is only in the last 10 years or so that machine vision experts started to tackle the idea of using ears as a biometric. French criminologist Alphonse Bertillon was the first to recognize the biometric potential of human ears [12]. Empirical evidence supporting the ear's uniqueness was later provided in studies by Iannarelli [11]. The structure of the ear is rich and stable, and does not change radically over time; the ear is less variability with expressions, and has a more uniform distribution of color than faces. These unique characters of the ear make it possible to make up the drawbacks of other biometrics and to enrich the biometrics identification technology [5].

    Ear Recognition is a new type of biometric recognition method. Meanwhile, ear recognition is also one of the most challenging subjects in the area of computer vision

and pattern recognition. Due to the unique location and structural features of human ears, the research in the above field has prospective potential in public safety and information security applications, which has drawn an increasing attention [1][2]. There are varieties of systems for human identification from two-dimensional images of ears. Burge *et al.* [3] proposed an approach based on Vornoi diagrams. The system proposed by Moreno *et al.* [4] is based on neural network. Hurley *et al.* [5] developed a system based on a force field feature extraction. Zhang *et al.* [6] developed a system combining independent component analysis (ICA) with radial basis function (RBF) network. The system proposed by Choras *et al.* [7] is based on geometric feature extraction. However, there are common problems in all of these systems: 1- If in the input images of these systems, the head of people is tilted upwards or downwards. 2- If in the input images of these systems, the distance of the individual from camera changes, the sizes of ears in the saved images are varied in proportion to this distance. In both of these cases, these systems wouldn't work properly and couldn't recognize the individual from their input ear images. So far no algorithm was suggested for normalizing the ear images to compensate these changes and solving these problems. In this paper we proposed a method for normalizing ear images which can detect any rotation and scaling changes in these images and normalize the images using the proposed algorithm. In our method, firstly all edges of ear images are detected using the Canny edge detection algorithm with different thresholding [8]. The fundamental problem in all these binary images resulted from Canny algorithm [8] is that we can't detect the part of outer edges of ear which attach to head thorough Canny algorithm [8]. In [9], a manual approach has been proposed which is resistant to the scale and rotation changes of input ear images. However, it does not work with any existing database, because in this method for each ear image, all the outer edges of ear image must be present. According to the authors [9], one special database has been collected to be used only by their algorithm. But, in our proposed method for normalizing ear images, without any need of a special kind of database and without any need to have all outer edges of an ear image, specially the part of outer edges of ear which attach to the head, we can normalize ear images to standard form. Figure 1, shows an input ear image and one obtained from Canny algorithm [8].

The proposed approach is explained in section 2, noise robustness of proposed method is explained in section 3, experimental results are given in section 4, and conclusion is given in the last section.

(a)   (b)
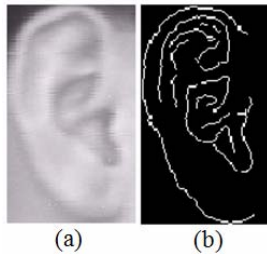
Fig. 1. (a): Input ear image. (b): Ear image obtained from Canny algorithm [8].

2   Ear Image Normalization

To test the proposed method, we have used a database of 60 individuals, selected from the USTB database [10]. To present a complete description of the proposed algorithm, we must first consider two definitions: Max-Line and Ear Main Angle

A. Max-Line

The longest line that can be drawn with both of its endpoints on the outer edge of the ear is called "Max-Line". The length of a line can be measured in terms of Euclidean distance [9]. The beginning and ending points of this line (Max-Line) are located on the outer edge of ear. It is important for us to precisely calculate length of this line. In the rest of article, we will explain the method of calculating this parameter.

B. Ear Main Angle

The angle between a horizontal line confluent with Max-Line is called "Ear Main Angle". We present Ear Main Angle as an important feature for human ear. In ear images, we can normalize rotational change with this new feature. This angle changes within a certain range for right and left ears, that's why we called it, Ear Main Angle. Figure 2, from left to right respectively shows the binary image of ear, Max-Line obtained from binary image, and Ear Main Angle.
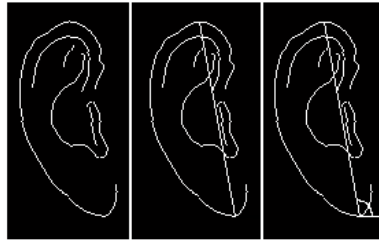
Fig. 2. From left to right respectively: the binary image of ear and Max-Line obtained from binary image and Ear Main Angle

Having known these definitions, we can now explain the normalization process of ear image which include four steps:

1- Finding Max-Line and calculating its length.
2- Calculation the value of Ear Main Angle
3- Normalizing the changes resulting from scaling changes
4- Normalizing the changes resulting from rotation changes

2.1   Finding Max-Line and Calculating Its Length

Finding Max-Line can be done firstly by finding all edges in an ear image. After finding all the edges of ear image, we must detect the outer edges of an ear in the image. For this purpose, we start to scan pixels from four sides of edge detected binary image (up to down, down to up, right to left, left to right)to be able to find the outer edges of ear. For example, from right to left: we start to scan the pixels in rows. As we reach

the pixels related to one outer edge of ear (first pixels form right to left in the row for which the value is 1 in binary image), for the first row we save the coordinate of this pixel and go to the next row. This process should be apply on all the rows. We do the same on three other direction (left to right and up to down and down to up), quite the same as the first direction. But for the two directions (up to down and down to up) we should scan through each column instead of each row. In each direction, one side of outer edge of ear can be determined. Finally, the result of each direction which includes one side of outer edge of ear image is integrated with others to form final result which includes a complete outer edge of ear image. Figure 3, shows all the proposed algorithm steps to detect the outer edge of ear image.
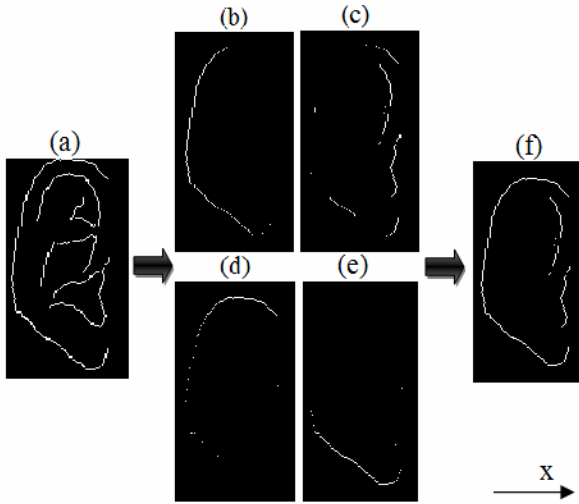
Fig. 3. a: Ear binary image, b: Left side of outer edge of ear, c: Right side of outer edge of ear, d: Upper side of outer edge of ear, e: Down side of outer edge of ear, f: Outer edge of ear

According to our algorithm, we can detect the outer edge of ear. However, in the case that part of outer edge of ear which attached to the head can't be detected through Canny algorithm as the edge, maybe some points inside the ear are detected as the outer edge of ear when scanning the pixels (for example: from right to left for right ear). But this problem does not have any effect on our proposed method in finding Max-Line, because pixels of these points do not have maximum distance with other pixels of outer edge of ear. So, none of these points are recognized as the starting or ending point of Max-Line. Finally, we measure the Euclidean distance of each pixel with all another pixels in the binary image of outer edge of ear and find the maximum value of these distances will give us the length of Max-Line and then we can locate the starting and ending points of this line.

2.2   Calculation the Value of Ear Main Angle

With finding Max-Line, we now can calculate the Ear Main Angle, which is the angle between a horizontal line confluent with Max-Line in the image. If (x_1,y_1) and (x_2,y_2) are the starting or ending point of the Max-Line, Ear Main Angle is given by the equation 1.

$$
\theta = \begin{cases}
\tan^{-1}(\dfrac{x_1 - x_2}{y_2 - y_1}) & \text{if} \quad x_1 > x_2 \quad \text{and} \quad y_1 < y_2 \\[2mm]
180 - \tan^{-1}(\dfrac{x_2 - x_1}{y_2 - y_1}) & \text{if} \quad x_1 < x_2 \quad \text{and} \quad y_1 < y_2 \\[2mm]
180 - \tan^{-1}(\dfrac{x_1 - x_2}{y_1 - y_2}) & \text{if} \quad x_1 > x_2 \quad \text{and} \quad y_1 > y_2 \\[2mm]
\tan^{-1}(\dfrac{x_2 - x_1}{y_2 - y_1}) & \text{if} \quad x_1 < x_2 \quad \text{and} \quad y_1 > y_2
\end{cases}
\tag{1}
$$

where θ is the Ear Main Angle of image.

2.3 Normalizing the Changes Resulting from Scaling Changes

After we obtained the Ear Main Angle, in order to solve the scaling problem, we rotate all 60 images of the database to set the Ear Main Angles to 90 degrees. To obtain a standard scale for ear, it is essential to find the bounding rectangle which includes only human ear in the image. After fixing the Ear Main Angle to 90°, the length of bounding rectangle is equal to the length of Max-Line. So, we intend now to find the maximum width of the bounding rectangle, which includes entire width of human ear within the image. To calculate this width, after changing the Ear Main Angle to 90°, we must obtain the opposing vertical sides (right and left sides) which are parallel to the length of this bounding rectangle(Max-Line), in such a way that the entire width of human ear locates between these two sides. The distance between these two sides gives us the width of bounding rectangle. We assume the x axe perpendicular to Max-Line (After changing Ear Main Angle to 90°, Max-Line is equal to the length of bounding rectangle). In the right side of outer edge of ear, (Fig.3c), the line parallel with the length of bounding rectangle and tangent with the pixel which has the highest value of x, gives us the right side of the bounding rectangle. In the left side of outer edge of human ear (Fig.3b), the line parallel with the length of bounding rectangle and tangent with the pixel which has the lowest value of x, give us the left side of the bounding rectangle. The distance between these two sides is the width of the bounding rectangle. We apply this process on all the 60 images of human ear database. We average all the lengths values calculated from the bounding rectangles of these 60 images, to gives us an average value of length for bounding rectangle. Also we average all the widths values which calculated from the bounding rectangles of these 60 images, to gives us an average value of width for bounding rectangle. Now, each other input image can be normalized through the proposed algorithm first by finding
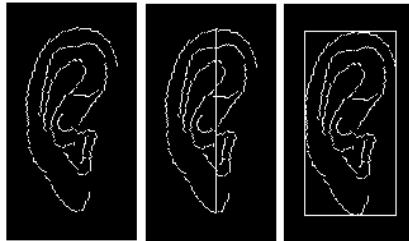
Fig. 4. Bounding rectangle obtained from ear image

the bounding rectangle of the ear and then by scaling the length and width of this bounding rectangle to standard values. Figure 4, shows the result of our proposed algorithm in finding the bounding rectangle.

2.4   Normalizing the Changes Resulting from Rotational Changes

To normalize the changes which are resulted from rotation, we first calculate the Ear Main Angle for all the 60 images. Then we average them to obtain a standard angle value for all the normalized images which can be done for all another input images through adjusting this angle to the value of standard angle. In section 2.3, for obtaining the bounding rectangle length, we rotated the Ear Main Angle to $90°$. But this couldn't be a reason that we didn't need the normalization process for the rotational changes. Because when we fix Ear Main Angle to $90°$, it results in a non-standard ear shape in ear image, so after applying normalization process without normalizing the rotational changes, if we want to detect ear by some ear detection algorithms, specially the algorithms working according to shape of the ear, like ASM (Active Shape Model), we can't detect ear shape correctly from that non-standard image. Figure 5, gives a scheme of our overall proposed method.

Ear Image

```
┌─────────────────────┐          ┌─────────────────────┐
│   Canny Edge        │          │    Finding          │
│ Detection Algorithm │          │   Bounding          │
└─────────────────────┘          │   Rectangle         │
          │                      └─────────────────────┘
          ▼                                │
┌─────────────────────┐          ┌─────────────────────────┐
│   Max-Line          │          │ Scale the Length and Width │
│   Detection         │          │ Of Bounding Rectangle   │
└─────────────────────┘          │ To 146 , 76 Pixel       │
          │                      └─────────────────────────┘
          ▼                                │
┌─────────────────────┐          ┌─────────────────────────┐
│   Calculate Ear     │          │ Rotate Image to Set Ear │
│   Main Angle        │          │ Main Angle to 102.2 °   │
└─────────────────────┘          └─────────────────────────┘
          │                                │
          ▼                                ▼
┌─────────────────────┐          Normalized Ear Image
│   Rotate Image to   │
│ Set Ear Main Angle  │
│     to 90 °         │
└─────────────────────┘
```
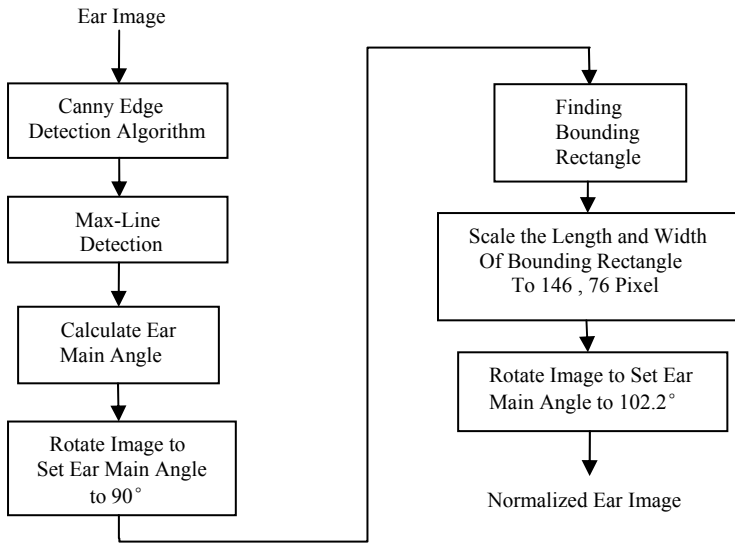
Fig. 5. A scheme of the proposed method

3   Robustness against Noise

To check the robustness of proposed method against noise, we add Gaussian noise with zero mean and different variances to input ear images and then apply overall proposed algorithm to noisy images. Gaussian noise is given by the equation 2.

$$P(z) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(z-\mu)^2}{2\sigma^2}} \qquad -\infty < z < +\infty \; , \tag{2}$$

where μ is the mean of noise and σ is the standard deviation of noise. When we add Gaussian noise which has a variance less than 0.001 to input images, the detection rate of bounding rectangle and Ear Main Angle do not change considerably with respect to the images with no noise. But when we add Gaussian noise which has a variance between 0.001 and 0.01 to input images, the detection rate of bounding rectangle and Ear Main Angle decrease at a maximum of 20%. If we use Gaussian noise which has a variance larger than 0.01, the detection rate of bounding rectangle and Ear Main Angle is not acceptable. It means that the detection rate of bounding rectangle and Ear Main Angle were lower than 40%. The detection rate of Ear Main Angle and bounding rectangle was calculated by comparing the results from applying proposed algorithm to the noisy images with the results from applying proposed algorithm to original images without noise. Table 1 shows detection rate of bounding rectangle and Ear Main Angle for different applied noise variances.

Table 1. Detection rate of bounding rectangle and Ear Main Angle in 60 images

Noise variances	Detection rate of bounding rectangle and Ear Main Angle
0.0008	100%
0.001	100%
0.005	85%
0.008	80%
0.01	80%
0.1	not acceptable
0.5	not acceptable

4 Experimental Results

This method was tested on one part of the USTB database [10] including 60 images of human ear of 60 individuals. The images in this database are taken under standard conditions. The size of all 60 images is 80*150 pixels. We have two goals in this paper, firstly, to find the amount of scale which input ear image was changed in proportion to its standard scale, secondly, to find the value of the angle which Ear Main Angle was rotated in proportion to its standard position. The range of Ear Main Angle for all 60 image changes from 92.47 to 110.87 degrees (for right ears). The standard value of Ear Main Angle calculated from the average of these angles is 102.26 degree. The range of bounding rectangles width in these 60 images changed from 62 to 87 pixels. The range of bounding rectangles lengths in these 60 images changed from 138 to 153 pixels (Table 2). The standard Length of the bounding rectangle is 146

pixels and its width is 76 pixels. In section 3, in a case we use input images without any noises, in all 60 images, Ear Main Angles and bounding rectangles were detected correctly (100%). But, for example, in a case we add Gaussian noise with variance of 0.005 to input ear images, only in 51 of 60 ear images, Ear Main Angles and bounding rectangles were detected correctly (85%). Also we showed that our proposed method can detect bounding rectangle and Ear Main Angle correctly even if we added Gaussian noise which has the variance less than 0.001 to input images. As mentioned before, normalizing ear images as a preprocessing algorithm, plays an important role for human identification system. Because, firstly, we proved that the person whether near to the camera or far from it, our proposed algorithm, by normalizing the scaling changes, could be able to change the ear image to its standard scale. So human identification systems could easily work with these standard images. Secondly, we proved that if the head of the person in taken images is tilted upwards or downwards, our proposed algorithm, by normalizing the rotational changes, could be able to rotate Ear Main Angle to its standard value. So human identification systems could easily work with these standard images.

Table 2. Range of variations for three parameters in 60 images

Parameter	Unit	Lower Bound	Upper Bound
Ear Main Angle	Degree	92.47	110.87
Width of Bounding Rectangles	Pixel	62	87
Length of Bounding Rectangles	Pixel	138	153

5   Conclusions

In this paper, we proposed a new method for normalizing images of human ears which can detect each rotation and scaling change and to normalize ear images to their standard form according to the proposed algorithm. Unlike other method, our method can work with all ear databases and all images which have been taken from the front side of ears (either right or left). Concluding, this method of normalization is scale invariant and rotation invariant.

References

1. Zhang, H., Mu, Z., Qu, W., Liu, L., Zhang, C.: A Novel Approach for Ear Recognition Based on ICA and RBF Network. In: Proceeding of ICMLC 2005 Conference, pp. 4511–4515 (2005)
2. Wang, Z., Mu, Z., Wang, X., Mi, H.: Ear Recognition Based on Invariant Moment Matching. Pattern Recognition and Artificial Intelligence 17(4), 502–505 (2004)

3. Burge, M., Burger, W.: Ear Biometrics in Computer Vision. In: Proc. of ICPR 2000, pp. 822–826 (2002)
4. Moreno, B., Sanchez, A.: On the use of outer ear images for personal identification in security applications. In: Proc. IEEE 33rd Annual Intl. Conf. on Security Technology, pp. 469–476 (1999)
5. Hurley, D.J., Nixon, M.S., Carter, J.N.: Force Field Feature Extraction for Ear Biometrics. Computer Vision and Image Understanding 98, 491–512 (2005)
6. Zhang, H., Mu, Z., Qu, W., Liu, L., Zhang, C.: A Novel Approach for Ear Recognition Based on ICA and RBF Network. In: Proc. of the Fourth International Conference on Machine Learning and Cybernetics, pp. 4511–4515 (2005)
7. Choras, M.: Ear Biometrics Based on Geometrical Feature Extraction. Electronic Letters on Computer Vision and Image Analysis 5(3), 84–95 (2005)
8. Canny, J.: A Computational Approach to Edge Detection. IEEE Transaction on Pattern Analysis and Machine intelligence 8(6), 679–698 (1986)
9. Shailaja, D., Gupta, P.H.: A Simple Geometric Approach for Ear Recognition. In: ICIT 2006, pp. 164–167 (2006)
10. Yuan, L., Mu, Z., Xu, Z.: Using Ear Biometrics for Personal Recognition. In: Li, S.Z., Sun, Z., Tan, T., Pankanti, S., Chollet, G., Zhang, D. (eds.) IWBRS 2005. LNCS, vol. 3781, pp. 221–228. Springer, Heidelberg (2005)
11. Iannarelli, A.: Ear Identification. Paramount Publishing Company, Freemont, California (1989)
12. Bertillon, A.: La photographie judiciaire, avec un appendice sur la classification et l'identification anthropometriques. Gauthier-Villars, Paris (1890)

An Ensemble Classifier Based on Kernel Method for Multi-situation DNA Microarray Data

Xuesong Wang, Yangyang Gu, Yuhu Cheng, and Ruhai Lei

School of Information and Electrical Engineering
China University of Mining and Technology, Xuzhou, Jiangsu 221116, P.R. China
`wangxuesongcumt@163.com`

Abstract. In order to deal with the interaction between genes effectively, a kernel technology was adopted into a subspace method in our study. A linear subspace classifier was generalized to a nonlinear kernel subspace classifier by using a kernel principle component analysis method to constitute nonlinear feature subspaces. Because DNA microarray data have characteristics of high dimension, few samples and strong nonlinearity, three types of classifiers based on kernel machine learning method were designed, i.e., support vector machine (SVM), kernel subspace classifier (KSUB-C) and kernel partial least-squares discriminant analysis (KPLS-DA). But the performances of these classifiers lie on the optimum setting of kernel functions and parameters. Therefore, to avoid the difficulty of selecting optimal kernel functions and parameters and to further improve the accuracy and generalization property of the classifiers, an ensemble classifier based on kernel method for multi-situation DNA microarray data was proposed by adopting the idea of ensemble learning. The ensemble classifier combines the classification results of the SVM, KSUB-C and KPLS-DA classifiers. Experimental results involving three public DNA microarray datasets indicate that the proposed ensemble classifier has high classification accuracy and perfect generalization property.

Keywords: Multi-situation; DNA microarray data; Kernel method; Ensemble learning; Classifier.

1 Introduction

As an advanced experimental technique, DNA microarray experiment has a profound influence on the development of bio-informatics. It is a high-throughout technique, i.e., it allows the recording of expression levels of thousands of genes simultaneously, which provides experts with access to comprehensive information about diseases. However, high dimension, few samples and strong nonlinearity of the DNA microarray make data mining from gene expression data a very challenging investigation [1].

In recent years, the design of appropriate machine learning methods suitable for DNA microarray analysis has become a hot topic. Many classification methods have been proposed in the literature for DNA microarray analysis; for example, decision tree [2], Bayes classifier [3], partial least-squares discriminant analysis (PLS-DA) [4],

artificial neural network [5], and support vector machine (SVM) [6]. SVM belongs to nonparametric method. It can effectively solve the problems of over-fitting, curse of dimensionality and poor generalization property of neural networks. But the classification accuracy of SVM depends on an optimal selection of a kernel function and parameters. As a classical recognition method, subspace classifier (SUB-C) has already been applied to pattern recognition field in the end of the 1980s. The SUB-C generally extracts useful information from original feature vectors to realize pattern classification. Like the PLS-DA, the SUB-C is a linear analysis method and it is difficult to deal with complex nonlinear cases. Therefore, we tried to extend a linear SUB-C to a nonlinear kernel SUB-C (KSUB-C) by drawing an idea from the construction of kernel PLS-DA (KPLS-DA) [7], i.e., by introducing kernel technology into SUB-C and using a kernel principle component analysis (KPCA) method to construct a nonlinear feature subspace.

SVM, KPLS-DA and KSUB-C are all kernel-based machine learning methods. They can implement a linear classification in a high-dimension feature space by adopting an appropriate kernel function to fulfill a nonlinear transformation without any increase of computational complexity. Therefore, they can avoid the curse of dimensionality. However, they are inevitably confronted with an optimal selection of a kernel function and parameters, which will affect the classification accuracy and generalization property. It is generally easier to obtain much better performances through ensemble learning than from a single learning machine, since ensemble learning results are not sensitive to the setting of the kernel function and parameters. In order to improve the classification accuracy and to avoid the optimal selection of kernel function and parameters of the kernel-based classifiers, an ensemble classifier based on kernel method for multi-situation DNA microarray data was proposed by integrating the classification results of the SVM, KSUB-C and KPLS-DA base classifiers in this study.

2   An Ensemble Classifier for DNA Microarray

There are two main aspects in the field of DNA microarray analysis. The first is how to remove those genes whose expression levels are indistinctive by gene selection operation. The other aspect is how to design an appropriate classifier based on the preprocessed data to diagnose and predict unknown diseases. The architecture of an ensemble classifier for DNA microarray is described in Fig. 1, where Kruskal-Wallis ranksum test was used to select P^* genes whose expression levels are distinctive.
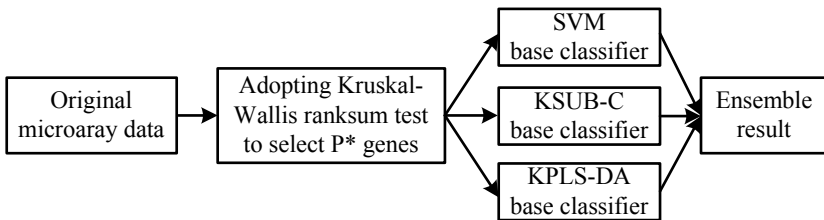
Fig. 1. Architecture of an ensemble classifier for DNA microarray data

Based on the selected gene data, three types of base classifiers, i.e., SVM, KSUB-C and KPLS-DA were designed. At last, the final classify result can be obtained by integrating the results of different classifiers.

2.1 Gene Selection

DNA microarray experiments allow the simultaneous recording of expression levels of thousands of genes. But usually only a small set of genes are suitable in disease identification and too large a dataset will increase computational complexity and reduce classification speed. Therefore, it is necessary to select a subset of useful genes before disease classification. There are some gene selection methods suitable for binary problems, such as Golub method, t test and Wilcoxon test. In order to realize pattern recognition of multi-class diseases, we used the Kruskal-Wallis test to select those genes whose expression levels fluctuate markedly in our study.

Considering a $N \times p$ data matrix $V = (x_{ij})_{N \times p}$, where N is the number of samples, p is the number of original genes, x_{ij} is the expression level of the jth gene in the ith sample. Suppose there are k independent classes of samples in V_c, i.e., $V_c \sim F(x - \theta_c)$, $c = 1, 2, \cdots, k$. These distributions F are continuous functions with the same form but different location parameters θ_c. Suppose $x_1^c, \cdots, x_{n_c}^c$ are samples from V_c, then N can be represented as $N = \sum_{c=1}^{k} n_c$ and the rank of x_q^c in V_c is R_{cq}. If the sum and average rank of V_c are denoted as, respectively,

$$R_c = \sum_{q=1}^{n_c} R_{cq} \text{ and } \overline{R}_c = \frac{R_c}{n_c}, \text{ then the average rank of } V_c \text{ is } \overline{R} = \sum_{c=1}^{k} \frac{R_c}{N} = \frac{N+1}{2}.$$

The Kruskal-Wallis test uses $H = \frac{12}{N(N+1)} \sum_{c=1}^{k} n_c (\overline{R}_c - \overline{R})^2$ to illustrate the expression diversity of the same gene among different classes. We calculated the H value of each gene and selected P^* genes with relatively large H values for later operations.

2.2 Support Vector Machine

The support vector machine is a machine learning method that can effectively deal with small-scale sample problems as well as with samples of large dimensions. Therefore the SVM is suitable for analyzing gene expression data. The SVM uses a hypothetical space of linear functions in a kernel-induced feature space that respects the insights provided by a generalization theory and exploits an optimization theory.

Suppose that a binary dataset is $\{(x_i, y_i) | i = 1, 2, \cdots, N\}$, where $x_i \in R^p$ is an input vector and $y_i \in \{-1, +1\}$ contains the corresponding labels. In the end, the training of the SVM is transformed into the following optimization problem:

$$\min \frac{1}{2}\sum_{i=1}^{N}\sum_{i'=1}^{N} y_i y_{i'}\alpha_i\alpha_{i'}K(x_i,x_{i'})-\sum_{i'=1}^{N}\alpha_{i'}$$
$$\text{s.t. } \sum_{i=1}^{N} y_i\alpha_i = 0, \quad 0 \leq \alpha_i \leq C \tag{1}$$

where α_i is a Lagrangian multiplier and C a regularization factor which determines the tradeoff between the empirical error and the complexity of the model. $K(x_t,x_{i'}) = \varphi(x_t)\cdot\varphi(x_{i'})$ is a kernel function that should satisfy the Mercer theorem. There are many types of kernel functions, such as the polynomial, Gaussian and sigmoid kernels which can be used [8]. Compared with other kernel functions, the Gaussian kernel has the advantages of a simple model, light computational burden, high computational efficiency and ease of realization. Therefore, we have opted for the Gaussian kernel in our study.

The optimal values of $\alpha^* = (\alpha_1^*,\cdots,\alpha_N^*)^{\mathrm{T}}$ and b can be obtained by solving Eq. (1) and then the classification model of the SVM can be described as follows.

$$f(x) = \mathrm{sgn}\left[\sum_{i=1}^{N}\alpha_i^* y_i K(x,x_i)+b^*\right]. \tag{2}$$

The SVM was initially proposed to deal with binary classification problems. This paper solved the multi-classification problems by training and combining a number of binary SVMs.

2.3 Kernel Subspace Classifier

Traditional SUB-C is a linear classifier, therefore, it cannot effectively extract nonlinear characteristics of samples. In machine learning, the kernel trick is a method for using a linear classifier algorithm to solve a non-linear problem by mapping the original nonlinear observations into a higher-dimensional feature space, where the linear classifier is subsequently used. It makes a linear classification in the new space equivalent to nonlinear classification in the original space. Therefore, we applied the kernel trick to SUB-C method and designed a nonlinear kernel SUB-C by using KPCA to construct a nonlinear feature subspace.

The KSUB-C method can be described as follows:

$$\overline{C} = \frac{1}{n_c}\sum_{q=1}^{n_c}\varphi'(x_q)\varphi'(x_q)^{\mathrm{T}}. \tag{3}$$

The eigenvector W is the nonlinear principal vector of the original samples which satisfies Eq. (4).

$$\lambda W = \overline{C}W. \tag{4}$$

where $\overline{CW} = \dfrac{1}{n_c} \sum\limits_{q=1}^{n_c} (\phi'(x_q) \cdot W) \phi'(x_q)$. The eigenvector W must lie in a space spanned by $\phi'(x_1), \cdots, \phi'(x_{n_c})$, i.e., W is a linear combination of $\varphi'(x_1), \cdots, \varphi'(x_{n_c})$. Accordingly, we can obtain the following equation.

$$\lambda \varphi'(x_q) \cdot W = \varphi'(x_q) \cdot \overline{CW} . \tag{5}$$

In addition, there is a coefficient α'_q which satisfies $W = \sum\limits_{q=1}^{n_c} \alpha'_q \varphi'(x_q)$, therefore, we can obtain

$$\lambda \sum\limits_{q=1}^{n_c} \alpha'_q (\varphi'(x_l) \cdot \varphi'(x_q)) = \dfrac{1}{n_c} \sum\limits_{q=1}^{n_c} \alpha'_q \left(\varphi'(x_l) \cdot \sum\limits_{g=1}^{n_c} \varphi'(x_g) \right) (\varphi'(x_g) \cdot \varphi'(x_q)) . \tag{6}$$

where $l = 1, 2, \cdots, n_c$. Eq. (6) can be converted into Eq. (7) by defining a $n_c \times n_c$ matrix $K'_{lq} = \varphi'(x_l) \cdot \varphi'(x_q)$.

$$n_c \lambda K' \alpha' = (K')^2 \alpha' . \tag{7}$$

where $\alpha' = (\alpha'_1, \alpha'_2, \cdots, \alpha'_{n_c})^T$. Because K' is a symmetry matrix and one of its eigenvector is the based variable of the whole spanned space, Eq. (7) can be rewritten as $n_c \lambda' \alpha' = K' \alpha'$. The former nb principal components $U^c = U(u_1^c, u_2^c, \cdots, u_{nb}^c)$ of K' is also the principal vector of the space $\varphi'(x)$. Consequently the projection matrix $P_c = U^c (U^c)^T$ of each subspace can be obtained.

For the unlabeled sample x, its projection residue distance in each subspace U^c is

$$D(x, U^c) = \left\| \varphi'(x) - M_c \right\|^2 - \left\| P_c (\varphi'(x) - M_c) \right\|^2 . \tag{8}$$

where, M_c is a projection of the mean of samples of class c into the cth subspace. Larger $D(x, U^c)$ indicates better separability between x and samples of class c. Therefore x can be classified into the class corresponding to the smallest value of $D(x, U^c)$.

$$\text{class}(x) = \arg \min_c D(x, U^c) . \tag{9}$$

2.4 Kernel Partial Least-Squares Discriminant Analysis

Partial least-squares discriminant analysis is a robust discriminant analysis and statistical method, which is particularly suitable for small-scale sample problems with multiple correlated variables. The PLS-DA method was initially used for the analysis of chemical spectroscopy, and later was applied to engineering as well as economic management fields [9].

The original PLS-DA method transforms a n dimensional observation vector into a group of new features through linear transformation. Every new feature is a linear combination of original features. But the new features exhibit a poor intelligibility. In addition, the large dimensions of the feature space will lead to low accuracy. Therefore we introduced a kernel function into the linear PLS-DA to extend it to a nonlinear kernel PLS-DA. We used the kernel PLS to extract the main potential factors from the original variables, and then we can obtain the classification results by discriminant analysis on the new features.

The basic operation steps of KPLS-DA are described as follows. Firstly, we extract the first potential factors t_1 (which is a linear combinations of $x_1, x_2, \cdots, x_p$) and u_1 from independent and dependent variables respectively. It is necessary to ensure the degree of correlation between t_1 and u_1 is maximum. Then we should construct a regression model between y and t_1. If this regression model reaches a satisfactory accuracy, the algorithm can be finished. Otherwise, the next run of potential factor extraction begins using the residual information of x and y explained by t_1 and u_1. The above operation was repeated until the required accuracy is met. Suppose that the principal component matrix is $w = (t_1, t_2, \cdots, t_T)$, where T is the number of principal components. The response matrix of new samples is $y_t = x_t * w$ and the response result of sample x_i is $y_i = (y_{v_1}, y_{v_2}, \cdots, y_{v_k})$. Therefore, the classification result of sample x_i is $\text{class}(x_i) = \arg\max_c (y_{v_c})$.

3   Case Study

Three typical DNA microarray datasets were used to verify the ensemble classifier, namely Leukemia, Mixed-Lineage Leukemia (MLL), and Small Round Blue Cell Tumor (SRBCT) [10]. An overview of the characteristics of all the datasets can be found in table 1. They are binary, three-class and four-class problems respectively.

Because the number of DNA microarray samples is very small, we used the leave-one-out cross validate (LOOCV) method to assess the generalization property of classifiers. Suppose there are total N samples and the number of the misclassified samples is N^*, the correct rate of the classifier is $r = 1 - \dfrac{N^*}{N}$. The larger r, the stronger the generalization property.

Table 1. DNA microarray dataset

Dataset	Number of samples	Number of genes	Sorts of cancers and the distributions
Leukemia	72	7129	ALL(47), AML(25)
MLL	72	12533	ALL(24), MLL(20), AML(28)
SRBCT	88	2308	63 labeled samples: EWS(23), BL(8), NB(12), RMS(20); 25 unlabeled samples

Considering the homology of the Leukemia and MLL datasets, we compared these two experiments and results. For these two datasets, we selected 10, 50, 100 or 200 genes using Kruskal-Wallis test to construct SVM classifiers respectively. The classification results show that if the number of the selected genes is larger than 50, all samples in the Leukemia dataset will be ascribed into one class and the samples in the MLL dataset could not be classified correctly. Therefore, we should reduce the number of selected genes. Experimental results show that the selection of 2 leukemia genes and 5 mixed-lineage leukemia genes can achieve the best classification results, with the accuracy rates of 0.9583 and 0.9444 respectively. Results on these two DNA microarray datasets using the SVM classifiers are shown in tables 2 and 3.

When we designed a KSUB-C base classifier for the Leukemia and MLL datasets, the value of P^* was selected from the range of [50, 200] with an interval of 10 and the dimension L was set as 10, 15 or 30 using the KPCA method respectively. The experimental results show that the stability and accuracy of the classification results are satisfactory.

The experimental results show that the effect of the number of potential factors on KPLS-DA is relatively prominent, which can be explained by the classification results on the MLL dataset. For the Leukemia dataset, the correct rate is 0.9722 if 2 potential factors are extracted. Many experimental results on the MLL dataset show that results corresponding to 2 or 3 potential factors are acceptable and results corresponding to only one potential factor cannot identify the MLL at all, i.e., the KPLS-DA will ascribe all of the samples of ALL and MLL to one class with a significant distinction between AML. Therefore, we used a KPCA method to reduce the dimension as 500, and then extracted potential factors. Results show that this operation can improve the correct rate on the MLL dataset by about 10 percentages. Because the KPLS-DA with only one potential factor can distinguish samples of the MLL and ALL classes from that of the AML class, we can use the KPLS-DA to pre-determine whether a sample is AML and to further determine whether it is MLL or ALL, which plays a role as auxiliary therapy.

The parameters of SVM, KSUB-C and KPLS-DA including the number of selected genes, the number of principle components of KPCA and the number of potential factors were updated to train these base classifiers. For each type of base classifier, 5 classifiers were trained and the final result was integrated by the total 15 base classifiers through the majority voting method. Tables 2 and 3 present the highest correct rates, the average correct rate of each type of base classifier and the final correct rate after integration. It can be seen that the correct rate of the ensemble classifier is higher than the average correct rate of each type of base classifier. The classification accuracy has been improved.

Table 2. Results on the Leukemia dataset

Base classifier	The optimal operation (the highest correct rate)	Average correct rate	Final correct rate
SVM	2 genes were selected (0.9583)	0.9333	
KSUB-C	100 genes were selected and 10 KPCA principle components were extracted (0.9722)	0.9531	0.9861
KPLS-DA	2 potential factors were extracted (0.9722)	0.9070	

Table 3. Results on the MLL dataset

Base classifier	The optimal operation (the highest correct rate)	Average correct rate	Final correct rate
SVM	5 genes were selected (0.9444)	0.9110	
KSUB-C	120 genes were selected and 15 KPCA principle components were extracted (0.9167)	0.9056	0.9206
KPLS-DA	3 potential factors were extracted (0.7778)	0.7292	

The 63 SRBCT samples that have already been labeled were served as training samples to train SVM, KSUB-C and KPLS-DA base classifiers, and then the left 25 unlabeled samples were served as testing samples.

First of all, genes with the number of 10, 20, 50 or 100 were selected to construct a SVM classifier. Many times of experiments show that if the number of genes exceeds 20, the correct rate of the SVM is less than 50%. Therefore, we narrowed the scope of gene selection within the range of [1, 20]. Results show that 4 genes can guarantee the best classification accuracy.

When we designed a KSUB-C base classifier for the SRBCT dataset, the number of the selected genes was set as 50, 100, 150 or 200 and the reduced dimension was 10, 20 or 30. Many results show that the classification results are relatively stable.

The performances of KPLS-DA on each type of SRBCT samples are quite different, i.e., one EWS sample was misclassified as NB and 7 RMS samples were misclassified as NB, while the correct rate of all BL samples was 100%. Although the KPLS-DA has a poor performance on the SRBCT dataset, but it can provide important reference value for the classification of RMS samples. For example, if a sample is classified as a RMS case by both the KPLS-DA and other classifiers, then the sample is likely to belong to RMS case. It should be noted that because the number of training samples is relatively small, the above conclusion still needs to be verified by a large number of medical diagnosis.

Followed the experimental steps on the Leukemia and MLL datasets, we carried out many experiments on the SRBCT dataset. For each type of base classifier, 5 classifiers were trained and the final result was integrated by the total 15 base classifiers through the majority voting method. Table 4 shows the results of 63 labeled samples in the SRBCT dataset.

Table 4. Results of 63 labeled samples in the SRBCT dataset

Base classifier	The optimal operation (the highest correct rate)	Average correct rate	Final correct rate
SVM	4 genes were selected (0.9206)	0.8611	
KSUB-C	100 genes were selected and 20 KPCA principle components were extracted (0.9365)	0.9325	0.9524
KPLS-DA	4 potential factors were extracted (0.8730)	0.7064	

5   Conclusions

Kernel machine learning methods can effectively deal with small-scale sample problems as well as with samples of large dimensions. In addition, they can effectively extract nonlinear characteristics of samples. Therefore, they are suitable for analyzing gene expression data. By borrowing the idea of ensemble learning, we put forward an ensemble classifier based on kernel methods for analyzing multi-situation DNA microarray data. The detailed design steps are as follows. At first, we used the Kruskal-Wallis test to select a subset of genes whose expression levels fluctuate markedly. Then, three types of base classifiers are designed using the selected DNA microarray data, i.e., SVM, KSUB-C and KPLS-DA classifiers. In a third step, the results of the base classifiers are integrated to obtain the final classification result. The experimental results on Leukemia, mixed lineage leukemia, as well as small round blue cell tumor of DNA microarray datasets show that the proposed ensemble classifier not only makes multi-class data much separable, but also has an excellent generalization property.

Acknowledgements

References

1. Huang, D.S., Zheng, C.H.: Independent component analysis based penalized discriminant method for tumor classification using gene expression data. Bioinformatics 22(15), 1855–1862 (2006)
2. Chen, H.C., Liao, H.C., Huang, C.L.: Predicting breast tumor via mining DNA viruses with decision tree. In: Proceedings of the IEEE Int. Conf. on Systems, Man and Cybernetics, Taipei, Taiwan, pp. 3585–3589 (2006)
3. Asyali, M.H.: Gene expression profile class prediction using linear Bayesian classifiers. Computers in Biology and Medicine 37(12), 1690–1699 (2007)
4. Tan, Y.X., Shi, L.M., Tong, W.D.: Multi-class tumor classification by discriminant partial least squares using microarray gene expression data and assessment of classification models. Computational Biology and Chemistry 28, 235–244 (2008)
5. Narayanan, A., Keedwell, E.C., Gamalielsson, E., Tatineni, S.: Single-layer artificial neural networks for gene expression analysis. Neurocomputing 61(1-4), 217–240 (2004)
6. Shen, L., Tan, E.C.: Reducing multiclass cancer classification to binary by output coding and SVM. Computational Biology and Chemistry 30(1), 63–71 (2006)
7. Li, S.T., Liao, C., James, J.T.: Gene feature extraction using T-test statistics and kernel partial least squares. In: King, I., Wang, J., Chan, L.-W., Wang, D. (eds.) ICONIP 2006. LNCS, vol. 4234, pp. 11–20. Springer, Heidelberg (2006)
8. Vapnik, V.: The nature of statistical learning theory. Springer, New York (1995)
9. Baek, J.S., Son, Y.S.: Local linear logistic discriminant analysis with partial least square components. In: Li, X., Zaïane, O.R., Li, Z.-h. (eds.) ADMA 2006. LNCS (LNAI), vol. 4093, pp. 574–581. Springer, Heidelberg (2006)
10. http://www.ailab.si/orange/datasets.asp

Signaling Pathway Reconstruction by Fusing Priori Knowledge

Shan-Hong Zheng[1,2], Chun-Guang Zhou[1], and Gui-Xia Liu[1,*]

[1] Computer Science and Technology College, Jilin University, 130012, Changchun, China
[2] Computer Science and Engineering College, Changchun University of Technology,
130012, Changchun, China
Bioszsh2007@yahoo.cn, lgx1034@163.com

Abstract. Signaling pathway construction is one of hotspots in the present bio-informatics. A novel approach where priori knowledge is fused is proposed, called Dk-NICO, where partial missing regulation relationships and regulation directions are used as data samples, and biological experiment result as priori knowledge, while HMM is used as a model for reconstructing the signaling pathway, so as to predict signaling pathway. By reconstructing MAPK pathway, it is showed that the proposed approach not only is capable of predicting gene regulation relationships, but also is capable of identifying gene regulation directions. Moreover, we apply the approach to MAPK pathway reconstruction in the case of no priori knowledge and demonstrate that, by introducing priori knowledge from direct biochemical reaction experiment, the prediction accuracy is improved.

Keywords: Signaling transduction, Pathway prediction, HMM model, Biochemical reaction information.

1 Introduction

The signaling pathway describes the signal transduction process that causes certain endpoint biological function by modeling interactions among a series of cell components. Most biological functions of a cell are coordinated by the signaling pathways and regulatory mechanisms, and then cells respond to and accommodate oneself to an ever-changing environment. In recent years, theoretical and computational analysis of biochemical networks has been applied to well-defined metabolic pathways, signal transduction, and gene regulatory networks[1-3]. The existing theoretical literature on signaling is focused on networks where the elementary reactions and direct interactions are known; however quantitative characterization of every reaction and regulatory interaction participating even in a relatively simple function of a single-celled organism requires a concerted and decades-long effort[4], whereas these are important information of understanding the complex life process. Therefore, as an essential means to understand the complex life process and the pathogenesis of disease, signaling pathway construction becomes one of hotspots in the present bioinformatics.

---

* Corresponding author.

One typical method is to construct a molecular interaction network according to the gene-protein interaction or protein-protein interaction, and predict the signaling pathway using the graph searching [5,6]. Experiment results demonstrate how large-scale genomic approaches may be used to uncover signaling and regulatory pathways in a systematic, integrative fashion. However, there is no reliable evidence to testify if a specific molecular interaction exists in a specific signaling pathway, so that the reliability of the prediction can not be assured. With a rapid progress of the high-throughput technology, people start to study all kinds of methods of predicting signaling pathway utilizing expression profile [7-11]. The expression profile produced by the high-throughput measurement technology has been applied to identify the components in different signaling pathways. However, expression profile can only reflect sequence information in a coarse and unreliable level. Therefore, it is important to explore new way to predict signaling pathway.

Generally, as experimental means differs from each other, experimental information about a specific signal transduction can be partitioned into three categories such as biochemical reaction information, genetic experiment information and pharmacological experiment information. First, biochemical reaction information refers to enzymatic activity or protein-protein interactions. Second, genetic experiment information implicates the product of the mutated gene in the signal transduction process because of differential responses to a stimulus in wild-type organisms versus a mutant organism. Third, pharmacological experiment information provides observed relationships that are not direct interactions but indirect causal effects most probably resulting from a chain of interactions and reactions. Here, biochemical reaction information embodies direct interactions between the cell components. In signaling pathway prediction, this kind of information may be used as key information.

Based on the idea, a novel method of predicting signaling pathway, called Dk-NICO, is proposed, which is based on research from zhu et al[10]. The method is not limited to the numerical or categorical data used by previous approaches, while it integrates data from multiple data sources and multiple data type to predict signaling pathway, which can be obtained from biological experiments, so that prediction effect can be increased to higher extent. In this method, HMM is taken as a signaling pathway construction model, the data with genes regulation directions are used as samples, EM algorithm is used to optimize the model parameters with missing values. Finally, this method is testified by taking an example of MAPK signaling pathway.

2   Dk-NICO Method

2.1   Problem Proposition and Formal Description

All of the regulation relationship and regulation direction are necessary to understand signaling pathway. In this article, regulation relationship denotes that a cellular component regulates another component in certain signaling pathway, while regulation direction represents positive and negative regulation between the cellular components. Here we denoted positive and negative regulation as the verbs "promote" and "inhibit" and represented graphically as $\rightarrow$ and $\dashv$ respectively.

Based on above-mentioned knowledge, a signaling pathway can be abstracted as a weighted directed graph $DWG =< V, E, W >$, wherein V is vertex in the graph, E is edge, and W is weight of the edge. Obviously, V is corresponding to genes of signaling pathway, E is corresponding to the regulation relationship, W is corresponding to the regulation direction. That W exceeds a certain threshold such as 0.5 is deemed as positive regulation that is to activate expression of next gene, otherwise as negative regulation that is to inhibit next gene expression. Thus, the signaling pathway construction is to determine V, E and W in the DWG.

According to the method's idea that we presented, owing to taking regulation relationship and regulation direction information, coming from biological experiment, text mining, and protein interaction analysis, from the literature, as data sample, the data samples could contain missing data. Therefore, EM algorithm is used to optimize the model parameters. In EM algorithm, the initial value affects the correctness of the result to a certain extent. In this paper, we introduce part of biochemical reaction information as priori knowledge, so that accuracy of prediction can be improved.

Under the direction of these ideas, HMM is used as a model of weighted directed graph, the genes serves as HMM state, the regulation direction information is used as observation in HMM, and HMM parameters is optimized by Dk-NICO where priori knowledge is added, which is based on EM algorithm. Furthermore, in order to simplify complexity of the problem, we assume that the genes appearing in each sample should have same trend about regulation relationship and regulation direction. In fact, according to current observations the reality is not far: the average in/out degree of the mammalian signal transduction network is close to 1[12].

2.2   Dk-NICO Method

HMM is defined by the initial state distribution π, the state transition matrix A, and the observation distribution B, that is $HMM = (A, B, \Pi)$, wherein $\Pi = \{\pi_i \mid i = 1, 2, ..., G\}$, $A = \{a_{ij} \mid i = 1, 2, ..., G; j = 1, 2, ..., G\}$, and $B = \{b_j(k) \mid j = 1, 2, ...G; k = 1, 2\}$, wherein π_i represents the probability of gene i as initial state, a_{ij} represents the probability of gene i regulating j, and $b_j(k)$ represents the probability of gene j having regulation direction k. Here, $k = 1$ represents regulation direction is positive, and $k = 2$ represents regulation direction is negative. Obviously, the following conclusion is drawn.

$$\sum_{1 \le i \le G} \pi_i = 1 \text{ and } \pi_i \ge 0; i = 1, 2, ...G. \tag{1}$$

$$\sum_{1 \le j \le G} a_{ij} = 1 \quad \text{and } a_{ij} \ge 0; i, j = 1, 2, ..., G. \tag{2}$$

$$\sum_{1 \le k \le 2} b_j(k) = 1 \quad \text{and } b_j(k) \ge 0, j = 1, 2, ..., G; k = 1, 2. \tag{3}$$

Provided the data sample is $Y = \{y^{(m)} \mid m = 1, 2, ..., M\}$, wherein M is the sample size. It is worth mentioning that the data sample Y contains regulation direction information. Here RS is used to represent the regulation relationship as $RS = \{rs^{(m)} \mid m = 1, 2, ..., M\}$, while D is defined to represent regulation direction

as $D = \{d^{(m)} \mid m = 1, 2, ..., M\}$.   Here, $rs^{(m)} = < rs_1^{(m)}, rs_2^{(m)}, ..., rs_{Nm}^{(m)} >$,
$d^{(m)} = < d_1^{(m)}, d_2^{(m)}, ..., d_{Nm}^{(m)} >$, wherein $rs_n^{(m)}$ and $d_n^{(m)}$ are the regulation relationship and
direction of the corresponding gene n contained in the mth sample. Nm is the total
number of genes contained in the mth sample. For conveniently deducing the algo-
rithm, a matrix representation equivalent to RS in the algorithm is defined as X.
$rs^{(m)}$ is expressed as an equivalent matrix form $x^{(m)} = \{x_{ij}^{(m)} \mid i = 1, 2, ..., Nm; j = 1, 2, ..., G\}$,
wherein G is the total number of genes in the signaling pathway. If $rs_i^{(m)} = j$, then
$x_{ij}^{(m)} = 1$, or else $x_{ij}^{(m)} = 0$. The permutation corresponding to RS is defined
as $\Gamma = \{\tau^{(m)} \mid m = 1, 2, ..., M\}$.   Each $y^{(m)}$ is   defined   as   a   permutation
set $\tau^{(m)} = \{\tau^{(m)(k)} \mid k = 1, 2, ..., Nm!\}$,   wherein $\tau^{(m)(k)} = < \tau_1^{(m)(k)}, \tau_2^{(m)(k)}, ..., \tau_{Nm}^{(m)(k)} >$. Like-
wise,   each   permutation $\tau^{(m)(k)}$ is   expressed   as   an   equivalent   matrix
form $r^{(m)} = \{r_{ij}^{(m)} \mid i, j = 1, 2, ..., Nm\}$. If $\tau_i^{(m)(k)} = j$, then $r_{ij}^{(m)} = 1$, or else $r_{ij}^{(m)} = 0$. $d^{(m)}$ is ex-
pressed as an equivalent matrix form $\beta^{(m)} = \{\beta_{ij}^{(m)} \mid i = 1, 2, ..., Nm; j = 1, 2\}$. If $d_i^{(m)} = j$,
then $\beta_{ij}^{(m)} = 1$, or else $\beta_{ij}^{(m)} = 0$. The permutation corresponding to D is defined
as $C = \{c^{(m)} \mid m = 1, 2, ..., M\}$.   Each $c^{(m)}$ is   defined   as   a   permutation
set $c^{(m)} = \{c^{(m)(k)} \mid k = 1, 2, ..., cNm\}$, wherein $c^{(m)(k)} = < c_1^{(m)(k)}, c_2^{(m)(k)}, ..., c_{cNm}^{(m)(k)} >$, and $cNm = 2^{Nm}$.
Likewise,   each   permutation $c^{(m)(k)}$ is   expressed   as   equivalent   matrix
form $\gamma^{(m)} = \{\gamma_{ij}^{(m)} \mid i = 1, 2, ..., Nm; j = 1, 2\}$. If $c_i^{(m)(k)} = j$, then $\gamma_{ij}^{(m)} = 1$, or else $\gamma_{ij}^{(m)} = 0$.

Based on the above definition, priori knowledge is introduced to our method in or-
der to improve algorithm's accuracy. Here we use the parameter uk_i to represent priori
value of gene i as initial state, vk_{ij} of gene i regulating gene j , and p_{jk} of gene j
having regulation direction k . Obviously, the parameter uk_i, vk_{ij} and p_{jk} should be
non-negative in order to have proper priors.

In our method, using EM algorithm is to optimize model parameters in case of hav-
ing missing or false data. The EM algorithm alternates between the expectation or
E-step and the maximization or M-step. More precisely, the computing method of
E-step and M-step is given as below:

(1) E-step

$$\bar{r}_{1t'}^{(m)} = E[r_{1t'}^{(m)} \mid x^{(m)}, \hat{A}, \hat{\Pi}] = \frac{\sum\limits_{r \in \tau^{(m)}} r_{1t'}^{(m)} P[x^{(m)} \mid r, \hat{A}, \hat{\Pi}]}{\sum\limits_{r \in \tau^{(m)}} P[x^{(m)} \mid r, \hat{A}, \hat{\Pi}]} \tag{4}$$

$$\bar{\alpha}_{t't''}^{(m)} = E[\sum_{t=2}^{Nm} r_{tt'}^{(m)} r_{(t-1)t''}^{(m)} \mid x^{(m)}, \hat{A}, \hat{\Pi}] = \frac{\sum\limits_{r \in \tau^{(m)}} r_{tt'}^{(m)} r_{(t-1)t''}^{(m)} P[x^{(m)} \mid r, \hat{A}, \hat{\Pi}]}{\sum\limits_{r \in \tau^{(m)}} P[x^{(m)} \mid r, \hat{A}, \hat{\Pi}]} . \tag{5}$$

$$\bar{\beta}_{t't''}^{(m)} = E[\beta_{t't''}^{(m)} \mid \gamma, \hat{B}] = \frac{\sum\limits_{\gamma \in \Phi_{Nm}} \gamma_{t't''}^{(m)} P[\beta^{(m)} \mid \gamma, \hat{B}]}{\sum\limits_{\gamma \in \Phi_{Nm}} P[\beta^{(m)} \mid \gamma, \hat{B}]} . \tag{6}$$

Here,

$$P[x^{(m)} \mid r, \hat{A}, \hat{\Pi}] = P[rs^{(m)} \mid \tau^{(m)}, \hat{A}, \hat{\Pi}] = \hat{\pi}_{rs_{\tau_1^{(m)(k)}}^{(m)}} \prod_{t=2}^{Nm} \hat{a}_{rs_{\tau_{t-1}^{(m)(k)}}^{(m)} rs_{\tau_t^{(m)(k)}}^{(m)}} . \tag{7}$$

$$P[\beta^{(m)} \mid \gamma, \hat{B}] = P[d^{(m)} \mid c, \hat{B}] = \prod_{t=1}^{Nm} b_{d_{c_t^{(m)}}^{(m)}} . \tag{8}$$

(2) M-step

$$(\hat{\pi}_i)_{new} = \frac{uk_i + \sum\limits_{m=1}^{M} \sum\limits_{t'=1}^{Nm} \bar{r}_{1t'}^{(m)} x_{t'i}^{(m)}}{\sum\limits_{i=1}^{G} (uk_i + \sum\limits_{m=1}^{M} \sum\limits_{t'=1}^{Nm} \bar{r}_{1t'}^{(m)} x_{t'i}^{(m)})} ; i = 1, 2, ..., G . \tag{9}$$

$$(\hat{a}_{ij})_{new} = \frac{vk_{ij} + \sum\limits_{m=1}^{M} \sum\limits_{t',t''=1}^{Nm} \bar{\alpha}_{t't''}^{(m)} x_{t'i}^{(m)} x_{t'j}^{(m)}}{\sum\limits_{j=1}^{G} (vk_{ij} + \sum\limits_{m=1}^{M} \sum\limits_{t',t''=1}^{Nm} \bar{\alpha}_{t't''}^{(m)} x_{t'i}^{(m)} x_{t'j}^{(m)})} ; i, j = 1, 2, ..., G . \tag{10}$$

$$(\hat{b}_j(k))_{new} = \frac{p_{jk} + \sum\limits_{m=1}^{M} \sum\limits_{t'=1}^{Nm} \bar{\beta}_{t'k}^{(m)} x_{t'j}^{(m)}}{\sum\limits_{k=1}^{2} (p_{jk} + \sum\limits_{m=1}^{M} \sum\limits_{t'=1}^{Nm} \bar{\beta}_{t'k}^{(m)} x_{t'j}^{(m)})} ; j = 1, 2, ...G, k = 1, 2 . \tag{11}$$

For the signaling pathway construction, in many cases, the start gene and terminal gene can be known. In this case, the above computing process can be simplified, and the computing efficiency can be improved. Then, the initial state distribution Π can be directly obtained according to the sample observation. The formula of computing Π is given as blow:

$$\pi_i = \frac{\sum\limits_{m=1}^{M} x_{1i}^{(m)}}{M} ; i = 1, 2, ..., G . \tag{12}$$

The computing method of the state transition matrix A and observation distribution B does not change. But while computing E-step, for the permutation matrix r, only $r_{11}^{(m)} = 1$ and $r_{NmNm}^{(m)} = 1$ have to be considered. So, state transition matrix A meets the following restriction condition:

$$\sum_{1 \le j \le G} a_{ij} = 1 \text{ and } i, j = 1, 2, ..., G; i \ne l , \tag{13}$$

Where l is the row number corresponding to the terminal gene in the matrix A.

3   Results

Appling the above method to the construction of representative signaling pathway (MAPK pathway) illustrates that in case of missing partial gene regulation data, the method still can effectively predict the signaling pathway structure. Moreover, MAPK pathway is reconstructed in the case of no priori knowledge again. The result shows that fusing priori knowledge into the presented algorithm improves algorithm's accuracy.

3.1   MAPK Pathway Construction with Priori Knowledge

In the mammalian organism, five different MAPK pathways have been found. MAPK pathway can foster the vascular endothelial cell proliferation and angiogenesis which can provide more nutriment to the tumor, so as to foster the growth of tumor and proliferation for cancer cell. SAPK/JNK pathway is a pathway in the MAPK family that is suffered from various environment stress, inflammatory cytokines, growth factor and GPCR agonist[10]. Recent research demonstrates that, in addition to immune responses, SAPK/JNK activation plays essential roles in organogenesis during mouse development by regulating cell survival, apoptosis, and proliferation[13].

Fig. 1 gives the sample data for SAPK pathway construction. The first place on the left indicates a kind of extracellular signals or membrane proteins essential to activate a special signaling pathway, which is identified as start gene; the last place indicates transcription factor activated by the signaling pathway when extracellular signals are transmitted into the cell, which is identified as terminal gene. Red indicates regulation direction is negative; green indicates regulation is positive. Fig. 2 gives priori value used in the method.

(1) GF,HPK,MEKK,MKK,JNK
(2) GF,MEKK,HPK,MKK,JNK
(3) GF,RAC,RAS,MEKK,MKK,JNK
(4) GF,RAS,CDC42,RAC,MKK,MEKK,JNK
(5) GF,RAS,RAC,RHO
(6) CS1,RAC,MEKK,MKK,CDC42,JNK
(7) CS2,MEKK,MKK,RAC,JNK
(8) FASL,GCKs,MKK,MEKK,JNK
(9) OS,ASK1,MEKK,MKK,JNK

$(1) HPK \rightarrow MEKK$
$(2) RAS \rightarrow RAC$
$(3) CDC42 \rightarrow MEKK$
$(4) ASK1 \rightarrow MKK$

Fig. 1. Samples essential to reconstruct SAPK pathway

Fig. 2. Priori vlaue essential to reconstruct SAPK pathway

For program to process conveniently, the genes in the sample are numbered, as shown in Tab. 1.

Setting the iteration to 100 times, and computing with Dk-NICO method, the state transition probability A and observation distribution B are obtained as shown in Fig. 3. According to the definition of Dk-NICO, each element in matrix A represents the regulation relationship between the related genes, wherein a_{ij} represents the probability of gene i regulating j. The row number with of all zero value in matrix A represents the corresponding number of terminal gene; the value of each element in matrix

$$
A = \begin{bmatrix}
0 & 0.4 & 0 & 0 & 0 & 0 & 0.6 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0.120026 & 0.666667 & 0 & 0.213307 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0.181102 & 0 & 0 & 0.818898 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\
0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\
0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0
\end{bmatrix}
\qquad
B = \begin{bmatrix}
0.823495 & 0.176505 \\
1 & 0 \\
1 & 0 \\
0.992187 & 0.007813 \\
0 & 0 \\
1 & 0 \\
1 & 0 \\
1 & 0 \\
0 & 0 \\
1 & 0 \\
1 & 0 \\
1 & 0 \\
1 & 0 \\
1 & 0 \\
1 & 0
\end{bmatrix}
$$

Fig. 3. The running results of Dk-NICO

B represents the probability of regulation direction of the corresponding genes being positive or negative, that is $b_j(k)$ represents the probability of gene j being k. Likewise, the row number with of all zero value in matrix B represents the corresponding number of terminal gene.

The SAPK pathway structure constructed by Dk-NICO as shown in Fig. 4.
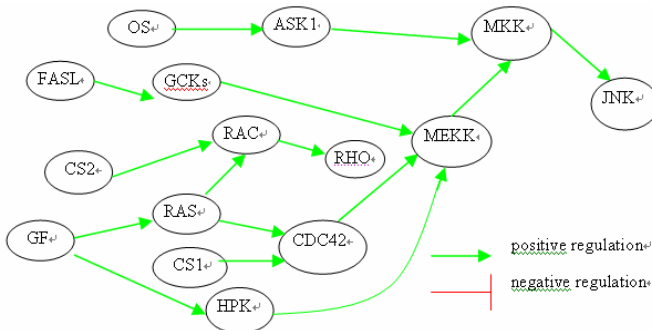
Fig. 4. SAPK signaling pathway topology reconstructed from Fig.3

Table 1. Comparison table of gene number in SAPK pathway

gene number	gene name	gene number	gene name
1	GF	9	RHO
2	HPK	10	CS1
3	MEKK	11	CS2
4	MKK	12	FASL
5	JNK	13	GCKs
6	RAC	14	OS
7	RAS	15	ASK1
8	CDC42		

3.2 MAPK Pathway Construction without Priori Knowledge

Similar to above process, we still use the samples in Tab. 1. Setting the iteration to 100 times, and not fusing priori knowledge, the result is obtained as shown in Fig. 5.

A SAPK pathway structure is shown in Fig. 6 based on the above computing result.

$$
A = \begin{bmatrix}
0 & 0.40 & 0 & 0 & 0 & 0.6 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0.120026 & 0.666667 & 0 & 0.213307 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0.210924 & 0 & 0 & 0.789076 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\
0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\
0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0
\end{bmatrix}
\qquad
B = \begin{bmatrix}
0.823495 & 0.176505 \\
0.450027 & 0.549973 \\
1 & 0 \\
0.992187 & 0.007813 \\
0 & 0 \\
1 & 0 \\
0.012094 & 0.987906 \\
1 & 0 \\
0 & 0 \\
1 & 0 \\
1 & 0 \\
1 & 0 \\
1 & 0 \\
1 & 0 \\
1 & 0
\end{bmatrix}
$$

Fig. 5. The results of A and B under the condition of not having priori knowledge
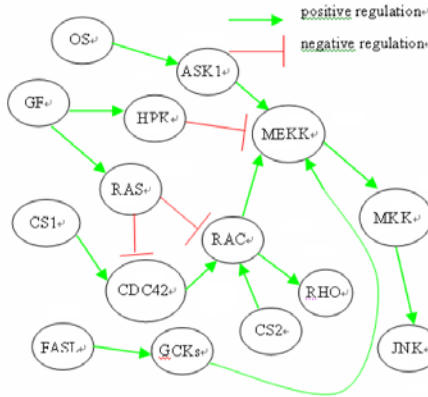
Fig. 6. SAPK signaling pathway topology reconstructed from Fig. 5

4 Discussion

Contrasting the SAPK pathway structure in Fig. 4 with the cellsignaling database (http://www.cellsignal.com), it can be seen that regulation relationship and regulation direction predictions are correct on the basis of priori knowledge. However, contrasting the SAPK pathway structure in Fig. 6 with it in Fig. 4, it can be seen that most regulation relationship and regulation direction predictions are correct. But there are several errors. In Fig. 6, for instance, the regulation relationship between CDC42 and RAC is not predicted correctly. For CDC42 and RAC, the regulation direction between RAS and them is predicted wrongly because of data samples' errors under the condition of no priori knowledge. The main reason is that some data samples

are probably from analytic results of high-throughput data. By introducing priori knowledge from biochemical reaction information – 'RAS→RAC', the problem has been solved. This proved our method's efficiency further. Besides, some errors consist in prediction of the regulation direction relevant to MEKK and RAS. The comparison result demonstrates that there are eleven being exactly predicted while four not being predicted in fifteen paths and three is wrongly predicted. We compute accuracy rate on the basis of the above data, and the computation result is shown in Tab. 2.

Table 2. the computation result of accuracy rate

condition	accuracy rate
not considering regulation direction	93.3%
considering regulation direction	73.3%

5   Conclusion

This article proposes a Dk-NICO method which is based on EM algorithm. Based on the regulation relationship and regulation direction, the computing method for signaling pathway construction based on HMM is demonstrated. Taking partial missing data as samples and fusing priori knowledge from direct biological experiments, MAPK pathway are reconstructed. The results are testified according to the data in the cellsignaling database. This method can predict not only the regulation relationship between the genes in the signaling pathway, but also the regulation directions, which more specifically describe the pathway.

In our method, as matrix calculation and recursive algorithm are used, algorithm's efficiency can be fallen when the number of data is large and dimensionality is high. Our result of simulation experiments indicated this. The running time is about 20 seconds when the number of genes is 15, however, when the number of genes is increased to 50, the running time climbed to 5 minutes. In addition, when determining the regulation direction, a specific gene regulating only one gene in a specific signaling pathway is considered. In a specific signaling pathway, however, possibility of one gene regulating two or more genes exists. Under this circumstance, the regulation direction of genes maybe varies with different regulation genes. Therefore, integrating more complicated network model to the current method can solve this problem.

Acknowledgement

References

1. Fall, C., Marland, E., Wagner, J., Tyson, J.: Computational Cell Biology. Springer, New York (2002)
2. Voit, E.: Computational Analysis of Biochemical Systems. Cambridge University Press, Cambridge (2000)
3. Bower, J., Bolouri, H.: Computational Modeling of Genetic and Biochemical Networks. MIT Press, Cambridge (2001)
4. Albert, R., DasGupta, B., Dondi, R., Sontag, E.D.: Inferring (biological) Signal Transduction Networks via Transitive Reductions of Directed Graphs. Algorithmica 51(2), 129–159 (2008)
5. Ideker, T., Ozier, O., Schwikowski, B., Siege, A.F.: Discovering Regulatory and Signaling Circuits in Molecular Interaction Networks. Bioinformatics 18, S233–S240 (2002)
6. Rajagopalan, D., Agarwal, P.: Inferring Pathways from Gene Lists Using a Literature-Derived Network of Biological Relationships. Bioinformatics 21(6), 788–793 (2005)
7. Li, Z., Chan, C.: Inferring Pathways and Networks with a Bayesian Framework. Faseb J. 18(6), 746–748 (2004)
8. Novak, B., Jain, A.: Pathway Recognition and Augmentation by Computational Analysis of Microarray Expression Data. Bioinformatics 22(2), 233–241 (2006)
9. Darvish, A., Najarian, K.: Prediction of Regulatory Pathways Using mRNA Expression and Protein Interaction Data: Application to Identification of Galactose Regulatory Pathway. Biosystems 83, 125–135 (2006)
10. Zhu, D., Rabbat, M., Hero, A.O., Nowak, R., Figueiredo, M.: De Novo Signaling Pathway Reconstruction from Multiple Data Sources. In: Columbus, F. (ed.) New Research on Signal Transduction. Nova Publishers (2006)
11. Li, Z., Srivastava1, S., Mittal, S., Yang, X., Sheng, L.F., Chan, C.: A Three Stage Integrative Pathway Search (TIPS) Framework to Identify Toxicity Relevant Genes and Pathways. BMC Bioinformatics 8, 202–209 (2007)
12. Ma'ayan, A., Jenkins, S.L., Neves, S., et al.: Formation of Regulatory Patterns During Signal Propagation in a Mammalian Cellular Network. Science 309(5737), 1078–1083 (2005)
13. Nishina, H., Wada, T., Katada, T.: Physiological Roles of SAPK/JNK Signaling Pathway. J. Biochem. 36(2), 123–126 (2004)

Dynamic Identification and Visualization of Gene Regulatory Networks from Time-Series Gene Expression Profiles

Yu Chen and Kyungsook Han[*]

School of Computer Science and Engineering, Inha University, Incheon, Korea
`khan@inha.ac.kr`

Abstract. Recent improvements in high-throughput proteomics technology have produced a large amount of time-series gene expression data. The data provide a good resource to uncover causal gene-gene or gene-phenotype relationships and to characterize the dynamic properties of the underlying molecular networks for various biological processes. Several methods have been developed for identifying the molecular mechanisms of regulation of genes from the data, but many of the methods consider static gene expression profiles only. This paper presents a new method for identifying gene regulations from the time-series gene expression data and for visualizing the gene regulations as dynamic gene regulatory networks. The method has been implemented as a program called DRN Builder (Dynamic Regulatory Network Builder; http://wilab.inha.ac.kr/drnbuilder/) and successfully tested on actual gene expression profiles. DRN Builder will be useful for generating potential gene regulatory networks from a large amount of time-series gene expression data and for analyzing the identified networks.

1   Introduction

The large amount of gene expression data has contributed to revealing many mechanisms of biological processes. Most mechanisms are controlled by complex regulatory interactions between genes rather than by a single gene [1]. Therefore, fully identifying the regulations of the molecular determinants and their interplays has become the focus of the modern biology [2]. Several methods have been proposed to identify gene regulations, but most methods consider static gene expression profiles at a time point, and cannot directly handle the changes in gene regulations over time.

Gene regulation often varies over time rather than being constant [3]. Suppose that there is a gene (g_1) whose inhibitory effect (g_2) depends on an inducer (g_3) to be able to bind to the inhibition site on g_2. In this case, there can be a significant delay between the expression of the inhibitor gene g_1 and its observed effect, i.e., the inhibition of g_2. A Boolean model proposed by Silvescu et al. [1] considers the regulation with multiple-time delays but the Boolean model is restricted to logical relationships between variables and relies on the assumption that the expression of a gene is likely to be controlled by other genes. Model-based approaches such as probabilistic

---

[*] Corresponding author.

Boolean networks [4] and dynamic Bayesian networks [5] can reconstruct longitudinal regulatory networks from a set of mathematical equations if the equations precisely specify the networks, but may fail when the underlying model is not correct [6]. A more reliable and general method is required to identify the gene regulations from time-series gene expression profiles.

In this study, we propose a new method for identifying gene regulations from the temporal gene expression profiles and for visualizing the identified gene regulations in a network. We implemented the method in a program called DRN Builder (Dynamic Regulatory Network Builder) and tested it on actual gene expression data.

2   Methods

2.1   Gene Expression Data

The gene expression data of m genes with n samples is represented as an m×n matrix, where rows represent genes and columns represent various samples such as experimental conditions or time points in a biological process. Each element of the matrix represents the expression level of a particular gene in a particular sample. Fig. 1 shows an example of the gene expression matrix for yeast genes during the yeast cell cycle, obtained from the Yeast Cell Cycle Analysis Project [7].
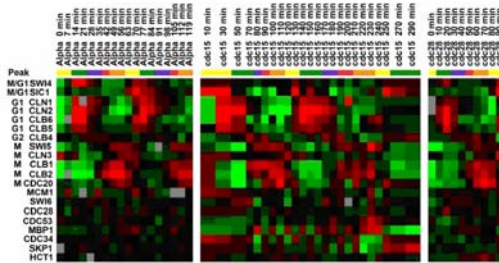
Fig. 1. Gene expression during the yeast cell cycle [7]. Each row represents a gene and each column represents a time point. Red color indicates mRNA abundance and green color indicates absence of mRNA. The color at the top bar represents the cell cycle phases (M/GI in yellow, GI in green, S in purple, G2 in red, and M in orange).

2.2   Weight of Gene Regulations

Once the raw image of gene expressions is represented as a matrix, the matrix is analyzed for similarity between gene expressions at different time points. Two genes with highly similar expression patterns tend to be co-expressed. Three metrics are often used to measure the similarity of genes: Pearson correlation [8], Euclidean distance [9] and Spearman correlation [10]. Pearson correlation is useful for discovering the gene regulations with a significant change in expression level, but not for describing the level of gene expression. Euclidean distance is useful for scaling the level of gene expression. For example, two genes with a similar expression at time point k will have a large Euclidean distance if they have different levels of gene expression. Spearman correlation acts similarly as the Pearson correlation, but it is insensitive to probing volatile extreme values.

We compute a weight of a regulation between genes X and Y by Equation 1, which is based on the Pearson correlation. R in Equation 1 denotes the dependency relation between two genes. Equation 2 formulates a modified weight $R(X, i, Y, i+p, p)$ to include a directional and temporal difference in gene regulation. $R(X, i, Y, i+p, p)$ represents a weight of a regulation between gene X at time point i and gene Y at time point $i+p$. Among $(i \times p)$ candidates, the regulation with the maximum absolute value of $R(X, i, Y, i+p, p)$ will be selected as the regulation between X and Y.

$$R = \frac{\sum X_i Y_i - N \overline{X}\,\overline{Y}}{(N-1) S_x S_y} = \frac{N \sum X_i Y_i - \sum X_i \sum Y_i}{\sqrt{N \sum X_i^2 - (\sum X_i)^2} \sqrt{N \sum Y_i^2 - (\sum Y_i)^2}} \tag{1}$$

$$R(X, i, Y, i+p, p) = \frac{\sum_{i=\max[1,p]}^{\min[N, N+p]} (X_i - \overline{X})(Y_{i+p} - \overline{Y})}{\sqrt{(\sum_{i=1}^{N} (X_i - \overline{X})^2)(\sum_{i=1}^{N} (Y_i - \overline{Y})^2)}} \tag{2}$$

In Equations 1 and 2, N is the total number of time points, X_i and Y_i are the expression levels of genes X and Y at time i, and $\overline{X}$ and $\overline{Y}$ are the average gene expression levels at all time points. p is the time delay of a gene regulation. For example, p=3 for the regulation between gene A at time 0 and gene B at time 3.

2.3 Identification of Gene Regulations

In our study, transcription of a gene is assumed to be in a finite number of different states such as 'up-regulated' and 'down-regulated'. The gene expression level represents the log-ratio of the red and green intensities, so we consider genes with positive expression levels are up-regulated and that genes with negative expression levels are down-regulated.

From the gene expression profiles in a series of time points, DRN Builder identifies the gene regulations and includes them in the regulation list. In the regulation list, +A(t) indicates that gene A is up-regulated at time t, and –A(t) indicates that gene A is down-regulated at time t. The symbol '->' represents a directional relationship between genes. For the expression levels of two genes at different time points, there are four possible relations:

1. +A(t_1) -> +B(t_2): up-regulation of A at time t_1 is followed by up-regulation of B at time t_2 ($t_2>t_1$).
2. –A(t_1) -> +B(t_2): down-regulation of A at time t_1 is followed by up-regulation of B at time t_2 ($t_2>t_1$).
3. +A(t_1) -> –B(t_2): up-regulation of A at time t_1 is followed by down-regulation of B at time t_2 ($t_2>t_1$).
4. –A(t_1) -> –B(t_2): down-regulation of A at time t_1 is followed by down-regulation of B at time t_2 ($t_2>t_1$).

The relations +A(t_1) -> +B(t_2) and –A(t_1) -> +B(t_2) have positive R values, whereas +A(t_1) -> –B(t_2) and –A(t_1) -> –B(t_2) have negative R values. The weight of each gene regulation is then iteratively calculated using Equation 2. For genes g_i and g_j, the regulation with the largest absolute value of the weight is chosen for the regulation between the genes and represented as R(A, t_1, B, t_2, p), where p=t_2-t_1. The algorithm for constructing the regulation list can be described as follows:

1. Compute the weight of the regulation `R(A, t₁, B, t₂, p)` between gene A at time point t_1 and gene B at time point t_2 for all pairs of genes. `p=t₂–t₁.`
2. Select the regulation with the largest absolute value of `R(A, t₁, B, t₂, p)`.
3. If p>0, classify the regulation into one of the four types, `+A(t₁) -> +B(t₂)`, `–A(t₁) -> +B(t₂)`, `+A(t₁) -> -B(t₂)`, `-A(t₁) -> -B(t₂)`, and add it to the regulation list.
4. If p=0, two genes are co-expressed or co-inhibited, and such gene regulation is not added to the regulation list.
5. If the new gene regulation is already in the regulation list, merge it with the previous regulation.
6. Go to step 2 to find the next gene regulation until no more regulation found.

2.4   Visualization of Gene Regulatory Networks

All regulations identified in the previous step are visualized by DRN Builder as a 2-dimensional gene regulatory network. In the gene regulatory network, nodes that are associated with the cell/environment interfaces are put along vertical lines. Arrows and blocked arrows represent inductive interactions and inhibitory interactions, respectively. The regulator gene, type of regulation (+ for induction and – for inhibition), and time delay of the regulation are annotated as arrow labels.

   Two layout algorithms were implemented for drawing a gene regulatory network by DRN Builder: grid layout and layered graph layout. The grid layout algorithm can be described as follows:

1. Find the node with the highest degree, and place the node in the center grid (node S in Fig. 2A). If there is a tie, select a node with a higher out-degree.
2. Position all nodes connected to the center node in the adjacent grids. Nodes with a higher degree are positioned earlier than those with a lower degree in the east, north, west, south, northeast, northwest, southwest, and south east grid of the center grid (node numbers 1-8 of Fig. 2A show the order). If more than 8 nodes are connected to the center node, put the 9th node to the east of the 1st node, the 10th node to the north of the 2nd node, and so on.
3. Repeat step 2 for nodes connected to nodes that are already positioned. If an adjacent grid is occupied, move to the next possible position until it is available.
4. If there are disconnected nodes, repeat step 1 for the nodes and put them to the right of the previous subgraph.
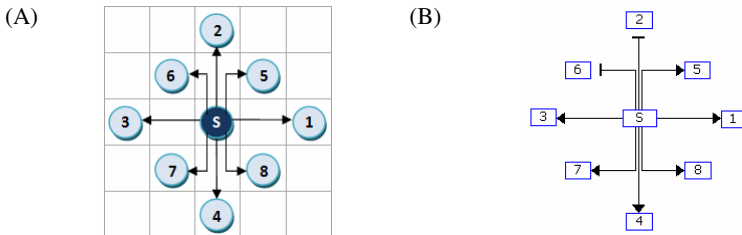
Fig. 2. Example of the grid layout. (A) Node S with the highest degree is placed in the center grid, and the nodes connected to S are placed in the adjacent grids in the specified order. (B) Grid drawing by DRN Builder.

In the layered graph layout, all nodes are assigned to a layer and the node with the maximum degree is assigned to the top layer. The degree of a node includes both in-degree and out-degree. The layered drawing usually takes more time to generate than the grid drawing. The layered drawing is generated in the following way:

1. Put the node with the maximum degree at layer 1. If there is a tie, select a node with a higher out-degree.
2. Assign the nodes connected the nodes at layer i to layer $i+1$ (see Fig. 3A).
3. Repeat steps 1 and 2 for the remaining nodes (see Fig. 3B).
4. If two nodes at the same layer are connected to each other, make a new layer between the layer and the upper layer and move the node with a smaller degree to the new layer (node 4 in Fig. 3C). Nodes with 0 out-degrees (node 1 in Fig. 3C) are also moved to the new layer.
5. Order the nodes in each layer by the Barycenter method [11] to reduce the number of edge crossings (Fig.3D).
6. If there are disconnected nodes, repeat steps 1-5 for the nodes and put them to the right of the previous subgraph.
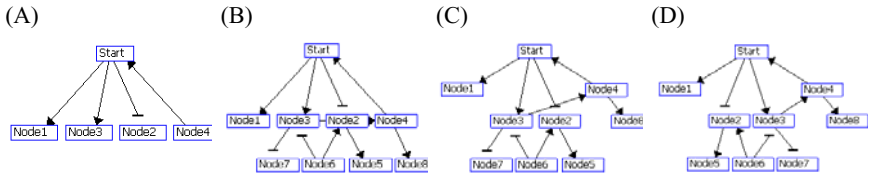
Fig. 3. Example of the layered graph

3 Experimental Results

DRN Builder was tested on two data sets of gene expression profiles. The first data set, shown in Fig. 1, is the data of yeast cell cycling, which was prepared by Spellman et al. [7] from the Saccharomyces cerevisiae cell cultures that were synchronized by three different methods: cdc15, cdc28 and alpha-factor. There are 19, 17, 18 time points for cdc15, cdc28 and alpha-factor, respectively. The 19 yeast cycling genes, CLN1, CLN2, CLN3, CLB1, CLB2, CLB4, CLB5, CLB6, CDC28, MBP1, CDC53, CDC34, SKP1, SWI4, SWI5, SWI6, CDC20, SIC1, and MCM1 are known to be involved in the cell-cycle regulations.

The second data set is the Northern blot image of 8 Arabidopsis response regulators (ARR) of type A in response to exogenous cytokinin in etiolated Arabidopsis seedlings [12]. Fig. 4A shows ARR genes of types A and B, which differ in their sequence and domain structure [12]. We quantified the image of Fig. 4B using ImageQuant for testing.

In the data of yeast cell cycling, DRN Builder identified a total of 40 potential gene regulations (Table 1). To assess the 40 regulations, we searched them in the KEGG, SGD and CYGD databases and literatures. KEGG currently has 29,471 pathways whereas SGD (http://www.yeastgenome.org/) and CYPD (http://mips.gsf.de/genre/proj/yeast/) provide genetics and functional networks of the budding yeast Saccharomyces cerevisiae, respectively.

(A)   (B)

Fig. 4. (A) Gene family of Arabidopsis response regulators. Genes from ARR 3 to ARR 9, and ARR 15, ARR 16, ARR 17 are of type A. The remaining genes are of type B. (B) Northern-blot analysis of genes from ARR at various times (indicated in minutes above each line).

33 out of the 40 potential gene regulations were found in at least one of KEGG, SGD and CYGD. 21 of the 33 gene regulations, which are underlined in Table 1, show exact agreement with the data of the databases, and these regulations were determined by experimental methods (see supplementary data at http://wilab.inha.ac.kr/DRNBuilder). 12 gene regulations, written in italics in Table 1, were not determined directly by experimental methods but were implied by several studies (supplement data). The remaining 7 regulations had no supporting database or literature, so more work should be done to verify them. Even if the 7 regulations are not correct regulations, at least 82% (33 out of 40) of the regulations found by DRN Builder correspond to known gene regulations.

Fig. 5. A grid layout of the yeast cell cycle regulatory network visualized by DRN Builder. This network shows all the gene regulations of Table 1. Edge labels represent the regulators and time points of the regulation.

Table 1. Gene regulations identified in the time-series expression profiles of yeast cell cycling. Underlined entries denote the regulations determined by experimental methods, and italicized entries denote the regulations implied by previous studies. The KEGG pathway ID is placed within brackets.

Genes	T	T+1	T+2	T+3
CLN1	-CLN1(T) -> -CLB4(T+1)	-CLN1(T+1) -> -CLB4(T+2)		
CLN2	-CLN2(T) -> -SWI6(T+1)		-CLN2(T+2)-> +SIC1(T+5)	
CLN3	+CLN3(T) -> +SIC1(T+1) [CYGD] +CLN3(T)-> +CLB6(T+3)			
CLB1	+CLB1(T) ->-CLN2(T+1) +CLB1(T)-> +SWI5(T+1) [KEGG:ko04110] -CLB1(T) -> -CDC20(T+1)	+CLB1(T+1)-> +CDC20(T+3)	+CLB1(T+2)-> +CDC20(T+4)	
CLB2		*+CLB2(T+1)-> +SIC1(T+3)*	+CLB2(T+2)-> +SWI4(T+5)	+CLB2(T+3)-> +SWI4(T+6)
CLB6	+CLB6(T) -> -CLB1(T+1) +CLB6(T) -> -CLB2(T+1)	+CLB6(T+1) -> -CLB1(T+2) +CLB6(T+1) -> -CLB2(T+2)	+CLB6(T+2) -> -CLB1(T+3) +CLB6(T+2) -> -CLB2(T+3)	
MCM1		+MCM(T+1)-> -MBP1(T+2)		
SIC1		*-SIC1(T+1)-> +SWI5(T+3) [CYGD] -SIC1(T+1)-> -CLN2(T+3)*	*-SIC1(T+2)-> +SWI5(T+4) [CYGD] -SIC1(T+2)-> -CLN2(T+4)*	*-SIC1(T+3)-> +SWI5(T+5) [CYGD] -SIC1(T+3)-> -CLN2(T+5)*
SWI6		-SWI6(T+1)-> -SKP1(T+2)		*-SWI6(T+3) -> -CDC20(T+6)* [KEGG: ko04110]
CDC28	+CDC28(T)-> +SWI4(T+1)			
CDC53	*+CDC53(T) ->-CLN3(T+1)*			
MBP1	-MBP1(T)-> +CDC28(T+1) [SGD]		+MBP1(T+2)-> +SKP1(T+5)	
CDC34			-CDC34(T+2)-> +CDC34(T+5)	
SWI5	*+SWI5(T) -> -CLN1(T+1) [KEGG:ko04110]* -SWI5(T) -> +CLB6(T+1) [KEGG: ko04110]	+SWI5(T+1)-> -CLN1(T+2) [KEGG:ko04110]		*+SWI5(T+3)-> -CLN1(T+4) [KEGG:ko04110]*
SKP1	-SKP1(T) -> -SWI4(T+1)			-SKP1(T+3)-> -MBP1(T+4)

Table 2. Comparison of DRN Builder and others

	KEGG	ASIAN	DRN Builder
Source data	KEGG PATHWAY database	gene expression data	gene expression data
# Regulations confirmed/found	33	18/43	33/40
Node representation	compound	gene cluster	gene
Graph type/ Edge label	directed graph/ EC number	undirected graph/ correlation coefficient	directed graph/ regulator, regulation type, time
Visualization	static	dynamic	dynamic

Table 1 shows the regulations for the first four time points only, but there are two more regulations found by DRN Builder actually finds at time points T+4 and T+5: –CLB6(T+4)->+CDC28(T+5) [13] and –CDC20(T+4)->-CLB6(T+5) [14].

Yeast cells replicate and divide their genetic material in two distinct but coordinated processes [7]. Many of the gene regulations found by DRN Builder are consistent with the phase characteristics of the cell cycle: G1, S, G2, M, M/G1 (see Fig. 1). For example, Schwob et al. [15] found that CLB6 promotes progression of cells into the S phase and expresses periodically throughout the cell cycle. CLB1 and CLB2 both promote cell cycle progression into mitosis and their transcripts accumulate during G2 and M. These processes can be characterized by two regulations, +CLB6->-CLB1 and +CLB6->-CLB2, which are included in the regulation list found by DRN Builder (Table 1).

Fig. 5 shows the gene regulatory network of all the gene regulations of Table 1. DRN Builder provides a time bar to browse the changes in the network at different time points, and can show a network at a single time point.

There are a few methods developed to analyze the time-series gene expression data [16, 17, 18]. For comparative purposes, we executed DRN Builder and ASIAN [18] on the same data of yeast cell cycling with 19 genes, and compared the results with the KEGG data. With default parameter values, ASIAN found 43 correlations between gene clusters, and 18 of them (cln1->clb5, clb6, cln2->clb5, clb6, cln2->swi6, clb5, clb6 -> cln3, clb5, clb6 -> sic1, swi5 -> clb1, clb2, clb1, clb2 -> skp1, clb1, clb2 -> sic1, swi6 -> skp1, cdc53 -> sic1, cdc34 -> sic1) were found in KEGG and literatures. As we discussed earlier, DRN Builder found 40 gene regulations, and 33 of them agree with the KEGG data. DRN Builder visualizes a gene regulatory network as a directed graph, in which a node represents a gene and an edge represents a regulation type and time point. Therefore, the network clearly shows what genes activate or inhibit other genes and when the regulations occur. On the other hand, ASIAN visualizes a gene regulatory network as undirected graph, in which a node represents a gene cluster and an edge represents a non-zero correlation coefficient between the clusters. Thus, the gene regulatory network visualized by ASIAN cannot show what genes activate or inhibit other genes and when the regulations occur, but shows the correlation relationships between gene clusters.

To identify regulations between ARR genes, the ARR gene expression data shown in Fig. 4 was first quantified by ImageQuant. By applying the same strategy, the gene regulations in the ARR genes group were also identified by DRN Builder. The threshold to distinguish the up or down regulate is set as the ratio of the volume of
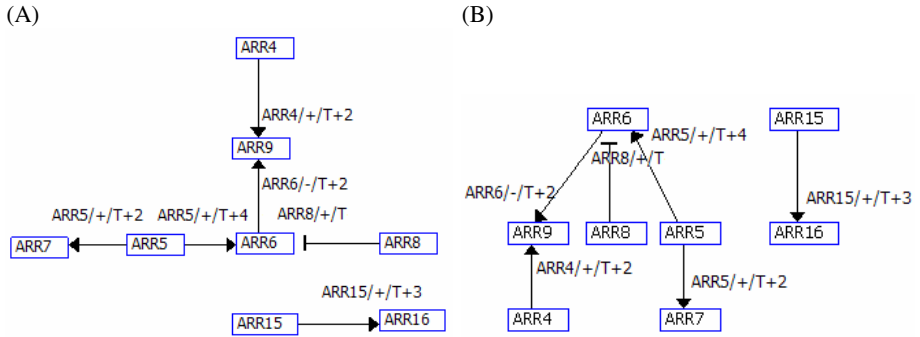
Fig. 6. Gene regulatory network for the ARR gene regulations. (A) Grid layout by DRN Builder. (B) Multi-layered layout in the DRN Builder.

gene expression to the volume of *β-tubulin* at a target time. A regulation is up-regulated if the ratio exceeds 1.0, down-regulated otherwise. Fig. 6 shows a network for the gene regulations identified by DRN Builder.

4   Conclusion

We developed an algorithm for identifying and visualizing the gene regulatory network in the time-series gene expression profiles and implemented the algorithm in a program called DRN Builder. We tested DRN Builder on two datasets, the yeast cell cycling data and Arabidopsis response regulator data. Experimental results show that DRN Builder is powerful for finding gene regulations from time-series data of gene expression profiles as well as for dynamically visualizing gene regulatory networks. Many regulations found by DRN Builder are consistent with the regulations determined by experimental methods. In particular, the potential gene regulatory networks constructed by DRN Builder can be used to find novel gene regulatory networks or to refine known networks.

Acknowledgments

References

1. Slivescu, A., Honavar, V.: Temporal Boolean Network Models of Genetic Networks and Their Inference from Gene Expression Time Series. Complex Systems 13, 54–70 (2001)
2. D'Haeseleer, P., Wen, X., Fuhrman, S., Somogyi, R.: Linear Modeling of mRNA Expression Levels during CNS Development and Injury. In: Pac. Symp. Biocomput., pp. 41–52 (1999)

3. Yeang, C.H., Jaakkola, T.: Time Series Analysis of Gene Expression and Location Data. In: Thrid IEEE Symposium on Bioinformatics and BioEngineering (BIBE 2003), pp. 305–312. Institute of Electrical and Electronics Engineers, Inc., Bethesda (2003)

4. Shmulevich, I., Dougherty, E.R., Kim, S., Zhang, W.: Probabilistic Boolean Networks: a Rule-based Uncertainty Model for Gene Regulatory Networks. Bioinformatics 18, 261–274 (2002)

5. Hartemink, A.J., Gifford, D.K., Jaakkola, T.S., Young, R.A.: Using Graphical Models and Genomic Expression Data to Statistically Validate Models of Genetic Regulatory Networks. In: Pac. Symp. Biocomput., pp. 422–433 (2001)

6. Pomerening, J.R., Kim, S.Y., Ferrell, J.E.J.: Systems-level Dissection of the Cell-cycle Oscillator: Bypassing Positive Feedback Produces Damped Oscillations. Cell 122, 565–578 (2005)

7. Spellman, P.T., Sherlock, G., Zhang, M.Q., Iyer, V.R., Anders, K., Eisen, M.B., Brown, P.O., Bostein, D., Futcher, B.: Comprehensive Identification of Cell Cycle-regulated Genes of the Yeast Saccharomyces cerevisiae by Microarray Hybridization. Molecular Biology of the Cell 9, 3273–3297 (1998)

8. Rodgers, J.L., Nicewander, W.A.: Thirteen Ways to Look at the Correlation Coefficient. The American Statistician 42, 2 (1988)

9. Danielsson, P.E.: Euclidean distance mapping. Computer Graphics and Image Processing (1980)

10. Spearman, C.: The proof and Measurement of Association between Two Things. American Journal of Psychology (1904)

11. Sugiyama, K.: Graph Drawing and Applications for Software and Knowledge Engineers. World Scientific, Singapore (2002)

12. D' Agostino, I.B., Deruere, J., Kieber, J.J.: Characterization of the Response of the Arabidopsis Response Regulator Gene Family to Cytokinin. Plant Physiology 124, 1706–1717 (2000)

13. Kuhne, C., Linder, P.: A New Pair of B-type Cyclins from Saccharomyces Cerevisiae that Function early in the Cell Cycle. Embo. J. 12(9), 3437–3447 (1993)

14. Alexandru, G., Zachariae, W., Schleiffer, A., Nasmyth, K.: Sister Chromatid Separation and Chromosome Re-duplication are Regulated by Different Mechanisms in Response to Spindle Damage. Embo. J. 18(10), 2707–2721 (1999)

15. Schwob, E., Nasmyth, K.: CLB5 and CLB6, A New Pair of B Cyclins Involved in DNA Replication in Saccharomyces Cerevisiae. Genes Dev., 1160–1175 (1993)

16. Yeung, M.K.S., et al.: Reverse Engineering Gene Networks Using Singular Value Decomposition and Robust Networks. Proc. Natl. Acad. Sci., U.S.A 99, 61–63 (2002)

17. Wang, Y., et al.: Inferring Gene Regulatory Networks from Multiple Microarray Datasets. Bioinformatics 22, 2413–2420 (2006)

18. Aburatani, S., et al.: ASIAN: A Web Server for Inferring a Regulatory Network Framework from Gene Expression Profiles. Nucleic Acids Res. 33, W659–W664 (2005)

On-Line Signature Verification Based on Spatio-Temporal Correlation

Hao-Ran Deng and Yun-Hong Wang

Intelligent Recognition and Image Processing Laboratory, School of Computer
Science and Engineering, Beihang University, 100083 Beijing, China
`dhr@cse.buaa.edu.cn, yhwang@buaa.edu.cn`

Abstract. In this paper, a novel signature verification algorithm based
on spatio-temporal correlation and an improved template selection
method are proposed. The proposed algorithm computes the spatio-
temporal matrices of signatures and considers them as images, then ap-
plies image processing methods to obtain features. The spatio-temporal
matrices will enlarge the variations of the signatures, so an improved
template selection method considering the intra-class variation of enroll-
ment signatures is also adopted. The algorithms are validated on the
SVC 2004 database and inspiring results are obtained.

Keywords: On-Line Signature, Verification, Biometrics, Spatio-
Temporal, Correlation, SIFT, Multiple Templates.

1 Introduction

Signature verification is one of the widely accepted approaches confirming indi-
vidual's identity. It has been used to validate cheques, financial and legal docu-
ments, and many other applications. On-line signature verification uses dynamic
information of signatures, such as speed, pressure and so on, which are ob-
tained by some special instrument during signing process, and is more robust
than off-line signature verification. Sice 1960s, dedicated efforts of a great many
researchers have been spend on this filed and numerous methods have been pro-
posed [1], [2], [3], [4], [5], such as Dynamic Time-Warping (DTW) [6], Hidden
Markov Models (HMM) [7] and so forth.

Signatures of each person contain some habitual features that are as peculiar
as one's habits, so they can be used as a biometric for identity verification. An
on-line signature can be considered as a sequence of sample points and each
sample point contains some dynamic information. So, there are some kind of
spatio-temporal inner-relationship between the sample points, which contain the
habitual features for signature verification. This paper propose a novel signature
verification algorithm extracting and making use of these relationships. The
algorithm computes the spatio-temporal correlation matrix from a signature and
considers it as an intensity image, then uses image processing methods to extract
features. In this paper scale invariant feature transform (SIFT) is adopted to
obtain the features, which is an excellent algorithm to detect and describe local

features in images and is used in many applications, such as object recognition, gesture recognition, and video tracking.

The signing process is effected by the signer's mood, physical condition, muscular coordination and other external influences such as writing conditions, drugs, alcohol, nervousness. Though signatures have enough consistency for personal identification, there are some natural variations in the signatures of a person. The signatures of some persons may have a large intra-class variability for the unstability of signatures. And the variation of the signatures will be enlarged in the spatio-temporal matrix, so matching the test signature with single template may result in an inferior performance. To improve the performance, an improved template selection method considering the intra-class variations is proposed.

The main contribution of this paper is proposing a novel approach of signature verification, which considering the spatio-temporal relationship of the sample points. This paper also proposed an improved template selection method considering the variations of the enrollment signatures.

This paper is organized as follows. The spatio-temporal correlation is introduced in section 2. SIFT algorithm is described in section 3. Template selection and matching are presented in section 4. The experiments and results are given in section 5. Section 6 concludes the paper.

2   Spatio-Temporal Correlation

As discussed above, an on-line signature contains a sequence of sample points and each sample point has some dynamic features, such as x-coordinate, y-coordinate, sample time, speed, pressure, and so on. During the signing process, the trajectory of the pen-tip locate on a two-dimensional writing plane, and the state of pen-tip, such as coordinates, pressure, azimuth, speed, can be considered as functions of time. The sample points are discrete points picked from the trajectory, so it's easy to understand that there are some spatio-temporal relationships between the sample points, which make the signatures of a individual unique.

To obtain the spatio-temporal relationships, the correlations between the sample points of a signature are computed and result in a correlation matrix R. The matrix R depicts the spatio-temporal relationships of the sample points of a signature and variations between the correlation matrices reflect the variations between the signatures, which are also enlarged in the matrices. As a result, the task of signatures verification become the verification of correlation matrices.

It is common knowledge that an intensity image can be represented as a matrix. There are a lot of biometric verification methods using iris, fingerprint or face images. So, the correlation matrices can be though of as intensity images and the variations of the matrices can be considered as the variation of the texture of corresponding images. Then image processing methods can be used to extract the texture features.

There are dozens of dynamic features in signature verification literature. In this paper, 11 dynamic features are selected [8], [9], which is shown in Table 1.

Table 1. Dynamic features used in this paper

Feature name	Formula
Horizontal position	x_t
Vertical position	y_t
Normal pressure	p_t
Path tangent angle	$\theta_t = \arctan(\dot{x}_t / \dot{y}_t)$
Total velocity	$v_t = \sqrt{\dot{x}_t^2 + \dot{y}_t^2}$
x velocity	$v_x = \dot{x}_t$
y velocity	$v_y = \dot{y}_t$
Total acceleration	$a = \sqrt{\dot{v}_t^2 + (v_t \cdot \dot{\theta}_t)^2}$
Log radius of curvature	$\rho = \log(v_t / \dot{\theta}_t)$
Pen azimuth	z_t
Pen altitude	l_t

So, each sample point of signatures contains these dynamic information. Formula (1) is adopted to obtain the correlation of any two sample points. f_i denotes the i-th dynamic feature of sample point. The dynamic features of signatures have been smoothed and normalized during preprocess stage. And ω_i represents the weight for the i-th feature. Each entry of R is in the range of [0,1].

$$r_{ij} = \frac{n \sum_{k=1}^{11} \omega_k^2 f_i^k f_j^k - \sum_{k=1}^{11} \omega_k f_i^k \sum_{k=1}^{11} \omega_k f_j^k}{S_i S_j} \tag{1}$$

$$S_i = \sqrt{n \sum_{k=1}^{11} (\omega_k f_i^k)^2 - (\sum_{k=1}^{11} \omega_k f_i^k)^2} \tag{2}$$

Just as shown in Figure 1, 1,3 rows are the images of the original signatures and 2,4 rows are the corresponding correlation images. 1-3 column images are generated from the genuine signatures and the rest is from skilled forgery signatures. From the images, it can be easily found that the genuine signature images of different users are very different, and the genuine signature images and skilled forgery images also have different texture. It can be conclude that the habitual features of the signature are represented as the local texture of the image of correlation image.

3   Scale Invariant Feature Transform

Scale invariant feature transform is an algorithm to detect and describe local features in images. The SIFT features are local and based on some particular interest points. The features are invariant to image scale and rotation.

Features are efficiently detected through a staged filtering approach that identifies stable points in scale space. This approach transforms an image into a large
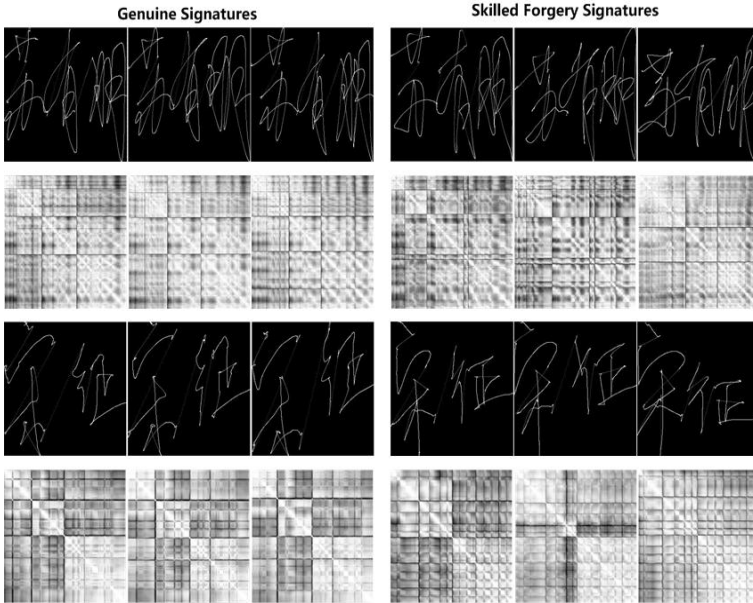
Fig. 1. 1,3 row are the images of the original signatures and 2,4 are corresponding correlation images. 1-3 column images are generated from genuine signatures and the rest is from skilled forgery signatures.

collection of local feature vectors called SIFT keys, each of which is invariant to image translation, scaling, and rotation [11].

The advantages of SIFT discussed above make it suitable for processing the correlation images of the signatures. We apply the SIFT algorithm to extract the local feature vectors and match these vectors with the ones obtained from other signatures. A matched vectors are called a pair of match point, so the number of the match points can estimate the similarity of two signatures.

The match points between signatures obtained by SIFT algorithm are shown in Figure 2. The first image of each row shows the match points between two genuine signature and the second image shows the match points between a genuine and a forgery image. It can be seen that genuine signatures have more match points, which can be used as a feature to classify genuine signature and forgery signatures.

4   Template Selection and Matching

Just as discussed above, the spatio-temporal matrix will enlarge the variations of the signatures. Subtle change of the signature will result in a great change in the texture of the corresponding correlation image. And there may be large intra-class variation in enrollment signatures, so a single template can not represent the signatures signed by an individual very well and then multiple templates for matching may promise better performance. Each template represent a sub-class of the signatures.
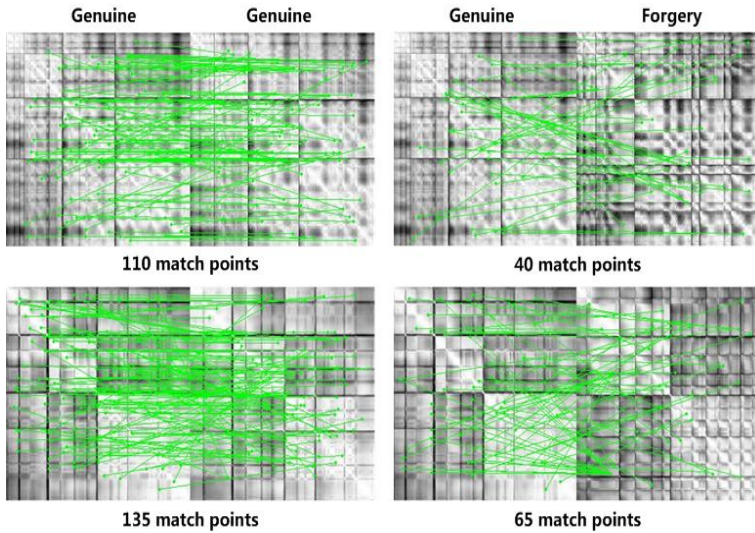
Fig. 2. The images in the first column are the match results of two genuine signatures and the images in second is the match results of a genuine signature and a skilled forgery signature

There are some popular biometric template selection methods [12], such as DEND, MDIST. In the enrollment stage of a signature verification system, only 3-8 enrollment signatures can be obtained from users. So, it's not realistic to adopt clustering algorithms like DEND to select the template signature. As for the MDIST method, it selects the template signatures using the following rules [12]:

1. Find the pair-wise similarity between the n enrollment signatures.
2. For the j-th signature, compute its average similarity with respect to other $(n-1)$ signatures.
3. Choose k signatures which have the maximum average similarity as templates.

The templates selected by MDIST have high similarity with each other. These templates represent the most common variations of the enrollment signatures instead of as many variations as possible.

In this paper, an improved template selection algorithm is proposed, which selects templates to represent as many variations as possible and achieve better performance. The selection rules:

1. Find the pair-wise similarity between the n enrollment signatures. So a similarity matrix M is obtained.
2. For every enrollment signature, compute its average similarity with respect to other $(n-1)$ signatures, that means computing the average value of each row of the matrix M.

3. Find the index of row which has maximum average value, select the corresponding signature as a template, delete the row and column contain the signature and results in a new matrix M. Repeat this step until templates is enough.

During the matching stage, the test signature is matched with each template, and each template will judge whether the test signature is genuine comparing the matching result with a certain threshold. Then the judgements of the templates are used to vote for the final decision by majority decision rule.

5   Experimental Results

SVC2004 database [10] is one of the widely used signature databases. The database is designed for Signature Verification Competition in 2004. It contains 40 sets of signatures collected from different people, and each set consists of 20 genuine signatures and 20 skilled forgery signatures. The database contains both English and Chinese signatures.

During the training process, we select 3 templates from 5 enrollment signatures, which are randomly picked from the genuine signatures of the database. The rest genuine signatures and skilled forgery signatures of the database are used as test samples. The user-dependent threshold, a threshold for a individual user, is used to make decisions.

In order to get the intuitive impression of the performance of the proposed method, we also repeat the experiment replacing the template selection with MDIST algorithm. Finally, the Equal Error Rates (ERR) of MDIST is observed to be 3.5%, and the result of using the proposed template selection method the ERR is 2.4%.

6   Conclusions

This paper has presented a novel signature verification algorithm which considers the spatio-temporal relationships of the sample points of signatures and adopts a multiple template selection method considering the intra-class variation of enrollment signatures to improve the performance.

In the experiments, we can find that the proposed template selection method achieve better performance than MDIST method, and also the performance of the presented signature verification algorithm achieve state-of-the-art performance.

References

1. Plamondon, R., Srihari, S.N.: On-Line and Off-Line Handwriting Recognition: A Comprehensive Survey. J. PAMI 22, 63–84 (2000)
2. Dimauro, G., Impedovo, S., Lucchese, M.G., Modugno, R., Pirlo, G.: Recent Advancements in Automatic Signature Verification. In: 9th International Workshop on Frontiers in Handwriting Recognition, pp. 179–184. IEEE Press, New York (2004)

3. Impedovo, D., Pirlo, G.: Automatic Signature Verification: The State of the Art. J. SMC 38, 609–635 (2008)

4. Jain, A.K., Griess, F.D., Connell, S.D.: On-line Signature Verification. J. PR 35, 2963–2972 (2002)

5. Tappert, C.C., Suen, C.Y., Wakahara, T.: The State of the Art in On-Line Handwriting Recognition. J. PAMI 12, 787–808 (1990)

6. Martens, R., Claesen, L.: On-Line Signature Verification by Dynamic Time-Warping. In: 13th International Conference on Pattern Recognition, pp. 38–42. IEEE Press, New York (1996)

7. Ignarza, J.J., Goirizelaia, I., Espinosa, K., Hernáez, I., Méndez, R., Sánchez, J.: Online handwritten signature verification using hidden markov models. In: Sanfeliu, A., Ruiz-Shulcloper, J. (eds.) CIARP 2003. LNCS, vol. 2905, pp. 391–399. Springer, Heidelberg (2003)

8. Richiardi, J., Ketabdar, H.: Local and Global Features Selection for On-line Signature Verification. In: 8th International Conference on Document Analysis and Recognition, pp. 625–629. IEEE Press, New York (2005)

9. Fierrez, J., Garcia, J.O., Ramos, D., Rodriguez, J.G.: HMM-based on-line signature verification: Feature extraction and signature modeling. J. PR Letters 28, 2325–2334 (2007)

10. Yeung, D.Y., Xiong, Y., George, S., Kashi, R.: SVC2004: First international signature verification competition. In: Zhang, D., Jain, A.K. (eds.) ICBA 2004. LNCS, vol. 3072, pp. 16–22. Springer, Heidelberg (2004)

11. Lowe, D.G.: Object Recognition from Local Scale-Invariant Features. In: 7th International Conference on Computer Vision, pp. 1150–1157. IEEE Press, New York (1999)

12. Jain, A., Uludag, U., Ross, A.: Biometric Template Selection and Update: A Case Study in Fingerprints. In: Kittler, J., Nixon, M.S. (eds.) AVBPA 2003. LNCS, vol. 2688, pp. 1061–1078. Springer, Heidelberg (2003)

Recovering Facial Intrinsic Images from a Single Input

Ming Shao and Yun-Hong Wang

School of Computer Science and Engineering, Beihang University, Beijing, China
`shaoming@cse.buaa.edu.cn`, `yhwang@buaa.edu.cn`

Abstract. According to Barrow and Tenenbaum's theory, an image can be decomposed into two images: a reflectance image and an illumination image. This midlevel description of images attracts more and more attentions recently owing to its application in computer vision, i.e. facial image processing and face recognition. However, due to its ill-posed characteristics, this decomposition remains difficult. In this paper, we concentrate on a slightly easier problem: given a simple frontal facial image and a learned near infrared image, could we recover its reflectance image? Experiments show that it is feasible and promising. Based on extensive study on hyperspectral images, skin color model and Quotient Image, we proposed a method to derive reflectance images through division operations. That is to divide visual frontal face images by learned near infrared images which are generated by super-resolution in tensor space. With the operation on grey distribution of frontal facial images, the results after division can represent the reflectance of skin, rarely bearing any illumination information. Experimental results show that our method is reasonable and promising in image synthesis, processing and face recognition.

Keywords: Intrinsic images, near infrared, super-resolution, Tensorfaces, Multi-Spectral Quotient Image.

1 Introduction

Lighting variation has been proved to be difficult problems in vision applications since variations of images under different illumination are even larger than images variations due to identity changes, especially in face recognition. Recently, a few methods have been proposed to deal with the problem generated by illumination variations: Illumination Cone [1], Quotient Image [2], Spherical Harmonic Subspace [3], [4], Morphable faces [5], Tensorfaces [6], Intrinsic Images [7], [8] and subspace model-based approach using BRDF [9]. All these methods can be clearly categorized based on whether 3-D shape model or the Lambertian assumption is used or not. Most of these methods are focusing on face relighting as well as face recognition and the factorization of a single or multiple images into lighting and reflectance components is a commonly used approach. Barrow and Tenenbaum (1978) introduced the term "intrinsic images" to refer to a midlevel decomposition of the sort depicted in Fig. 1. The observed image is the product of two images: an illumination image and a reflectance image. Because of a shortness of a full 3D description of the scene, we call this a midlevel description: the intrinsic images are viewpoint dependent and the physical causes of changes in illumination at different point are not made explicit. Owing to

Input image   Reflectance   Illumination
                        image   image

Fig. 1. The intrinsic images decomposition. The input image, from Yale B database, is decomposed into two parts, a reflectance image and an illumination image which are obtained according to [8].

this characteristic, the intrinsic images can be extremely useful for supporting a range of visual inferences, i.e. image segmentation and face recognition.

Recovering two intrinsic images from a single input image remains a difficult problem for computer vision systems for it is a classic ill-posed problem: the number of unknowns is twice the number of equations. In [8], Weiss focused on a slightly easier version of the problem. Given a sequence of T images whose reflectance images are constant over time and illumination images change, the decomposition could be derived. Nonetheless, this method will fail when few or only one image is available because the maximum-likelihood estimation for approaching this problem is not valid anymore. Other 3-D shape model based algorithms, for example, Morphable face, necessitates a 3-D training face database, leading to a relatively complex problem compared with 2-D condition.

The major contribution of this paper is to derive facial intrinsic images from single frontal face image under visible light. The reflectance image generated by our method is the quotient image of visual image (VIS) and its corresponding near infrared image (NIR) and this is what we call "Multi-Spectral Quotient Image". For the purpose of practice and simplicity, we use the learned NIR image which is generated by super-resolution in tensor space instead of ground truth NIR image.

The paper is organized as follows. In section 2, we introduce the multi-spectral properties and make an explanation of what we call "Multi-Spectral Quotient Image". Section 3 sets up two training tensors as the basic multilinear analysis tool, which include factors of identity, pixels as well as spectral and makes use of super-resolution to generate corresponding NIR image of the input. Section 4 discusses the experimental database and results before conclusions are drawn in section 5.

2   Multi-Spectral Quotient Image

Spectral measurements from human tissue have been used for many years in characterization and monitoring applications. The interaction of light with human tissue has been studied extensively [10], [11], [12] and a model for skin color in multi-spectral has been built [13]. In this paper, we only focus on two typical spectral, point visual light (VIS) source of 520nm and point near infrared (NIR) light source of 850nm in that their appearances are fairly distinct and each of them includes different facial features of the same subject, as depicted in Fig. 2.

Fig. 2. An example of NIR and VIS image pairs. The first row is a group of five NIR facial images under frontal illumination and the second row is the counterpart of VIS images which are taken simultaneously using a two CCD camera.

There are two reasons for this phenomenon. First, due to different reflectance between NIR and VIS on skin, eyes and lips, the intensity (I) of images is quite different, which indicates that the reflectance varies with spectral changes. This sort of changes regarding spectral has been explicitly revealed in [14]. Another reason is that under NIR, skin has a larger penetration depth than for visible wavelengths enabling imaging the subsurface characteristics. Comparing with the skin's surface characteristics included in VIS spectrum, the subsurface images obtained from NIR spectrum are relatively smooth, indicating that the subsurface parts share a more unified reflectance. These unique properties of multi-spectral images enlighten us to derive a novel but simple method to carry out intrinsic image decomposition in the following parts.

According to Lambertian assumption, when attached shadows are concerned, VIS and NIR images reflectance functions can be expressed as:

$$I_{VIS}(x, y) = R_{VIS}(x, y) \cdot \max(n_{VIS}(x, y) \bullet l_{VIS}, 0) , \tag{1}$$

$$I_{NIR}(x, y) = R_{NIR}(x, y) \cdot \max(n_{NIR}(x, y) \bullet l_{NIR}, 0) , \tag{2}$$

where R denotes the albedo, n normal direction, l a distant light source and " $\bullet$ " denotes dot product. Specifically, subscripts of parameters indicate that they pertain to NIR and VIS images respectively. In this paper, a pair of multi-spectral images is captured simultaneously with roughly the same light intensity and direction, which guarantees that the parameter l_{VIS} is close to l_{NIR} . Besides, since the images of one subject are taken at the same time, the shape n_{VIS} and n_{NIR} are exactly the same, leading to the following deduction of Multi-Spectral Quotient Image (MQI):

$$\text{MQI} = \frac{I_{VIS}(x, y)}{I_{NIR}(x, y)} = \frac{R_{VIS}(x, y)}{R_{NIR}(x, y)} \cdot \alpha , \tag{3}$$

where α is a constant, denoting the ratio of the intensity of l_{VIS} to that of l_{NIR} for they are not identical with each other, though very close. From [14] we can clearly learn that for a fixed subject, the VIS reflectance has some unknown relationship with NIR reflectance. If $R_{VIS}(x, y)$ is the square of the $R_{NIR}(x, y)$, the $R_{NIR}(x, y)$ can be obtained through dividing $R_{VIS}(x, y)$ by $R_{NIR}(x, y)$. Consequently, we make a nonlinear operation on $I_{VIS}(x, y)$, for example, $I_{VIS}(x, y)^{1.5}$ to achieve requirements mentioned above. Fig. 3 shows some MQI examples.

Fig. 3. Some examples of the Multi-Spectral Quotient Image. The first two columns are four illumination images of two subjects, while the middle column is their corresponding reflectance images gained by MQI. The last two columns are their illumination images.

Another way to understand the Multi-Spectral Quotient Image is considering the NIR image to be a "pseudo illumination image", which has some uncertain relationship with true illumination image. That means through some mathematical operation, a NIR image can be transformed into an illumination image. Meanwhile, the approximate appearance of these two sorts of images justifies our assumption, for both of them possess smooth surfaces which only reflect the shape and lights information on the face. Skin under NIR, as we have introduced, has a larger penetration depth than for visible wavelengths and we employ tissue thickness that reduces the light intensity to 37 percent of the intensity at the surface to describe this parameter. The optical penetration depth is defined as $1/\sqrt{3\mu_a\mu_s'}$ where μ_a and μ_s' are the absorption coefficient and reduced scattering coefficient of the tissue, respectively. For a typical person, we have μ_a =0.77mm^{-1} and μ_s' =1.89 mm^{-1} in the visual light and μ_a =0.02mm^{-1} and μ_s' =1.31 mm^{-1} in the near infrared, which means a 0.479mm and a 3.567mm penetration depth of facial skin respectively. At the latter depth, the tissue is smoother than the surface part, giving rise to a better expression of facial shape and illumination information.

3   Simulating Near Infrared Images in Tensor Space

In reality, gaining a pair of NIR and VIS images is relatively difficult than gaining only one image under normal illumination for NIR collecting device including active light source is not commonly used data acquisition equipment. Even though research on NIR images, i.e. NIR face recognition attracts more attentions recently, simultaneously obtaining two images in different spectral remains inconvenient. Therefore, a natural thought is to simulate NIR images other than employ ground truth NIR images to derive intrinsic images. In [15], multi-model Tensor Face is used to generate new faces under new conditions, for instance, new lighting and new pose. As a basic multi-linear algebra analysis tool, tensor can handle subjects with more than two modes, superior to commonly used principal component analysis (PCA), which only concerns pixels and identity of the subject. It is also suggested that, for a fixed subject, its downsampling image carries most of identity information. Therefore, by super-resolving a low-resolution image in tensor space using Bayesian inference, new images with high-resolution under new conditions can be rendered. Enlighten by
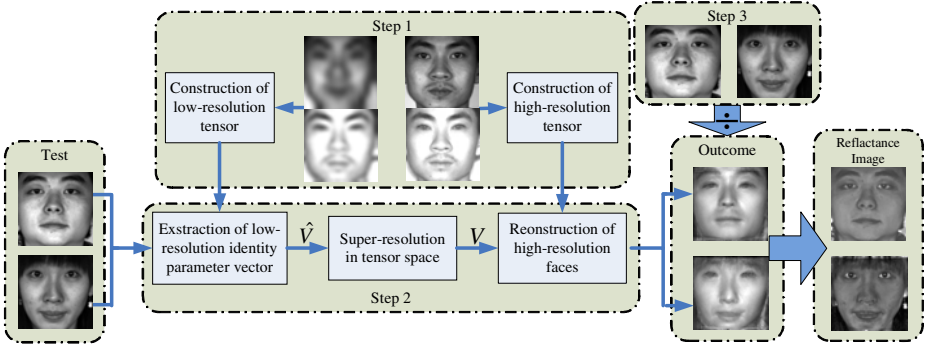
Fig. 4. The framework of our proposed method. Step (1) Building the training tensor, (2) super-resolving the low-resolution sample in order to obtain its identity parameter vector V in high-resolution tensor space and (3) the Multi-Spectral Quotient Image.

this approach, we can also simulating NIR images under this framework by constructing 3D tensors which consist of three factors, pixels, identity and spectral. Following the thought in [15], we first set up a high-resolution and a low-resolution tensors by several pairs of NIR images and their corresponding VIS images as well as their downsampling images. Then, we project downsampling tests into low-resolution tensor space to obtain its identity vector $\hat{V}$. Through super-resolution, the "true" identity vector V in high-resolution tensor space is achieved, giving rise to the simulated NIR images. Finally, the proposed MQI operation is implemented between input tests and simulated NIR images. Fig. 4 illustrates the whole procedure of the method mentioned above.

3.1 Modeling Face Images

First a tensor structure is constructed from multi-modal face images and HOSVD [6], [15] is used to decompose them. The decomposed model can be expressed as

$$D = Z \times_1 \mathbf{U}_{idens} \times_2 \mathbf{U}_{specs} \times_3 \mathbf{U}_{pixels} \,,\tag{4}$$

where tensor D groups the multi-modal face images into tensor structure, and core tensor Z governs the interactions among 3 mode factors. The mode matrix $\mathbf{U}_{idens}, \mathbf{U}_{specs}$ and $\mathbf{U}_{pixels}$ spans the parameter space of different people identities, spectral, and face images respectively. To start with, we will introduce how the pixel-domain image observation model works. Assuming D_H is the unknown true scene (the target NIR face image) and $\mathbf{A}$ is a linear operator that incorporates the motion, blurring and downsampling processes, the observation model can be expressed as

$$D_L = \mathbf{A}D_H + n \,,\tag{5}$$

where n represents the noise in these processes. Based on the tensor algebra, suppose we have a basis tensor

$$B = Z \times_2 \mathbf{U}_{specs} \times_3 \mathbf{U}_{pixels} \,.\tag{6}$$

The subtensor containing a specific sort of spectral and a particular individual can be expressed as

$$D_s = B_s \times_1 V^T + \varepsilon_s \, , \tag{7}$$

where B_s is the B's subtensor with particular spectral s, V^T represents the identity parameter row vector and ε_s stands for the tensor modeling error for modalities of spectral s. It is for the purpose of ease and readability that we use the mode-1 unfolding matrix to represent tensors. Then Eq. (7) can become

$$\mathbf{D}_s = V^T \mathbf{B}_s^{(1)} + e_s \, . \tag{8}$$

The counterpart of pixel-domain image observation model is then given as

$$\hat{\mathbf{B}}_s^{T(1)} \hat{V} + \hat{e}_s = \mathbf{A} \mathbf{B}_s^{T(1)} V + \mathbf{A} e_s + n \, , \tag{9}$$

where $\hat{\mathbf{B}}_s^{T(1)}$ and $\mathbf{B}_s^{T(1)}$ are the low-resolution and high-resolution unfolded basis subtensor, $\hat{V}$ and V are the identity parameter vectors for the low-resolution testing face image and unknown high-resolution image. Thus, through Eq. (9), we can naturally derived

$$\hat{V} = \mathbf{\Phi} \mathbf{A} \mathbf{B}^{T(1)} V + \mathbf{\Phi} \mathbf{A} E + \mathbf{\Phi} N \, , \tag{10}$$

where E is the combined tensor modeling error over all model face images and $\mathbf{\Phi}$ is the pseudoinverse of $\hat{\mathbf{B}}^{T(1)}$. Eq. (10) depicts the relation between the unknown "true" identity parameter V vector and the observed low-resolution counterpart $\hat{V}$.

3.2   A Bayesian Formulation

In order to obtain the estimation of Eq. (10), we use the Bayesian inference. The maximum *a posteriori* probability (MAP) estimation of the high-resolution identity parameter vector V can be expressed as

$$\tilde{V} = \arg\max_V \{ p(\hat{V} | V) p(V) \} \, , \tag{11}$$

where $p(\hat{V} | V)$ is the conditional probability modeling the relations between $\hat{V}$ and V, and $p(V)$ is a prior probability. We can assume the prior probability as Gaussian and Eq. (11) therefore degenerates to the maximum likelihood estimator

$$\tilde{V} = \arg\max_V \{ p(\hat{V} | V) \} \, . \tag{12}$$

Then we obtain the ML estimator $\tilde{V}$ as

$$\tilde{V} = \arg\min_V \left(\left(\hat{V} - \mathbf{\Phi} \mathbf{A} \hat{\mathbf{B}}^{T(1)} V - \mathbf{\Phi} \mu_F \right)^T \mathbf{Q}^{-1} \left(\hat{V} - \mathbf{\Phi} \mathbf{A} \hat{\mathbf{B}}^{T(1)} V - \mathbf{\Phi} \mu_F \right) \right) , \tag{13}$$

where the statistics of mean μ_F and matrix $\mathbf{Q}$ can be computed based on the training images. Using the iterative steepest descent method for ML estimation of $\tilde{V}$, we can obtain its value. For more details, readers can refer to [15]. The reflectance images
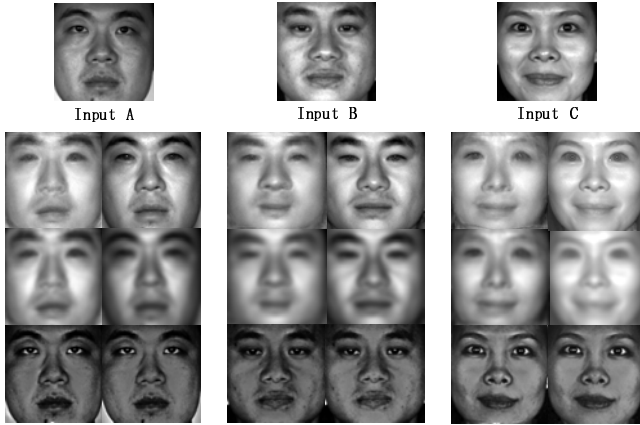
Fig. 5. The reflectance images derived by learned NIR images. For each input (Input A, B and C) the first image of left column is the generated NIR image and the second one is its Gaussian filter output. The final row of each column is the derived reflectance image. The right column of each input is the corresponding ground truth NIR image and its derived reflectance image. For the purpose of better appearance, we employ histogram equalization on the result reflectance images.

generated by learned NIR images are rendered in Fig. 5, from which we can conclude that the reflectance images derived by learned NIR images are quite similar with that produced by ground truth NIR images in both shape and details. The aim of using Gaussian filter is to eliminate the ghosting and blurring appeared in the learned NIR images due to lack of dense alignments between test and training samples.

4   Experiments and Results

The experiments aim at evaluating the proposed method of intrinsic image decomposition as well as showing the utility of this decomposition in image processing. The first part of this section is the extension of the proposed method to common objects, not faces. In Fig. 6, the first image is a cotton pillow captured under frontal visible light with obvious specularity (in the center) and shadows (around the corner). The second image is the same object under NIR illumination and little reflectance information is carried in this image due to the characteristics of NIR, left only the shape and lighting information. Therefore, after dividing the first image by the second one in Fig 6, most of specularity and shadows are removed.

The second part of this section reveals how the proposed intrinsic images decomposition method is applied in the image synthesis with an image addition sample. Given an image under arbitrary illumination, we find that a direct image addition will lead to an unreal result, for additional parts rarely carry any illumination information, i.e. specularity or shadows, inconsistent with the original image. However, if the additional part is added to the reflectance image at first, and then multiplied by the illumination image, the result seems fairly true, as Fig. 7 depicted.

Fig. 6. The application of the proposed intrinsic image decomposition in object image processing. VIS image (left), NIR image (middle) and reflectance image (right).

Fig. 7. Examples of image synthesis. The bend-aid image is directly added to the facial image under arbitrary illumination, resulting in an artificial synthesized image in the first row for the additional part rarely carries any illumination information. In the second row, the bend-aid is firstly added to the reflectance image, and then multiplied by the illumination image. Group A adopts the ground truth NIR image directly for MQI while Group B is the outcome by learned NIR image.

It is suggested that reflectance images contain much more intrinsic information that is helpful to face recognition and discrimination. To evaluate this performance of our method, a database of facial images including 70 subjects has been collected under indoor environment, all of which are captured with JAI AD080 digital camera, which can take NIR and VIS images at the same time. Each subject possesses 9 changes, normal, tilt -/+15~30°, yaw -/+15~30°, happiness, anger, frustration and surprise, under NIR and VIS respectively, therefore 18 images in all, as Fig 8 depicted.

When dealing with the learned NIR images, we separate the whole database into two parts, one for the purpose of training, the other for testing. Here we use leave-one-out scheme, i.e. each image as test in turn, others for training. In training process, we build a high-resolution tensor with 69 subjects and then, by registering all training images with the test one and downsampling them, we also acquire the low-resolution tensor.

Two experiments are carried out, one aimed at testing the within-class performance, the other aimed at evaluating between-class performance. For the first experiment, comparisons among four methods proceed by calculating matching scores which

Fig. 8. VIS examples of our experimental database

equals to $x \bullet y /(\|x\| \|y\|)$ when two vectors x, y undergoing matching procedure. Set 1-Set 9 in our experiments are corresponding to 9 changes of the fixed subject. Note that we only use normal set (Set 1) as the template to implement the matching procedure and results are shown in Table 1. From this table, we can see that MQI(L) and MQI(G) which denote MQI using learned NIR and ground truth NIR respectively, can improve matching scores for the impact of illumination is removed and only intrinsic components are retained. Self-Quotient Image (SQI) [16] is an effective approach on tackling illumination changes and it achieves the highest scores among four methods.

Table 1. Comparisions of matching scores among four methods

Methods	Matching Scores (%)							
	Set 2	Set 3	Set 4	Set 5	Set 6	Set 7	Set 8	Set 9
Appearance	96.4	97.0	97.4	98.2	96.8	97.2	97.6	97.2
MQI(L)	98.0	97.9	98.1	98.5	97.3	97.8	97.9	98.0
MQI(G)	98.2	98.1	98.4	98.8	98.3	98.3	98.4	98.2
SQI	99.9	99.9	99.9	99.9	99.9	99.9	99.9	99.9

Table 2. Comparisons of recognition rates among four methods

Methods	Recognition Rates (%)							
	Set 2	Set 3	Set 4	Set 5	Set 6	Set 7	Set 8	Set 9
PCA	71.4	75.7	77.1	87.1	75.7	64.3	74.3	68.6
MQI(L)+PCA	84.3	92.9	92.9	94.3	87.1	88.6	90.0	87.1
MQI(G)+PCA	94.3	98.6	88.6	94.3	85.7	85.7	92.9	91.4
SQI+PCA	65.7	71.4	65.7	80.0	65.7	45.7	57.1	58.6

For face recognition experiments, we also use normal image as template, others for testing and nearest neighbors (NN) method functions as classifier. In this experiment, MQI (G) outperforms other methods and SQI obtains the lowest recognition rates, even lower than PCA method. Essentially, SQI can actually reduce the lighting effect on face, but identity information is also weakened in this process. With the enhancement of within-class matching score, the between-class similarity increases too. When poses, expressions and other changes modes are involved, SQI's performance degenerates greatly. As to the proposed MQI, the intrinsic components are retained as many as possible, which are robust to poses and expressions changes.

5   Conclusions

In this paper, a method to extract facial intrinsic images from single frontal face image has been proposed. Firstly, the corresponding NIR image of the input VIS image can be generated through super-resolution in tensor space. Then, the reflectance image of a subject is simply recovered through dividing its frontal VIS by its learned NIR image. A few image synthesis, processing and recognition experiments clearly justify our method, which seems to be very promising in these areas. Besides, future work including applying the proposed method to image relighting will be implemented in order to evaluate its performance under various conditions.

Acknowledgements

References

1. Georghiades, A.S., Belhumeur, P.N., Kriegman, D.J.: From Few to Many: Illumination Cone Models for Face Recognition Under Variable Lighting and Pose. IEEE Trans. Pattern Anal. Mach. Intell. 23(6), 643–660 (2001)
2. Shashua, A., Riklin-Raviv, T.: The Quotient Image: Class-based Re-rendering and Recognition with Varying Illuminations. IEEE Trans. Pattern Anal. Mach. Intell. 23(2), 129–139 (2001)
3. Basri, R., Jacobs, D.: Lambertian Reflectance and Linear Subspaces. IEEE Trans. Pattern Anal. Mach. Intell. 25(2), 218–233 (2003)
4. Ramamoorthi, R., Hanrahan, P.: On the Relationship Between Radiance and Irradiance: Determining the Illumination from Images of a Convex Lambertian Object. JOSA A 18(10), 2448–2459 (2001)
5. Blanz, V., Vetter, T.: Morphable Model for the Synthesis of 3D Faces. In: Proc. ACM SIGGRAPH (1999)
6. Vasilescu, M.A.O., Terzopoulos, D.: Multilinear Analysis of Image Ensembles: Tensor-Faces. In: Heyden, A., Sparr, G., Nielsen, M., Johansen, P. (eds.) ECCV 2002. LNCS, vol. 2350, pp. 447–460. Springer, Heidelberg (2002)
7. Barrow, H.G., Tenenbaum, J.M.: Recovering Intrinsic Scene Characteristics from Images. Computer Vision System (1978)
8. Weiss, Y.: Deriving Intrinsic Images from Image Sequences. In: Proc. of IEEE ICCV (2001)
9. Shim, H., Luo, J., Chen, T.: A Subspace Model-Based Approach to Face Relighting Under Unknown Lighting and Poses. IEEE Trans. Image Process 17(8), 1331–1341 (2008)
10. Tuchin, V.: Tissue Optics: Light Scattering Methods and Instruments for Medical Diagnosis. SPIE Press, Bellingham (2000)
11. Anderson, R., Parrish, J.: The Optics of Human Skin. J. Investigative Dermatology 77(1), 13–19 (1981)
12. Gemert, M., Jacques, S., Sternborg, H., Star, W.: Skin Optics. IEEE Trans. Biomedical Eng. 36(12), 1146–1154 (1989)
13. Angelopoulou, E., Molana, R., Daniilidis, K.: Multispectral Skin Color for Modeling. In: Proc. of IEEE International Conference on Computer Vision and Pattern Recognition (2001)
14. Pan, Z., Healey, G., Prasad, M., Tromberg, B.: Face Recognition in Hyperspectral Images. IEEE Trans. Pattern Anal. Mach. Intell. 25(12), 1552–1560 (2003)
15. Jia, K., Gong, S.: Multi-modal tensor face for simultaneous super-resolution and recognition. In: Proc. IEEE Int. Conf. Computer Vision (2005)
16. Wang, H., Li, S.Z., Wang, Y.: Face Recognition Under Varying Lighting Conditions Using Self Quotient Image. In: International Conference on FGR, pp. 819–824 (2004)

Gender Recognition from Gait Using Radon Transform and Relevant Component Analysis

Lei Chen[1], Yunhong Wang[1], Yiding Wang[2], and De Zhang[1]

[1] School of Computer Science and Engineering, Beihang University, Beijing, China
[2] College of Information Engineering, North China University of Technology, Beijing
chenlei@cse.buaa.edu.cn, yhwang@cse.buaa.edu.cn

Abstract. In this paper, a new method for gender recognition via gait silhouettes is proposed. In the feature extraction process, Radon transform on all the 180 angle degrees is applied to every silhouette to construct gait templates and the initial phase of each silhouette in an entire gait cycle is also associated to the templates representing dynamic information of walking. Then the Relevant Component Analysis (RCA) algorithm is employed on the radon-transformed templates to get a maximum likelihood estimation of the within class covariance matrix. At last, the Mahalanobis distances are calculated to measure gender dissimilarity in recognition. The Nearest Neighbor (NN) classifier is adopted to determine whether a sample in the Probe Set is male or female. Experimental results in comparison to state-of-the-art methods show considerable improvement in recognition performance of our proposed algorithm.

Keywords: gender recognition, gait analysis, radon transform, relevant component analysis.

1 Introduction

Gait recognition has received close and continuous attention in the research fields of computer vision and biometric information over decades [1]. Unlike other biometrics, recognition based on human gait has several advantages related to the unobtrusiveness and the ease with which gait information can be captured.

In the study of gait analysis, gender recognition is a significant branch because of its validity, practicality and robustness. In medical study [2], Kozlowski and Cutting et al. have already examined recognizing the gender of walkers from moving light displays. Gender feature information in gait surveillance is further extracted and classified in later research. Lee and Grimson [3] propose a gait silhouette appearance representation and apply it to gender classification task in a database including 23 males and 2 females. In [4], Davis et al. present a three-mode expressive-feature model for recognizing gender from point-light displays of walking people. A later study in [5] by Yoo et al. utilizes even a larger gait database, consisting of 84 males and 16 females.

Several methods have tried to extract the important information from the silhouettes in a gait sequence by measuring quantities related to the projections of the

silhouettes in certain directions. In [6], the walking style of individuals is captured into patterns using the mathematical theory of Frieze patterns. The width of silhouette was proposed in [7] as a suitable feature for gait feature extraction and dynamic time warping (DTW) is employed. In [8], an angular transform is introduced and shown to achieve very encouraging results. N.V. Boulgouris et al. [9] form a Radon-transformed template for each cycle of a gait sequence and apply Linear Discriminant Analysis (LDA) to identify template coefficients for the purpose of gait recognition.

Relevant Component Analysis (RCA) algorithm [10] is a simple but efficient algorithm for learning a Mahalanobis metric. It is a solution for optimization problem, founded on an information theoretic basis. If dimensionality reduction is allowed within RCA, it is optimally accomplished by a version of Fisher's linear discriminant that uses constraints. Moreover, under certain Gaussian assumptions, RCA can be viewed as a Maximum Likelihood Estimation of the within class covariance matrix.

The Radon transform has its unique advantages suitable for gait recognition. Firstly, there is large variation in the angles of legs and arms axes between male and female during human walking [2]. Radon transform guarantees that much energy of the original silhouettes will appear in specific coefficients which will vary considerably through time. Secondly, gait representation and recognition with Radon transform take a low spatio-temporal cost that makes this algorithm with quite efficiency. In this paper, aiming at the target of gender recognition from gait, we make a further research on Radon transform under every angle degree ranging from 0 to 179. Section 2 shows the details about the construction of Radon-transformed templates for silhouettes in a gait cycle. Initial phase of each silhouette is associated to the template as the frequency component to represent the dynamic information of walking. The algorithm of Relevant Component Analysis (RCA) is introduced and employed in Section 3. The experiments on BUAA-IRIP Gait Database are described in details in Section 4, following with brief conclusion in Section 5.

2   Templates Construction with Radon Transform

In mathematics, Radon transform is the integral transform consisting of the integral of a function over straight lines. There are several different forms and notations, and the most popular one of definitions is:

$$R(\rho,\theta) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x,y)\delta(\rho - x\cos\theta - y\sin\theta)dxdy , \qquad (1)$$

where $R(\rho,\theta)$ is the line integral of a two-dimensional function f(x, y) along a line from $-\infty$ to ∞. The location and orientation of the line is determined by two parameters ρ and θ as depicted in Fig.1(a). In the gait analysis process, the two-dimensional function f(x, y) is the binary silhouette. Because of the discrete form of pixel intensity, a Radon transform consists in the summation of pixel intensities along
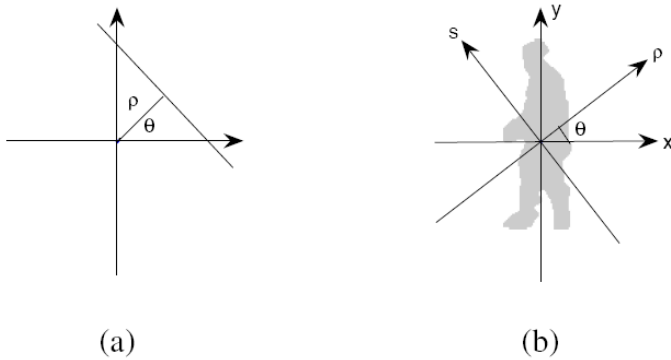
Fig. 1. The basic process of Radon transform on binary silhouette. (a) Two parameters (ρ, θ) determine the position of the integral line; (b) Calculation of Radon coefficients.

lines of different directions. The basic process of Radon transform on binary silhouette is shown in Fig.1(b).

Now we define the center of the binary silhouette as the reference point of the (ρ, θ) coordinates. So the value of ρ is equal to zero (ρ =0) in our method. Given a specific direction angle θ, the silhouette f(x, y) is projected onto the ρ axis. Therefore, pixel intensities along a set of lines parallel to s axis (Fig.1(b)) are summed together. By applying Radon Transform to silhouettes, we set up a mapping between the (x, y) coordinate system and (ρ, θ) coordinate system.
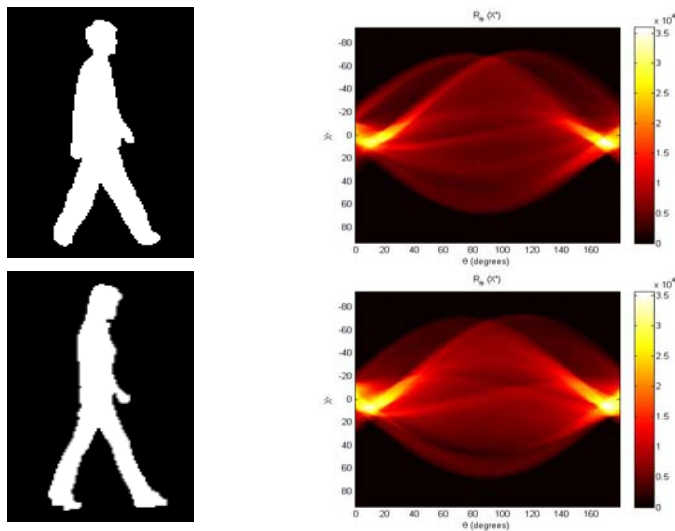
Fig. 2. Two sample silhouettes (on the left) from different genders and their corresponding radon transforms (on the right). In the right column, the transform in upper row comes from male and lower row comes from female.

An example of the Radon transform on gait silhouette is shown in Fig.2. Given a silhouette S, Radon transform on all the 180 angle degrees is applied to S, i.e., with θ ranging from 0° to 179°. Under each degree, a column vector of Radon coefficients is obtained. The 180 column vectors then make up a two-dimensional Radon-transformed coefficients matrix named as S_R. Since the sums of pixels can be seen as intensity values, the Radon transform of a silhouette can be displayed as an image. The image of S_R is shown on the right of Fig.2. As seen in directions in which the sums of pixels are greater, the corresponding Radon coefficients are greater as the same and, therefore, exhibit higher brightness. When referring to a certain direction of Radon transform, the coefficients are in a column vector. Fig.3 shows the coefficients of certain angle degree Radon transform vectors for three specific directions as $0°$, 45°and 90°.
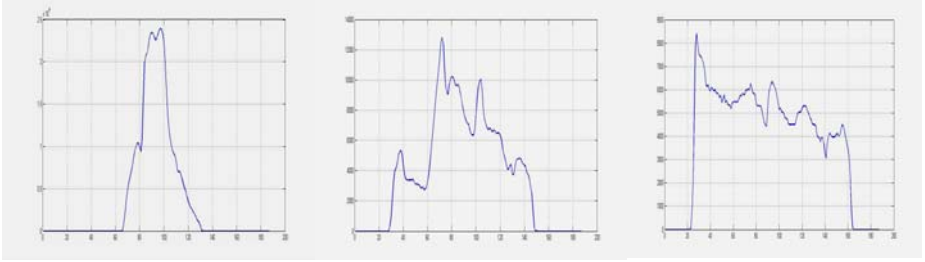
Fig. 3. Coefficients in Radon-Transformed vectors at different directions as θ = 0°, 45°, 90°from left to right. The dimension index of each vector is on X-axis and the value of coefficients is on Y-axis.

All the silhouettes in an entire gait cycle can be expressed as a single template T. To construct the template T, a simple scheme presented in [9] is employed in this paper. Let N denote the number of frames in one gait cycle, i denote the index in the range from 0 to N-1, and S_{R_i} denote the i'th Radon-transformed silhouette. The initial phase of the i'th silhouette φ_i can be defined as:

$$\varphi_i = \frac{2\pi}{4N} * i .$$ (2)

Let $\overline{S}_{R_i}$ denote the silhouette with initial phase associated to S_{R_i}, then $\overline{S}_{R_i}$ can be calculated as:

$$\overline{S}_{R_i} = S_{R_i} \cdot (1 + \cos \varphi_i + j \cdot (1 + \sin \varphi_i)) .$$ (3)

Therefore, the averaged Radon template T for a gait cycle of silhouettes is defined as the following formula:

$$T = \frac{1}{N} \sum_{i=0}^{N-1} \overline{S}_{R_i}$$

$$= \frac{1}{N} \sum_{i=0}^{N-1} S_{R_i} \cdot (1 + \cos \varphi_i + j \cdot (1 + \sin \varphi_i)) \qquad (4)$$

$$= \frac{1}{N} \sum_{i=0}^{N-1} (S_{R_i} \cdot (1 + j)) + \frac{1}{N} \sum_{i=0}^{N-1} (S_{R_i} \cdot (\cos \varphi_i + j \cdot \sin \varphi_i)))$$

According to *Euler Formula*:

$$e^{jx} = \cos x + j \cdot \sin x, x \in R. \qquad (5)$$

Finally the Radon template T can be constructed as:

$$T = \frac{1}{N} \sum_{i=0}^{N-1} (S_{R_i} \cdot (1 + j)) + \frac{1}{N} \sum_{i=0}^{N-1} (S_{R_i} \cdot e^{j \frac{2\pi}{4N} i}). \qquad (6)$$

3   Relevant Component Analysis

Relevant Component Analysis (RCA) is a method that seeks to identify and down-scale global unwanted variability within the data [10]. The method changes the feature space used for data representation, by a global linear transformation which assigns large weights to "relevant dimensions" and low weights to "irrelevant dimensions". These "relevant dimensions" are estimated using *chunklets*, that is, small subsets of points that are known to belong to the same although unknown class. The algorithm is presented below.

Given a data set $X = \{x_i\}_{i=1}^{N}$ and n chunklets $C_j = \{x_{ji}\}_{i=1}^{n_j}$, j=1…n, do:

1.   Compute the within chunklet covariance matrix

$$C = \frac{1}{N} \sum_{j=1}^{n} \sum_{i=1}^{n_j} (x_{ji} - m_j)(x_{ji} - m_j)^T, \qquad (7)$$

where m_j denotes the mean of the *j'th* chunklet.

2.   If needed, apply dimensionality reduction to the data using C.

3.   Compute the whitening transformation associated with C: $W = C^{-\frac{1}{2}}$, and apply it to the data points: $X_{new} = WX$, where X refers to the data points after dimensionality reduction when applicable. Alternatively, use the inverse of C in the Mahalanobis distance:

$$d(x_1, x_2) = \sqrt{(x_1 - x_2)^T C^{-1} (x_1 - x_2)}. \qquad (8)$$

The RCA transformation is intended to reduce clutter, so that in the new feature space, the inherent structure of the data can be more easily unraveled. To this end, the algorithm estimates the within class covariance of the data cov(X|Z) where X and Z describe the data points and their labels respectively. The estimation is based on positive equivalence constraints only, and does not use any explicit label information. In high dimensional data, the estimated matrix can be used for semi-supervised dimensionality reduction. Afterwards, the data set is whitened with respect to the estimated within class covariance matrix. The whitening transformation W assigns lower weights to directions of large variability, since this variability is mainly due to within class changes and is therefore "irrelevant" for the task of classification.

In our experiments, the dimensionality reduction by Principal Component Analysis (PCA) to matrix *C* in step 2 is carried on when using RCA on the Radon-transform templates. PCA may appear contradictory to RCA, since it eliminates principal dimensions with small variability, while RCA emphasizes principal dimensions with small variability. The principal dimensions are computed in different spaces. The dimensions eliminated by PCA have small variability in the original data space (corresponding to Cov(X)), while the dimensions emphasized by RCA have low variability in a space where each point is translated according to the centroid of its own chunklet (corresponding to Cov(X|Z)). As a result, the method ideally emphasizes those dimensions with large total variance, but small within class variance.

4   Experimental Results

According to all the above analysis, we design our experiment for gender recognition as shown in Fig.4. After data preprocessing and gait cycle detection, the sequence of silhouette in an entire gait cycle is Radon-transformed to construct a template with dynamic walking information associated by initial phase of each silhouette. After the features transform process of RCA, final feature vectors are obtained and outputted. Generally, we follow the tradition terminology in [11] that the reference database is termed Gallery Set and the test subjects are termed Probe Set. Fig.4(a) shows the basic procedure of our proposed method.

In Fig.4(b), the Mahalanobis distances between Gallery feature vectors and Probe feature vectors are calculated as the measure of dissimilarity. A Nearest Neighbor (NN) classifier is employed to determine whether the Probe samples are recognized to be right or wrong.

4.1   Database and Data Preprocessing

With further development of gait recognition, many gait databases have been created, such as UCSD database, MIT AI Gait Data, CASIA Gait Dataset, Georgia Tech Database, CMU Mobo Database, and Gait Challenge Database, etc. In these databases, even though some variances such as shoe types and carrying objects have been considered, they were mainly built for human identification other than gender classification, thus the amounts of males and females differ greatly. However, for the purpose

(a)

(b)
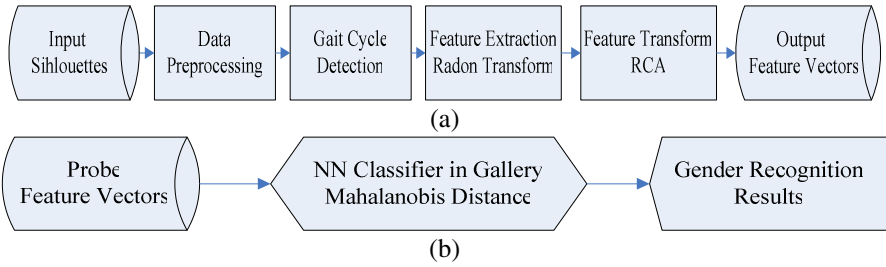
Fig. 4. Design of experiments. (a) Procedure of constructing feature vectors templates from silhouettes; (b) Procedure of gender recognition.

of gender classification, a database with nearly the same quantity of males and females is much more preferred.

For the purpose of gender recognition, we have built our BUAA-IRIP Gait Database for research and experiments. This is a gait database from Intelligent Recognition and Image Processing Laboratory of Beihang University. Totally there are 60 persons in this database, including 32 males and 28 females. Each person has five walking sequences, and several gait periods during every walking sequence. We select the sequences from side view in our experiment. The videos in this database are all captured in laboratory scenario indoor.

In the data preprocessing, we employ background subtraction method and then morphological filtering to reduce noises after binarization to get better quality of gait silhouettes. To reduce the noises in those images, erosion and dilation are used to erase the small spots on binarilized image and to fix discontinuous point on the contour. Then, the extracted silhouettes are centered and normalized to the same size as 155*100 pixels.

Gait cycle should also be detected and calculated for further process. We simply focus on the pixels number of foreground in the below-waist lower half of the silhouette in each frame over time. This number will reach to a maximum when the two legs are farthest apart and drop to a minimum when the legs overlap. An example of gait cycle detection is shown in Fig. 5.
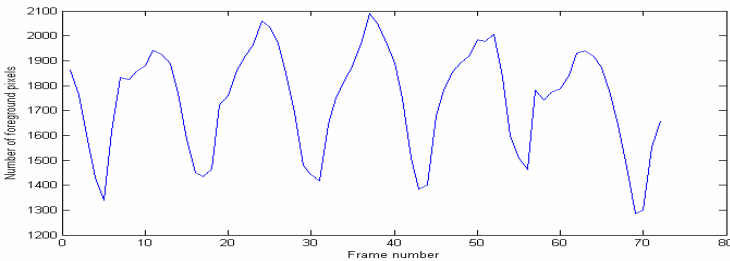
Fig. 5. Gait cycle detection by computing the number of foreground pixels (Y-axis) in each frame (X-axis) of gait sequences over time

4.2 Experimental Results and Analysis

As shown in Fig.4(b), we employ a Nearest Neighbor Classifier to determine the gender of Probe Set according to its Mahalanobis distance with Gallery Set. The reasons are, on the one hand, RCA is a method learning a Mahalanobis metric from equivalence constraints. Our experiment is a supervised recognition problem that each sample's gender information is labeled beforehand. On the other hand, gender recognition is a two-class classification problem that a Nearest Neighbor classifier is efficient and enough for this problem.

For all the 155*100 pixels silhouettes in one gait cycle, Radon transform generates 187 coefficients along one certain angle. Thus with θ ranging from 0 to 179 degrees, the size of the two-dimensional Radon-transformed coefficients matrix S_R can be 187*180, each column representing one angle degree of θ. Therefore, the Radon template T of samples is a complex matrix with each frame's initial phase (see Eq.(6)) and the size of T matrix is 187*180.

All the 60 persons, 32 males and 28 females, has five walking sequences, thus 300 samples are available totally. Since this is not a very large amount, we employ leave-one-out method to improve the accuracy of experimental results. Firstly, we select any five samples among the 300 samples belongs to the same person as Probe Set and leave the other 295 samples as Gallery Set. Secondly, we select the feature vector v_p and v_g for the both sets. Finally, we calculate the Mahalanobis distance between these two feature vectors.

$$d(v_p, v_g) = \sqrt{(v_p - v_g)^T \Sigma^{-1} (v_p - v_g)} \,. \tag{9}$$

The gender of v_p is determined as same as the gender of v_g which has the minimum Mahalanobis distance $d(v_p, v_g)$.

There are two schemes to select the feature vector. The first one is to select one certain angle degree of Radon-transformed template T, that is, certain column of T. The second one is the fusion of several or all the angle degrees of T, that is, fusion of several columns or all the 180 columns of T. We named the two feature select method as *FS1* and *FS2* separately.

The gender recognition correct rate results of scheme *FS1* is shown in Fig.6. Seen from this figure, generally the results grow better when the angle θ is next to 90° and grow worse when next to 0° and 180°. The best result is 95.67%, appearing in the angle of 124°, shown as the vertical line in Fig.6. And the worst result is 68.00% appearing in the angle of 173°.

In [9], N.V. Boulgouris et al. present their experimental results of individual identification probability ranging from 10% to 100% on USF's Gait Challenge database. Between our and their work, on the one hand, the research purposes differ much. Our work aims at gender recognition, while their work aims at individual recognition. On the other hand, the databases are also different. Nevertheless, our results shown in Fig.6 are still competitive to and even better than those results in [9].
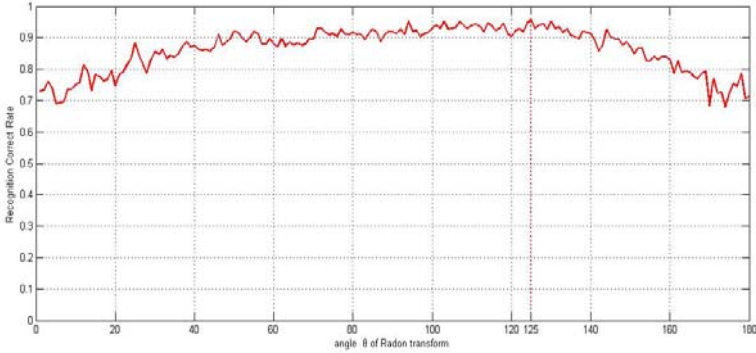
Fig. 6. Correct rates of gender recognition by selecting one certain degree of Radon-transformed template T. The angle degrees from $0°$ to $179°$ are on x-axis and the corresponding correct rates are on y-axis. The best correct rate, 95.67% appears when x=125.

The most common fusion method in scheme FS2 is to use all the columns in T. After concatenating each column in template T, a large feature vector is constructed. The size of columns in this feature vector is up to 187*180=33660. Therefore, the temporal and spatial cost must be considered seriously. Dimensional reduction is applied to reduce the size of each column from 187 down to 10-30. So size of fused feature vector is ranging from 1870*1 to 5610*1. But the recognition correct rates of this fusion scheme and vary much from 78.33% to 94.33%. This is not very stable or robust. There are also some other fusion scheme in FS2, which is a further work of this paper.

The following Table 1 shows the comparison between our result and former results on BUAA-IRIP Gait Database for the same purpose of gender classification.

Table 1. Comparison of experimental results for gender recognition

Correct Rate	Related Papers
89.5%	G.C Huang [12] ACCV2007
90.0%	D Zhang [13] ICPR2008
95.67%	this paper

5   Conclusion

We propose a new method for gender recognition via gait silhouettes. Firstly, in the feature extraction process, Radon transform on all the 180 angle degrees is applied to every silhouette to form a 187*180 coefficients matrix as S_R because of the advantages of Radon transform as its sensitiveness to angle variation and low spatio-temporal cost for calculating. All the silhouettes in one entire gait cycle construct the template T for one person. The dynamic information of walking is also associated to T by calculating the initial phase of each silhouette. RCA algorithm is employed on the radon-transformed templates T to get a maximum likelihood estimation of the within

class covariance matrix. Two kinds of feature selection method provide different gait recognition rates through the same Nearest Neighbor classifier by measuring the Mahalanobis distances on our BUAA-IRIP Gait Database for the reason of its quantity balance of different genders. How to make a proper fusion of these features to get better result is still underway.

Acknowledgments

References

1. Boulgouris, N.V., Hatzinakos, D., Plataniotis, K.N.: Gait Recognition: a Challenging Signal Processing Technology for Biometric Identification. IEEE Signal Processing Magazine 22(6), 78–90 (2005)
2. Kozlowski, L.T., Cutting, J.T.: Recognizing the Sex of a Walker from a Dynamic Point-Light Display. In: Perception and Psychology, pp. 575–580 (1977)
3. Lee, L., Grimson, W.: Gait Analysis for Recognition and Classification. In: IEEE Intl. Conf. on Automatic Face and Gesture Recognition, pp. 148–155 (2002)
4. Davis, J.W., Gao, H.: An Expressive Three-mode Principle Components Model for Gender Recognition. Journal of Vision, 362–377 (2004)
5. Yoo, J., Hwang, D., Nixon, M.S.: Gender classification in human gait using support vector machine. In: Blanc-Talon, J., Philips, W., Popescu, D.C., Scheunders, P. (eds.) ACIVS 2005. LNCS, vol. 3708, pp. 138–145. Springer, Heidelberg (2005)
6. Liu, Y., Collins, R., Tsin, Y.H.: Gait sequence analysis using frieze patterns. In: Heyden, A., Sparr, G., Nielsen, M., Johansen, P. (eds.) ECCV 2002. LNCS, vol. 2351, pp. 657–671. Springer, Heidelberg (2002)
7. Kale, A., Cuntoor, N., Yegnanarayana, B., Rajagopalan, A.N., Chellappa, R.: Gait Analysis for Human Identification. In: 4th Intl. Conf. on Audio- and Video-based Person Authentication, Guilford, pp. 706–714 (2003)
8. Boulgouris, N.V., Plataniotis, K.N.: An Angular Transform of Gait Sequences for Gait Assisted Recognition. In: IEEE Intl. Conf. on Image Processing, Singapore, pp. 857–860 (2004)
9. Boulgouris, N.V., Chi, Z.X.: Gait Recognition Using Radon Transform and Linear Discriminant Analysis. IEEE Trans. on Image Processing 16, 731–740 (2007)
10. Aharon, B.H., Tomer, H., Noam, S., Daphna, W.: Learning a Mahalanobis Metric from Equivalence Constraints. Journal of Machine Learning Research 6, 937–965 (2005)
11. Phillips, P.J., Moon, H., Rizvi, S., Raus, P.: The FERET Evaluation Methodology for Face Recognition Algorithms. IEEE Trans. Pattern Analysis and Machine Intelligence 22(10), 1090–1104 (2000)
12. Huang, G.C., Wang, Y.H.: Gender Classification Based on Fusion of Multi-view Gait Sequences. In: 8th Asian Conf. on Computer Vision, Tokoy, pp. 462–471 (2007)
13. Zhang, D., Wang, Y.H.: Investigating the Separability of Features from Different Views for Gait Based Gender Classification. In: 19th IEEE Intl. Conf. on Pattern Recognition, pp. 1–4 (2008)

Constrained Maximum Variance Mapping for Tumor Classification

Chun-Hou Zheng[1], Feng-Ling Wu[2], Bo Li[3], and Juan Wang[1]

[1] College of Information and Communication Technology, Qufu Normal University, Rizhao,
Shandong, 276826 China
Zhengch99@126.com
[2] College of Electrical Information and Automation, Qufu Normal University, Rizhao,
Shandong, 276826, China
[3] Intelligent Computing Lab, Hefei Institute of Intelligent Machines,
Chinese Academy of Sciences, Anhui 230031, China

Abstract. It is of great importance to classify the gene expression data into different classes. In this paper, followed the assumption that the gene expression data of tumor may be sampled from the data with a probability distribution on a sub-manifold of ambient space, an efficient feature extraction method named as Constrained Maximum Variance Mapping (CMVM), is presented for tumor classification. The proposed algorithm can be viewed as a linear approximation of multi-manifolds learning based approach, which takes the local geometry and manifold labels into account. The proposed CMVM method was tested on four DNA microarray datasets, and the experimental results demonstrated that it is efficient for tumor classification.

Keywords: Manifold learning, Gene expression data, Constrained maximum variance mapping, Tumor classification.

1 Introduction

DNA Microarray technology has made it possible to monitor simultaneously the expression levels for thousands of genes during important biological processes and across collections of related samples. Elucidating the patterns hidden in the gene expression data is a tremendous opportunity for functional genomics. By monitoring the expression levels in cells for thousands of genes simultaneously, microarray experiments may lead to a complete observation of the molecular variations among tumors, and hence result in a reliable classification. Recently, the gene expression data derived from such analyses have been employed to many cancer classification studies [1-3].

Gene expression data can be expressed as a gene expression matrix $X = (x_{ij})_{m \times n}$, where each row represents a gene, while each column represents a sample or a patient for tumor diagnosis. The numerical value of X_{ij} denotes the expression level of a specific gene i of a particular sample j. Generally, the very large number of variables (genes) with only a small number of observations (experiments) makes most of the data analysis methods, e.g. classification, inapplicable. Fortunately, this problem can be solved by selecting only the relevant features or extracting the essential features from the original data.

Feature extraction from the original data, which is often dictated by practical feasibility, is an important step in pattern recognition tasks. Over the past several years, the study on feature extraction methods has being conducted and many useful feature extraction techniques including linear and nonlinear methods, supervised or non-supervised ones have been well developed [4,5]. Among them, principal component analysis (PCA) and linear discriminant analysis (LDA) are two representative methods for linear feature extraction.

However, both PCA and LDA take the global Euclidean structure into account instead of the manifold geometry of the original data. If data points are resided on a manifold, the global Euclidean structure can not exactly describe the true distance between any two points. In order to characterize those nonlinear data, an idea for performing new method with local linear but global nonlinear transformation is put forward. Manifold learning based methods are representative ones for this idea. The basic assumption of manifold learning is that the input data lie on a smooth low dimensional manifold. Each manifold learning based method attempts to preserve a different geometrical property of the underlying manifold. Local linear approaches aim to preserving the proximity relationship among the input data.

In this paper, a Constrained Maximum Variance Mapping technique is put forward for tumor classification. Firstly an objective function is constructed to maximize the variance between different manifolds. In this objective function, the class information has been taken into account. Secondly in order to preserve the locality of each manifold, a constrained condition is appended to the objective function. Thirdly, we also adopted the strategy for linear approximation to the proposed one. On the one hand, the out-of-sample problem can be successfully overcome; on the other hand, due to introducing the linear transformation, the computational cost will be reduced greatly. Under such constraint, a projection can be found to map the local points with the same labels closer and the local points belonging to different classes farther, which contributes to classification. The prediction results on four tumor gene expression datasets show that our method is efficient and feasible.

2   Principle of Manifold Learning for Gene Expression Data

Most manifold learning algorithms assume that the input data resides on or close to a low dimensional manifold embedded in the ambient space. For most applications, the data is generated by continuously varying a set of parameters. Often the number of parameters is not very large. Otherwise, there will not be much benefit from dimensionality reduction. For example, a set of face images can be captured by varying the face pose, scale, position, and lighting conditions. With manifold learning, one can mapping the high dimension observe data to low dimensional manifold. As an example, Figure 1 (a) and (b) shows a 3D data set LLE-mapped to 2D. For manifold learning, dimensionality reduction can be achieved by constructing a mapping that respects certain properties of the manifold.
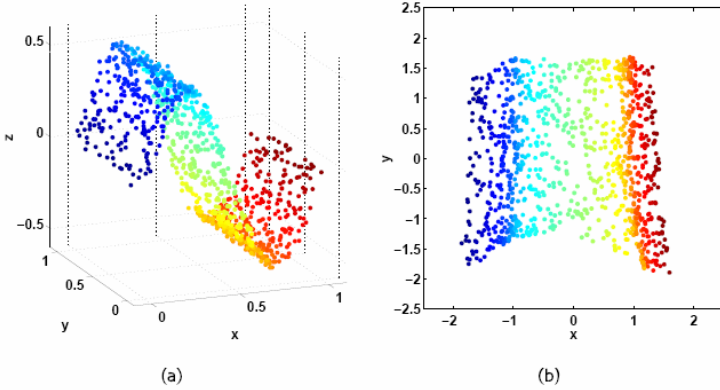
Fig. 1. (b) Data sampled from a sine-shaped manifold. (b) LLE-mapped data.

For gene expression data, we know that the cell state is concern to the environment and its own genes, so they should be determined by these outer and inner factors. Till now people still not clearly realize these factors. If we take the complicated environmental influence and inner gene factors as a set of parameters, then the gene expression data may be resides on or close to a low dimensional manifold embedded in the ambient space with these parameters. According to this assumption, the inherent nonlinear structure hidden in the observe gene expression data may be achieved by using manifold learning. Farther, these low dimensional manifolds may be used for tumor classification, gene regulatory networks construction, etc. In this paper, as an attempt in gene expression data analysis area, we will focus on the tumor classification using manifold learning.

3   Constrained Maximum Variance Mapping

For a general classification problem, the training sample points can be represented as a matrix $X = [X_1, X_2, ..., X_n], X_i \in \mathbb{R}^m, i = 1,...,n$, where n is the sample number and m is the feature dimensions. The class label of the sample X_i is denoted as $c_i \in \{1, 2,..., N_c\}$, where N_c is the number of the classes. In practice, the feature dimension m is often very high. The goal of the proposed algorithm is to transform the data from the original high-dimensional space to a low-dimensional one, i.e. $Y \in \mathbb{R}^{d \times n}$ with d<<m. Moreover, the transformation will separate the different manifolds farther under the constraint of local structure preserving.

3.1   Local Structure

In traditional linear dimensionality reduction methods, the global geometric structure of data is represented by Euclidean distance matrix. Correspondingly, the local scatter can be characterized by the Euclidean distance between any pair of the projected

sample points that are within any local k nearest neighbors or $\varepsilon - ball$ criterion. For example, if two points, X_i and X_j, are located in an $\varepsilon - ball$, then the locality can be expressed into the following form:

$$d_{ij} = \left(X_i - X_j \right)^T \left(X_i - X_j \right) \tag{1}$$

In terms of local relations of points, a local relation matrix L can be defined, the elements of which can be given as follows:

$$L_{ij} = \begin{cases} 1 & \left\| X_i - X_j \right\|^2 \leq \varepsilon \\ 0 & otherwise \end{cases} \tag{2}$$

Then the locality can be rewritten into:

$$\begin{aligned} J_L &= \sum_{ij} \left(Y_i - Y_j \right) \left(Y_i - Y_j \right)^T L_{ij} \\ &= 2\sum_i Y_i D_{ii} Y_i^T - 2\sum_{ij} Y_i L_{ij} Y_j^T = 2tr\left\{ Y(D-L)Y^T \right\} \end{aligned} \tag{3}$$

where $D_{ii} = \sum_j L_{ij}$

3.2 Dissimilarities between Manifolds

The Euclidean distance is often taken as a measure of the dissimilarity. If the Euclidean distance between two points is very large, we can say the probability of their dissimilarity is very high. Otherwise, they probably are similar to each other. So we can use the sum of the mean distance to measure the dissimilarities between manifolds. It must be noted that the dissimilarities also exist between different manifolds. How to distinguish one manifold from the others will heavily depend on the manifold labels. As mentioned above, the data distributed on a manifold are belonging to the same class. So we can construct a label matrix H to mark the label information of each point, where H shows the label information as follows:

$$H_{ij} = \begin{cases} 0 & \text{if } X_i \text{ and } X_j \text{ have the same class label} \\ 1 & otherwise \end{cases} \tag{4}$$

Then the dissimilarities between manifolds will be defined as the following equation:

$$\begin{aligned} J_D &= \sum_{i,j}^n H_{ij} (Y_i - Y_j)(Y_i - Y_j)^T \\ &= 2\sum_i Y_i Q_{ii} Y_i^T - 2\sum_{ij} Y_i H_{ij} Y_j^T = 2tr\left\{ Y(Q-H)Y^T \right\} \end{aligned} \tag{5}$$

where $Q_{ii} = \sum_j H_{ij}$

3.3 The Principle of Constrained Maximum Variance Mapping

Linear features Y are obtained by a linear transformation matrix $A = [A_1, A_2, ..., A_d]$, where $Y_i = A^T X_i$. In the proposed algorithm, a linear transformation is plug so that the out-of-sample problem can be successfully avoided. So J_L and J_D can be rewritten into the following form respectively:

$$J_L = tr\{Y(D-L)Y^T\} = tr\{A^T X(D-L)X^T A\} \tag{6}$$

$$J_D = tr\{Y(Q-H)Y^T\} = tr\{A^T X(Q-H)X^T A\} \tag{7}$$

The constrained objective function can be represented as follows:

$$J(A) = \max(J_D) = tr\{A^T X(Q-H)X^T A\}$$
$$s.t. \quad tr\{A^T X(D-L)X^T A\} = tr\{X(D-L)X^T\} \tag{8}$$

This constrained optimization problem can be figured out by enforcing Lagrange multiplier. Then we have:

$$X(Q-H)X^T A = \lambda X(D-L)X^T A \tag{9}$$

From Eqn. (9), it can be found that A is composed of the eigenvectors associated with the d top eigenvalues by solving the corresponding generalized eigen-equation.

  The detailed algorithm can be found in our previous work [13].

4 Experimental Results

In this section, the performance of CMVM is evaluated in comparison with other representative methods such as PCA, supervised LLE (SLLE) [6], partial least squares (PLS), LDA, etc. After extracting features with the proposed algorithm, different pattern classifiers can be adopted for recognition, including K-NN, neural networks, support vector machine (SVM), classification trees, etc. In this study, we apply the K-NN classifier for its simplicity. In addition, it may be the best classifier for tumor classification with low dimensional features [7]. In the K-NN classifier, the k was set to 5. Actually, we varied it from 3 to 7 and the experimental results were almost the same.

4.1 Two-Class Classification Experiments

In this subsection, two publicly available microarray datasets are used to study the tumor classification problem. They are Prostate cancer dataset [8] and Breast cancer dataset [9]. In these two datasets, the data samples have been assigned to a training set and a test set. The characteristics of the two datasets are listed in Table 1.

Table 1. Summary of the datasets for the two binary cancer classification problems

Datasets	Training set		Test set		Genes
	Class 1	Class 2	Class 1	Class 2	
Prostate cancer data	52	50	25	9	12600
Breast cancer data	34	44	12	7	24188

Table 2. The mean classification accuracies and deviations on the two datasets

No.	Methods	Prostate cancer dataset		Breast cancer dataset	
		Accuracy	Dimension	Accuracy	Dimension
1	CMVM+K-NN	92.65±3.37	2	72.63±8.88	1
2	PCA+K-NN	74.12±4.34	2	67.89±12.27	48
3	SLLE +K-NN	92.35±3.97	3	68.95±11.49	2
4	LDA+K-NN	90.00±2.84	1	67.89±11.76	1
5	SVM	88.10±4.93		68.42±7.62	

   To obtain reliable experimental results with comparability and repeatability for different numerical experiments, we not only used the original division of each dataset for training and testing, but also reshuffled all datasets randomly in the experiments. In total, the numerical experiments were performed with 20 random splitting of the two original datasets. In addition, the randomized training and test sets contain the same amount of samples of each class as those of the original training and test sets. We built the classification models using the training samples and estimated the correct classification rates using the test set. The best classification results for tumor and normal tissues using our proposed method (CMVM+K-NN) are listed in Table 2. For each classification problem, the experimental results were reported by the mean value and standard deviation of the accuracy on the original data set and the 20 randomly partitioned sets as described above. The corresponding dimensions are also listed in Table 2.

   To show the efficiency of the proposed method, the results by the other three methods (Methods 2-5) are also listed in Table 2 for comparison. In method 2, we first used PCA to reduce the dimensions of the gene expression data, and then used K-NN for classification. Methods 4 is similar to Method 2 except that PCA is replaced by LDA. Supervised LLE (SLLE) [6,10] is a variation of LLE and it has been successfully used for gene expression data classification [10]. In the experiments, the number of nearest neighbors for constructing the nearest neighbor graph of SLLE was set to 5. The experimental results of the four methods listed in Table 2 are their best classification results, i.e. the highest accuracy by varying the dimension of used features, with the statistical means and standard deviations of the accuracy on the original test data set and 20 randomizations splits.
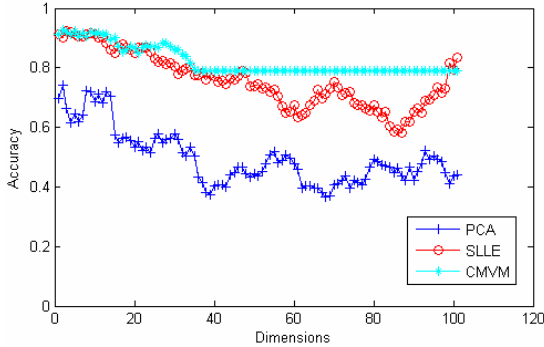
Fig. 2. The mean classification accuracy on the test set of the Prostate dataset
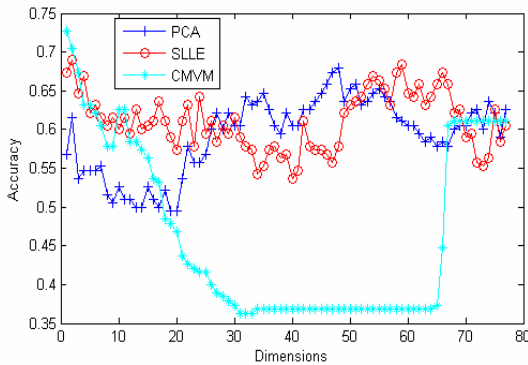
Fig. 3. The mean classification accuracy on the test set of the Breast dataset

To better illustrate the experimental results, we show the accuracies of Methods 1 to 3 in Figures2, 3, where the X-axis represents the d-dimensions, i.e. the number of used features extracted from the original data, and the Y-axis represents the mean accuracy of the 21 experiments.

From Table 2 and Figures 2~3 it can be seen that for the two datasets both SLLE and CMVM could achieve good classification results. And the classification result of CMCM is better than SLLE for Prostate dataset. Overall, the proposed method is very competitive in both the mean accuracy and the standard deviation.

4.2   Multi-class Classification Experiments

In this subsection, we used another two datasets to further investigate the classification ability of the proposed method. One is the Leukemia dataset [11], where the distinction between acute myelogenous leukemia (AML) and acute lymphoblastic leukemia (ALL) is known. The division of ALL into T and B cell subtypes is also known. The dataset contains 5000 genes in 38 cells and consists of 19 cases of B_cell acute lymphoblastic leukemia (ALL_B), 8 cases of T_cell acute lymphoblastic leukemia (ALL_T) and 11 cases of acute myelogenous leukemia (AML). We randomly selected 10 cases of ALL_B, 4 cases of ALL_T and 6 cases of AML as the training

set, and use the rest samples as test data. The experiment was run 21 times and the results are listed in Table 3 and Figure 4. For SLLE, the number of nearest neighbors for constructing the nearest neighbor graph was set to 3 since there are only 4 samples of normal in the training set (SLLE algorithm requires that the nearest neighbor of a sample should be of the same class). In this experiment, except PCA, SLLE and LDA, we also use PLS to do the experiment (Because of the very high dimensions and many samples, our computer (CPU 3.06GHz, Memory 1GB) can not process the Prostate cancer data set and the Breast cancer data set using PLS). Since the SVM is not specialize in multi-class classification, we have not used it to classify this dataset.

Table 3. The multi-class classification accuracies by different methods

No.	Methods	Leukemia dataset		Central Nervous System Tumors	
		Accuracy	Dimension	Accuracy	Dimension
1	CMVM+K-NN	98.33±3.75	2	94.38±3.55	3
2	PCA+K-NN	86.67±4.68	5	91.88±4.22	4
3	SLLE +K-NN	90.00±7.77	14	92.50±6.45	14
4	LDA+K-NN	92.78±6.44	2	95.63±4.22	3
5	PLS+K-NN	90.00±7.77	4	93.12±3.55	3

Another dataset is the central nervous system tumors dataset [12], which is composed of four types of central nervous system embryonal tumors. It contains 5597 genes in 34 samples representing four distinct morphologies: 10 classic medulloblasomas, 10 malignant gliomas, 10 rhabdoids and 4 normals. In this experiment, we randomly selected 5 medulloblasomas, 5 malignant gliomas, 5 rhabdoids and 3 normals as training set, and use the rest samples as test data. This experiment was run 21 times and the results are listed in Table 3 and Figure5. Since these are only 3 samples of normal in the training set of central nervous system tumors data, the number of nearest neighbors for constructing the nearest neighbor graph in the SLLE algorithm was set to 2.
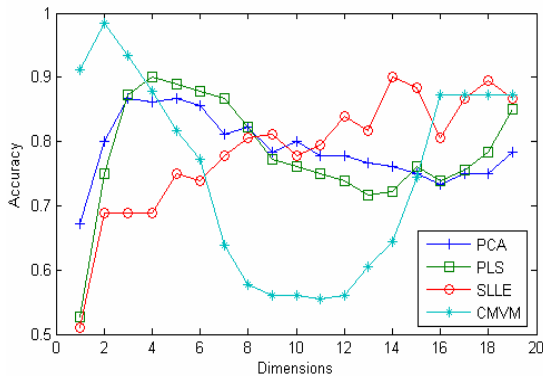
Fig. 4. The mean classification accuracy on the test set of the Leukemia data
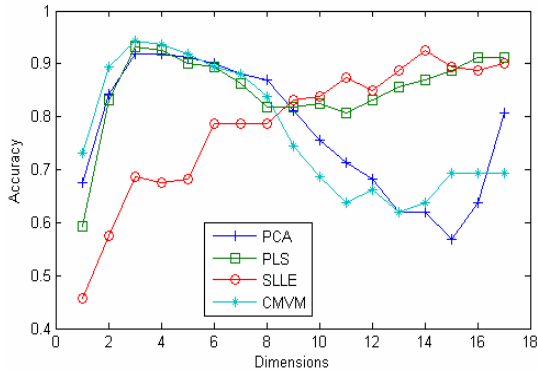
Fig. 5. The mean classification accuracy of Central Nervous System Tumors dataset

5   Conclusions

From all the four experiments, it can be seen that the proposed method achieves much higher accuracy than PCA, PLS and SLLE except Central Nervous System Tumors dataset. Besides, CMVM always achieves better results when the dimensions are lower. Of course, this is only some primary conclusion from our experiments. In the future more experiments will be done to verify the conclusion.

Feature extraction techniques have been widely employed to reduce the dimensionality of the original data and to enhance the discriminatory information so that different classes of patterns can be easily classified. In this study we proposed a constrained maximizing margin mapping for effectively and efficiently extracting the most discriminatory features from patterns. The experimental results demonstrated that it is very effective in extracting the most discriminant features from the gene expression data.

References

1. Kim, K.Y., Ki, D.H., Jeung, H.C., Chung, H.C., Rha, S.Y.: Improving the Prediction Accuracy in Classification Using the Combined Data Sets by Ranks of Gene Expressions. BMC Bioinformatic 9, 283 (2008)
2. Zhu, M.L., Martinez, A.M.: Using the Information Embedded in the Testing Sample to Break the Limits Caused by the Small Sample Size in Microarray-based Classification. BMC Bioinformatics 9, 280 (2008)
3. Yeoh, E.J., Ross, M.E., Shurtleff, S.A., Williams, W.K., Patel, D., Mahfouz, R., Behm, F.G., Raimondi, S.C., Relling, M.V., Patel, A., et al.: Classification, Subtype Discovery, and Prediction of Outcome in Pediatric Acute Lymphoblastic Leukemia by Gene Expression Profiling. Cancer Cell 1, 133–143 (2002)

4. Tenenbaum, J.B., de Silva, V., Langford, J.C.: A Global Geometric Framework for Nonlinear Dimensionality Reduction. Science 290, 2319–2323 (2000)
5. Belkin, M., Niyogi, P.: Laplacian Eigenmaps for Dimensionality Reduction and Data Representation. Neural Computatio 15, 1373–1396 (2003)
6. Bai, X.M., Yin, B.C., Shi, Q., Sun, Y.F.: Face Recognition Based on Supervised Locally Linear Embedding Method. Journal of Information & Computation Science 4, 641–646 (2005)
7. Dudoit, S., Fridyland, J.F., Speed, T.P.: Comparison of Discrimination Methods for the Classification of Tumor Using Gene Expression Data. Journal of the American statistical association 97, 77–87 (2002)
8. Singh, D., Febbo, P.G., Ross, K., Jackson, D.G., Manola, J., Ladd, C., Tamayo, P., et al.: Gene Expression Correlates of Clinical Prostate Cancer Behavior. Cancer Cell 1, 203–209 (2002)
9. van't Veer, L.J., Dai, H., Van De Vijver, M.J., He, Y.D., et al.: Gene Expression Profiling Predicts Clinical Outcome of Breast Cancer. Nature 415, 530–536 (2002)
10. Pillati, M., Viroli, C.: Supervised Locally Linear Embedding for Classification: an Application to Gene Expression Data Analysis. In: Proceedings of 29th Annual Conference of the of the German Classification Society (GfKl 2005), pp. 15–18 (2005)
11. Brunet, J.P., Tamayo, P., Golun, T.R., Mesirov, J.P.: Metagenes and Molecular Pattern Discovery Using Matrix Factorization. Proc. Natl. Acad. Sci. 101, 4164–4416 (2004)
12. Pomeroy, S.L., Tamayo, P., et al.: Prediction of Central Nervous System Embryonal Tumour Outcome Based on Gene Expression. Nature 415, 436–442 (2002)
13. Li, B., Huang, D.S.: Feature Extraction Using Constrained Maximum Variance Mapping. Pattern Recognition 41, 3287–3294 (2008)

A Hybrid Ant Colony Algorithm for the Grain Distribution Centers Location

Le Xiao[1,3] and Qiuwen Zhang[2]

[1] Department of Control Science and Engineering, Huazhong University of Science and Technology, Wuhan 430074, China
[2] China Airborne Missile Academy, Luoyang 471009, China
[3] College of Information Science and Engineering, Henan University of Technology, Zhengzhou 450001, China
zhangqwen@sohu.com

Abstract. Grain distribution center is the pivot of a grain logistics system. To define the location of grain logistics distribution center is the key of grain logistics system analysis. In this paper, according to the characteristics and requirements in the selection of the location, a mathematical model applied to the location selection was established on the basis of lowest transportation cost. A hybrid ant colony algorithm was then used to solve the model, the algorithm is based on the combination of genetic algorithm and ant colony clustering algorithm .First, it adopts genetic algorithm to give information pheromone to distribute. Second, it makes use of the ant colony clustering algorithm to give the precision of the solution. The algorithm can avoid premature convergence and prevent fast local optimal solution. The instance demonstrates that the hybrid algorithm can effectively get the grain logistics center optimal solution.

Keywords: ant colony algorithm (ACA); genetic algorithm (GA); grain distribution centers location; Optimization.

1 Introduction

Distribution center is a logistics base that specially focuses on distributes products. It can prepare products according to customers' ordering demand through transmission, classification, keeping, machining and information processing. It also can distribute products to customers quickly, exactly and cheaply [1].Distribution center is a goods delivery facility, which has multifunction, intensive and comprehensive service. Grain distribution center plays an important role in the grain logistics system. The better location scheme can save cost effectively, promote the coordination and combining of production and consumption, guarantee the balance development of grain logistics system. Once the distribution center is decided, it will run for a long time. The center has not only a direct connection with freight but also a big influence upon the work efficiency and the grain logistics control level. So the reasonable location of grain distribution center seems very important.

Because of the important function of logistics distribution center and its position, many scientific research personnel have developed the research work for this

problem, a series of location model and algorithm have been established, such as gravity method, numerical analysis method, linear programming method and heuristic algorithm and so on[2-4]. Where gravity method and numerical analysis method have mainly used in single distribution center location model, linear programming method and heuristic algorithm have mainly used in more distribution center location model, but linear programming method has strict requirement on linear of goal function in the problem of location, the large scale practical problem are difficult to solve in heuristic algorithm although which has overcome the deficiency of linear programming method. Some new intelligent calculation methods such as genetic algorithm [5], ant colony algorithm [6-9] etc have been developed which are enlightened by the mechanism of the biological evolution in recent years. The new methods have been used to solve some complex optimization problems constantly.

   This paper carries out research on the location problem of grain distribution, according to the characteristics and requirements of the location problem of grain distribution center, the mathematical model of the location problem is constructed based on the minimum transport cost, a hybrid algorithm is based on the combination of genetic algorithm and ant colony clustering algorithm is introduced for this model, and the validity of this algorithm is proved by a cases analysis of the site selection problem of grain logistics scheduling.

2   Problem Description

The location problem of grain logistics center can be description as follows: Address set of all grain demand point are given in a certain area, from which a certain number of addresses are selected to establish the grain logistics distribution center, then establish a series of distribution areas, and realize the distribution of each grain demand point, make the minimize total distribution fees between the grain logistics center established from selected grain demand point and the distribution system established from each grain demand point. It is with a view to speed the velocity of grain flowing and to avoid more unnecessary distribution cost.

2.1   Model Assumptions

In order to convenient for establishing the mathematical model of grain distribution center, make it not too complicated but have some practicality, we presume system meet some conditions as follows:

   1) Selected grain distribution center only in a certain alternate range;
   2) A grain distribution center can be served by multiple grain supplying goods point, also a grain demand point can be supplied by multiple grain logistics center;
   3) The capacity of grain distribution center can meet the demands;
   4) The demand of each grain demand point must have been known;
   5) Using the same transportation means at every point in the systems;
   6) System contains secondary transportation, namely contains the transportation from grain supplying goods point to grain distribution center and from grain logistics center to grain demand point;
   7) Transportation cost is proportional to traffic volume;

8) The demand quantity of grain demand point according to the regional amount which must be known constant;

9) The unit of transportation cost of grain distribution center and grain demand point, and the unit of management cost are known constant;

10) System cost only considers the fixed warehouse construction cost, transportation cost and the management cost of flowing through grain distribution center.

2.2 Mathematical Formulation and Column Generation

Suppose according to the above that we can draw the mathematical model expression of grain distribution center's location:

$$\min U = \sum_{k=1}^{l}\sum_{i=1}^{m} c_{ki} W_{ki} + \sum_{i=1}^{m}\sum_{j=1}^{n} h_{ij} X_{ij} + \sum_{k=1}^{l}\sum_{i=1}^{m} g_i \sqrt{W_{ki}}\,/2 + \sum_{i=1}^{m} z_i F_i \cdot \qquad (1)$$

$$A_k - \sum_{i=1}^{m} W \geq 0, (k = 1,2,\cdots,l) \cdot \qquad (2)$$

$$M_i - \sum_{k=1}^{l} W_{ki} \geq 0, (i = 0,1,2,\cdots,m) \cdot \qquad (3)$$

$$\sum_{i=1}^{m} X_{ij} - D_j \geq 0, (j = 1,2,\cdots,n) \cdot \qquad (4)$$

$$\sum_{j=1}^{n} X_{ij} - \sum_{k=1}^{l} W_{ki} = 0, (i = 1,2,\cdots,m) \cdot \qquad (5)$$

$$W_{ki} \geq 0, X_{ij} \geq 0, (k=1,2,\cdots,l; i=0,1,2,\cdots,m; j=0,1,2,\cdots,n) \cdot \qquad (6)$$

The symbols of model are illustrated as following:

U-total cost; l-the number of grain supplying goods point; m-the number of alternate grain distribution; n-number of grain demand point; C_{ki}-the unit of transportation cost from grain supplying goods point k to grain distribution center W_{ki}- the traffic volume from grain supplying goods point k to grain distribution center i; h_{ij}- the unit of transportation cost from grain distribution center i to grain demand point j; X_{ij}- the traffic volume from grain distribution center i to grain demand point j; g_i-the circulation of unit grain management cost of grain distribution center i; Zi-for 0,1 variable, when $Zi=1$, express grain distribution center i is selected; when $Zi=0$, express grain distribution center i is not selected; F_i-the fixed investment cost of grain distribution center i; A_k- supply ability of supplying goods point k; Mi-construction capacity of grain distribution center selected i; Di— the demand quantity of grain demand point j. In addition, formula (1) expresses the total cost of grain distribution center; formula (2) expresses the product population are sent to various grain distribution center from grain supplying goods point that can't over its supply ability; formula

(3) expresses the amount of goods are obtained by grain distribution center i that can't over its construction capacity; formula (4) expresses the product population are supplied from various grain distribution center to consumer j that should meet the demands of this grain demand point; formula (5) expresses the balance of grain in and our amount of grain distribution center; formula (6) expresses the nonnegative requirement of parameters in the practice applied problem.

3   Hybrid Ant Colony Algorithm Based on Genetic Algorithm for Grain Distribution Center Location

The location problem of grain distribution center is an optimization problem with many restraints and minimum cost problem. A hybrid algorithm is based on the combination of genetic algorithm and ant colony clustering algorithm is introduced for this model, this paper introduces the concept of monitor operator to search the location optimum scheme, which improves the chance to find the global optimal solution, accelerates the convergence rate of algorithm, and enhances the computing efficiency greatly.

3.1   Genetic Algorithm

The operators of standard genetic algorithm generally includes three basic types of selection crossover and mutation, they constitute the core of genetic algorithm possess powerful ability of searching. A new concept of monitor operator is put forward based on the fact in this paper; this operator can improve the convergence speed and searching performance of genetic algorithm much more.

l) Selection: this paper uses an optimal manner to implement the selection operation, that is, first of all, the largest chromosome to fitness value of parent-off springs occurs at least once in progeny- off springs, then performs the secretion operation according to the standard roulette way, thus, to ensure the best chromosome coming into next generation.

2) Crossover: the crossover operator for genetic algorithm is a gene recombination process to imitate sexual reproduction in nature, its function is to inherit original good gene to next generation individual, and generate new individuals which contain more complex gene structure. The two-point crossover method is adopted to perform the crossover operation, because the information of one-point crossover operation is small, and its important gene of the end of bit string always be changed. The linear descending function is used to generate crossover rate PC in this article, selected PC is 75% in the first generation, decrease linearly to the last generation the Pc is 25%. Finally, it is favorable to contain more much information in the original stage of the computation, and raise the convergence of algorithm in the late stage.

3) Mutation: this article adopts a linear function to produce mutation rate Pm, its equation is:

$$p_m = 0.001 + (0.20 - 0.001) \times current\,algebra / total. \tag{7}$$

PM increased gradually with the increase of the generations, the purpose is to accelerate convergence in the late stage of computation.

4) Monitor Operator: under the premise of unchanged the individual gene structure, put the former generation into the current population. Thus we can guarantee the best individual survived, ensure the best individual of current generation to enter future generation to avoid the lost of the best individual. The new individuals obtained will be sorted according to the fitness degree, the optimum chromosome at the first place; the best individual in population will be determined and stored. After selection crossover and motion, compared with the chosen new individual and the old best individual, if the fifteen degree of the new individual is superior to the old's, the new individual will be replaced and the optimum dividable of formerly storing will be inserted into the paired library. Thus, which is not only save the best individual, but also make the best individual have a chance to.

3.2 Improved Ant Colonies Algorithm Design

In the grain distribution center, first used genetic algorithm division of the customers, then used ant colony algorithm for each vehicle assigned to customers. At each iteration of the basic ant Colony method, each ant builds a solution of the problem step by step. At each step the ant makes a move in order to complete the actual partial solution choosing between elements of a set A_i of expansion states, following a probability function. For each ant k probability of moving from present state i to another state j is calculated taking into account [10]:

1. Attractiveness of the move according to the information of the problem.
2. Level of pheromone of the move that indicates how good the move was in the past.
3. A allowed$_k$ list of forbidden moves; in the ant algorithm original version formula for $p_{ij}^k(t)$ is:

$$
p_{ij}^k(t) = \begin{cases} \dfrac{[\tau_{ij}(t)]^\alpha [\eta_{ij}(t)]^\beta}{\sum\limits_{s\in allowed_k} [\tau_{ij}(t)]^\alpha [\eta_{ij}(t)]^\beta} & j \in allowed \\ 0 & otherwise \end{cases} , \tag{8}
$$

where α, β are parameters that are used to establish the relative influence of η_{ij} versus τ_{ij}. After iteration t is complete, that is when all the ants have completed their solutions, the pheromone levels are updated to:

$$
\tau_{ij}(t+1) = \rho.\tau_{ij}(t) + \Delta\tau_{ij} \quad \Delta\tau_{ij} = \sum_{k=1}^{m} \Delta\tau_{ij}^k , \tag{9}
$$

$$
\Delta\tau_{ij}^k = \begin{cases} Q/L_k & if \quad ant \quad k \quad pass \quad by \quad path \quad ij , \\ 0 & otherwise \end{cases} \tag{10}
$$

where Q is a constant that ants week cycle of the total amount of pheromone. L_k representing the *Kth* ant which walks path length in this cycle, it reflects the shortest path

within the overall situation, the search system to improve convergence speed. Q, C, α, β, ρ experimental methods can be used determine their optimal combination.

Pheromone Updating. In the real world, the higher concentration of pheromone, volatile sooner, and the lower concentration of pheromone, the more diffuse volatile, so effective in preventing some pheromone concentration on the path to unlimited growth, and the pheromone on some path reduce to zero, reducing the possibility of a local optimum[11]. In such circumstances volatile factor from a constant to a variable function by τ_{ij}, therefore, use the following update rules:

$$\tau_{ij}(t+N_{\max})=(1-\rho*\tau_{ij}(t))\cdot\tau_{ij}(t)+(\rho*\tau_{ij}(t))*\Delta\tau_{ij}(t,t+N_{\max}), \tag{11}$$

$$\Delta\tau_i(t,t+N_{\max})=\sum_{k=1}^{m}\Delta\tau_i^k(t,t+N_{\max}) \quad \Delta\tau_i^k(t,t+N_{\max})=\frac{Q}{L_{kl}}, \tag{12}$$

where $N_{\max}$ is a once cycle required the maximum time. Assume a unit time before ant's progress; L_{kl} representing total distance traveled of vehicle routing k.

Because of the constraint conditions of the location of grain distribution center corresponds to the mathematical model are much more, unsolvable of the adopted coding method take up a large proportions in the population, punishment strategy may be considered to use. The Essence of punishment is that the constrained problem is converted into unconstrained problem through publish unsolvable. We construct evaluation function by using the method of Michalewic and Attia, which is fitness function.

4   Simulation Results

Grain enterprises have a main grain processing factory, the production capacity of the factory is 3000 ton, another have ten grain demand points, the demand quantity is D_j(j=1,2.3…10), see table 1, the alternative places of grain distribution center are ten number, that is frontal ten grain demand points, see table 1, the fixed cost and the unit of management cost and the construction capacity of which are shown, in addition, the unit of transportation cost-generating which from grain factory to various grain logistics center are seen table 2, the unit of transportation cost-generating which from each grain distribution center to each customers are shown.

Table 1. The fixed cost and construction capacity

Grain Distribution Center	CS	NM	GZ	NN	FZ	WH	HZ	HF	NJ	ZZ
Fixed Investment（Wan）	75	50	100	60	50	70	100	50	100	50
Construction Capacity（tons）	20000	5000	10000	5000	3000	10000	3000	5000	2000	3000
Unit management costs（Yuan）	20	15	30	15	20	20	30	15	30	15

Table 2. The fixed cost and construction capacity

Unit freight (Tons/Yuan)	Grain Distribution Center										Factory Canpacity
	CS	NM	GZ	NN	FZ	WH	HZ	HF	NJ	ZZ	
Grain Factory	60	100	180	140	200	100	200	170	300	320	30000 tons

Table 3. The results of grain distribution center

	Grain Distribution Center			
	ChangSha	NanMu	GuangZhou	WuHan
Traffic Volume（tons）	7940	1012	1652	3757

The algorithm developed in this paper can be used to solve these problems, the program has been realized on P-4 PC, and by the Matlab7.0-Tool, the results are as follows: the average total cost is 440.820256, the minimum value of the total expense is 398.991005,the four alternate sites selected that changsha、nanchang、guangzhou and wuhan are the optimal grain distribution center. Meanwhile, the traffics volume which from grain factory to grain distribution center are obtained as table 3 shown.

5   Conclusions

The location of grain distribution center is a key link in grain logistics system. According to the characteristics and requirements of the location problem of grain logistics distribution center, this paper formulates a mathematical model of its based on the lowest transportation cost. On the basis of the complexity of this mathematical model, a hybrid algorithm is based on the combination of genetic algorithm and ant colony clustering algorithm is introduced for this model, to find the optimum scheme of the location. So it improves the chance to find out the global optimal solution and accelerates the convergent speed. The practical example of grain distribution site selection proves that the algorithm suiting for the development requirement of current various grain logistics scheduling, which may be have an important reference value for grain distribution location.

Acknowledgments

References

1. Kasilingam, R.G.: Logistics and Transportation: Design and Planning. Kluwer Academic Publisher, Boston (1998)
2. Yuan, Y.X.: A Scaled Central Path for Linear Programming. Journal of Computational Mathematics 19(1), 35–40 (2001)
3. Konstantinos, G.Z., Konstantinos, N.A.: A Heuristic Algorithm for Solving Hazardous Materials Distribution Problems. European Journal of Operational Research 152, 507–519 (2004)
4. Zhang, Q., Gao, L.Q.: A Two-phase Heuristic Approach to the Location Routing Problem. Control and Decision 19(7), 773–777 (2004)
5. Qian, J., Pang, X.H., Wu, Z.M.: An Improved Genetic Algorithm for Allocation Optimization of Distribution Centers. Journal of Shanghai Jiao tong University (Science) 9(4), 73–76 (2004)
6. Dorigo, M., Dicaro, G., Gambardella, L.M.: Ant Algorithms for Discrete Optimization. Artificial Life 5(2), 137–172 (1999)
7. Colorni, A., Dorigo, M., Maniezzo, V., Trubian, M.: Ant System for Job-shop Scheduling. Belgian Journal of Operations Research, Statistics and Computer Science (JORBEL) 34, 39–53 (1994)
8. Costa, D., Hertz, A.: Ants Can Colour Graphs. Journal of the Operational Research Society 48, 295–305 (1997)
9. Dorigo, M., Gambardella, L.M.: Ant Colony System: A Cooperative Learning Approach to the Traveling Salesman Problem. IEEE Transactions on Evolutionary Computation 1, 53–66 (1997)
10. Gambardella, L.M., Dorigo, M.: HAS-SOP: An Ant Colony System Hybridized with a New Local Search for the Sequential Ordering Problem. INFORMS, Journal on Computing 12(3), 237–255 (2000)
11. Gambardella, L.M., Taillard, E.D., Dorigo, M.: Ant Colonies for the Quadratic Assignment Problem. Journal of the Operational Research Society 50, 167–176 (1999)

Parallel Genetic Algorithms for Crystal Structure Prediction: Successes and Failures in Predicting Bicalutamide Polymorphs

Marta B. Ferraro[1], Anita M. Orendt[2], and Julio C. Facelli[2,3,*]

[1] Departamento de Física, Facultad de Ciencias Exactas y Naturales, Universidad de Buenos Aires, Ciudad Universitaria, Pab. I (1428), Buenos Aires, Argentina,
[2] Center for High Performance Computing and [3] Department of Biomedical Informatics, University of Utah, 155 South 1452 East Room 405, Salt Lake City, UT 84112-0190, US
julio.facelli@utah.edu

Abstract. This paper describes the application of our distributed computing framework for crystal structure prediction, Modified Genetic Algorithms for Crystal and Cluster Prediction (MGAC), to predict the crystal structure of the two known polymorphs of bicalutamide. The paper describes our success in finding the lower energy polymorph and the difficulties encountered in finding the second one. The results show that genetic algorithms are very effective in finding low energy crystal conformations, but unfortunately many of them are not plausible due to spurious effects introduced by the energy potential function used in the selection process. We propose to solve this by using a multi objective optimization GA approach, adding the unit cell volume as a second optimization target.

Keywords: Parallel genetic algorithms, crystal structure prediction, drug polymorphism.

1 Introduction

Bicalutamide (marketed as Casodex, Cosudex, Calutide, Kalumid) is an oral non-steroidal anti-androgen used in the treatment of prostate cancer [1] and hirsutism [2]. A recent study shows that bicalutamide can crystallize in at least two different polymorphic forms. [3] There are three structures deposited in the Cambridge Crystallographic Data Center for this compound, with one form reported by two research groups, JAYCES [4] and JAYCES01 [3] and the other polymorph JAYCES02 reported by Vega et al. [3] The first polymorph belongs to the $P_{21/c}$ symmetry group and the second to the P_{-1}. To study the ability of MGAC (Modified Genetic Algorithms for Crystal and Cluster Prediction) to predict these polymorphs is an especially interesting contribution to the understanding polymorphism in pharmaceutical compounds because bicalutamide not only shows packing polymorphism, but also conformational polymorphism. In the $P_{21/c}$ crystal the molecule is in an extended conformation while in the P_{-1} polymorph, it is in a very compact folded conformation.

---

[*] Corresponding author.

Polymorphism is the ability of a compound to crystallize in more than one crystal structure and thus having different values for properties such as solubility, bioavailability, shelf life, crystal size and color, vapor pressure, and shock sensitivity depending on which polymorph is present. The existence of polymorphic structures was originally thought to be a rarity but now it is known to be widely observed. [5-7]

The ability to readily and reliably predict crystal structures has become a desirable goal for the modeling and crystal engineering communities. The periodic blind tests [8-10] of crystal structure prediction (CSP) organized by the Cambridge Crystallographic Data Centre (CCDC) have been the focal point for this community and they reflect the overall progress in the field. The tests show a continuous improvement in the capabilities for predicting the crystal structures of simple rigid molecules and indicate that the methods should now be extended to more complex systems such as flexible molecules and co-crystals. However, questions remain on how effective these methods are in finding multiple polymorphs in pharmacologically active compounds.

Our research has been concentrated in using parallel genetic algorithms (GA), which are based on the idea of Darwinian natural evolution. [11, 12] Populations of candidate individuals (i.e., feasible solutions to the problem) compete with one another through selection, crossover, and mutation operations to produce individuals that have higher fitness, thereby concentrating the search towards the global minimum. The advantage of GA is that they extensively search the "good regions" of the configuration space because genetic operators create children whose structures can greatly differ from their parents, but belong to provable regions in the configurational space. [13] In addition, GA are naturally amenable to parallelization schemes, an important feature for computationally intensive problems like CSP. Our previous work [14-18] presented the development and use of the Modified Genetic Algorithms for Crystal and Cluster structures (MGAC) method, in which all the crystal structures considered by the GA are locally optimized, i.e., they correspond to a local minimum in the potential energy with respect to all intra- and inter- molecular parameters defining the crystal structure, even those not included in the GA global search.

Due to computational intractability as well as issues related to the proper description of the dispersion forces by DFT methods, [19-23] most of the work in CSP has been limited to using empirical force fields to calculate both inter- and intra- molecular interactions. A great deal of work has been done to improve the completeness and accuracy of force field descriptions by modeling the electrostatic interactions. In addition, improvements have been made to increase the speed of the necessary calculations. [24-27] Brodersen and his colleagues [28] tested distributed multipole models for evaluating electrostatic interaction between atoms in force field calculations. The methods were applied to large scale test sets and the results were compared to experiment. Neumann *et al.* [8, 29-31] presented a novel force field approach based on the detailed fitting of energies calculated using their dispersion corrected DFT calculations. [29] Similar work has been reported recently by Misquita *et al.* [32] for the prediction of the structure of 1,3-dibromo-2-chloro-5-fluorobenzene. In these approaches individual force fields have to be developed for each molecule; this is a time consuming and labor intensive process [33] that has been very successful, but it is unclear how they can be applied to any high throughput studies.

Our method, described in great detail in our previous publications [14-18] uses a standard force field (the general AMBER force field, known as GAFF) [34] and while

it is as computationally demanding as other methods in the literature, MGAC requires much less human labor. While this makes our method suitable for high throughput studies, the use of standard potentials may, to some extent, limit its predictive capabilities. The assessment of these limitations when attempting to predict polymorphism of pharmaceutical drugs is the thrust of the research presented here.

In this paper we describe the use of MGAC to predict the known polymorphic structures bicalutamide. The paper describes our success in finding the lower energy polymorph and the difficulties encountered in finding the second one.

2   Computational Methods

A full and more detailed description of the MGAC methods is given in Refs. [14-18] All of the energy calculations and local optimization were performed using CHARMM [35, 36] with the GAFF [34] parameters and RESP charges. [37, 38] These charges were calculated using the optimized HF/6-31G* geometries obtained when the experimental conformation was used as the starting one. But our previous work shows that these charges do not significantly depend on the molecular conformation. [17] A cutoff of 14 Å was used to compute short range non-bonded interactions and the Ewald technique was then applied to calculate the electrostatic interactions including at least two unit cells in the simulation box in every direction. While the crystal structures of the compound studied here are known, the calculations were done as if performing a blind test, i.e., no information of the experimental structure was used *a priori* in our calculations. A series of ten MGAC runs for each of the 14 most common space groups in organic molecules (P_1, P_{-1}, P_{21}, C_2, P_c, C_c, $P_{21/c}$, $C_{2/c}$, P_{212121}, P_{ca21}, P_{na21}, P_{bcn}, P_{bca}, and P_{nma}) were completed, with five done on structures with one molecule per asymmetric unit and five on structures with two molecules per asymmetric unit. The parameter values describing the initial population are randomly selected, including the dihedral angles considered in the global optimization. Each GA run produced 130 generations with 30 crystal structures each, using a crossover probability of 1.0 and a mutation probability of 0.001. This process generates approximately 500,000 structures for each compound studied here. To generate these structures requires running approximately 140 independent optimizations, each one taking between 12 to 72 hrs on 14 processors; this translates in a total of 23,000 to 140,000 processor hours. After these initial runs, which found only the P_{-1} polymorph using our standard search protocol, [17, 18] we performed numerous more detailed searches for the second $P_{21/c}$ polymorph that are described in the text below.

3   Results

The standard run of the MGAC protocol, described above, produced a list of 300 structures with energies ranging from -135.73 Kcal/mol to -129.47 Kcal/mol, that were compared with the experimental structures using the COMPACK methodology as implemented in COSET [39] with the standard settings. This comparison gives a very good match with the structure ranked #13 in the list. The pictorial comparison between the MGAC structure and the experimental one is depicted in Fig. 1. The comparison between the cell parameters of these structures is given in Table 1.
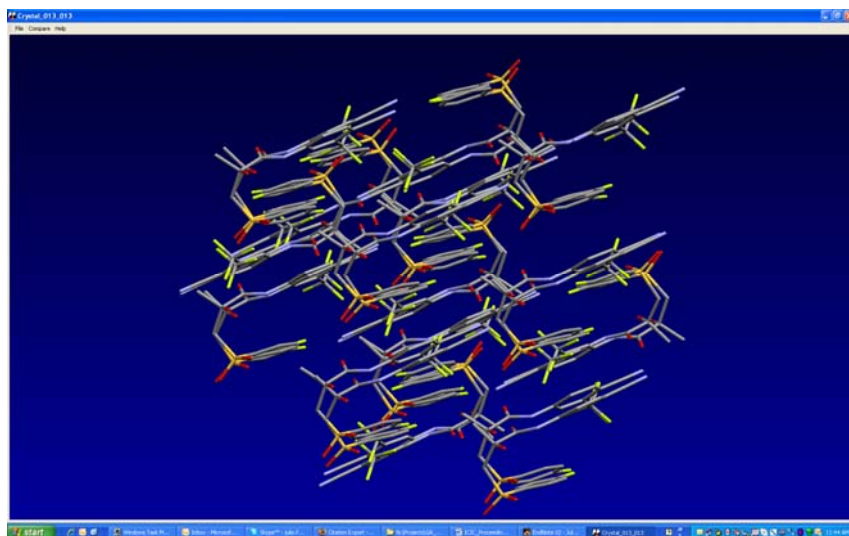
Fig. 1. Comparison of the experimental (P$_{-1}$ polymorph) and predicted structure of the bicalutamide (ranked #13)

Table 1. Comparison between the predicted (top row) and experimental (P$_{-1}$ polymorph) and cell parameters of bicalutamide

rms Å	a Å	b Å	c Å	α	β	γ	volume Å^3
0.642	7.494	10.666	11.434	89.93	77.13	80.11	877.083
	7.820	11.060	11.324	88.19	77.03	77.96	933.541

Clearly both the table and the figure show an excellent agreement between the predicted and experimental (P$_{-1}$ polymorph) structures, with both structures belonging to the P$_{-1}$ symmetry group and good showing agreement among the cell parameters. The predicted volume is approximately 6% smaller than the experimental one; this is consistent with our previous studies in flexible molecules and is arguable due to the neglect of thermal averaging in the calculated values. The results presented above are consistent with our previous experiences, [18] showing that when MGAC finds a good match of the experimental structure the quantitative agreement between the structures is remarkable good.

Unfortunately our standard search was not able to find the other known bicalutamide polymorph which belongs to the P$_{21/c}$ symmetry group. To determine the reasons for this failure it is much more challenging. There are essentially two reasons why MGAC cannot find a polymorph. Either the empirical potential function used to calculate the crystal energy is not an accurate representation of the molecular energetics or the GA method is not able to find the experimental configuration. In our previous studies the former cause was responsible for most of the failures in the MGAC searches, but this is not the case for bicalutamide. The locally optimized (using

CHARMM and the GAFF force field) structures of the known polymorphs of bicalu-tamide, reported in the CCDC as JAYCES [4] or JAYCES01 [3] and JAYCES02, [3] have energies of -133.62 kcal/mol and -134.34 kcal/mol respectively, which are both well within the range of energies of the structures obtained in the standard search and included in the list used to find matches with the experimental structures. It is impor-tant to note that these local optimizations do not change in any significant way the crystal structure from the starting experimental one, with rms between the experi-mental and locally optimized structures of 0.462 Å for JAYCES and 0.642 Å for JAYCES02.

To further enhance our searches for the second polymorph ($P_{21/c}$ symmetry and open conformation) we conducted MGAC searches in the $P_{21/c}$ symmetry group in-cluding in the GA search all dihedral angles and also restricting them to those of the molecular conformation observed in the $P_{21/c}$ symmetry polymorph. These searches produced lists of structures with the best 1000 energies ranging from -135.73 Kcal/mol to -129.47 Kcal/mol and -134.79 Kcal/mol to -128.35 Kcal/mol, respec-tively. Clearly the energy of the second polymorph belonging to the $P_{21/c}$ symmetry is well within the ranges of energies explored by the MGAC searches, but the method was not able to find the $P_{21/c}$ symmetry polymorph using our standard GA parameters. Following this finding, we ran a series of more exhaustive searches concentrating in the $P_{21/c}$ and P_{21} symmetry groups and changing different parameters in the GA op-timizations. The searches in the last group were motivated by our previous findings that it is common to find structures in P_{21} searches that belong to the more restrictive $P_{21/c}$ group. We have hypothesized that searches in a very restricted symmetry group may be more difficult because they may have a smaller attractive basin. The GA pa-rameters considered for changes were the mutation rate, the population size and the number of generations and the results from these searches are presented in Table 2.

The results in that table clearly show that more aggressive mutation rates, in-creases in the population size and number of generations do not lower the energy of the best physically plausible structure. In all cases the energy of the best physically feasible structure is higher than the energy of the locally optimized structure of the

Table 2. Comparison of the lowest energy of a physically plausible structure found by the MGAC searches using different GA optimization parameters, P: population size, G: number of generations, M: mutation rate. Only parameters that have been changed from the defaults are entered in column 1.

Simulation	Energy [a]	Rank [b]
Default (P=30, G=130, M=0.001)	-129.61	3
M=0.01	-126.09	10
P=100, G=500	-130.68	6
P= 200, G=1000	-130.67	2
P=100, G=2749	-130.74	8
P=100, G=1000 (P_{21})	-128.89	27

[a] Energy in Kcal/mol.
[b] Rank specifies the rank of the best structure that is physically feasible within the overall list of structures.

$P_{21/c}$ polymorph, -133.62 Kcal/mol. The evolution of the energy in the GA optimization as function of the generation for a $P_{21/c}$ run with P= 200 N= 1000 is depicted in Fig. 2. Clearly the figure shows that the GA converges very quickly and after 100 generations there are very few new lower energy structures entering into the population. Unfortunately, these new structures entering the population correspond to non-physical crystals (see Table 2).

The appearance of non-physical structures in the population is an artifact introduced by the use of empirical potentials. The empirical potentials are designed to reproduce the energy in the proximity of the equilibrium geometry, but because the potentials are expressed as continuous functions they can assume very low values for physically non-plausible structures. This is an artifact that introduces spurious structures into the population that cannot be eliminated when using only the energy of the structure as the single optimization objective.
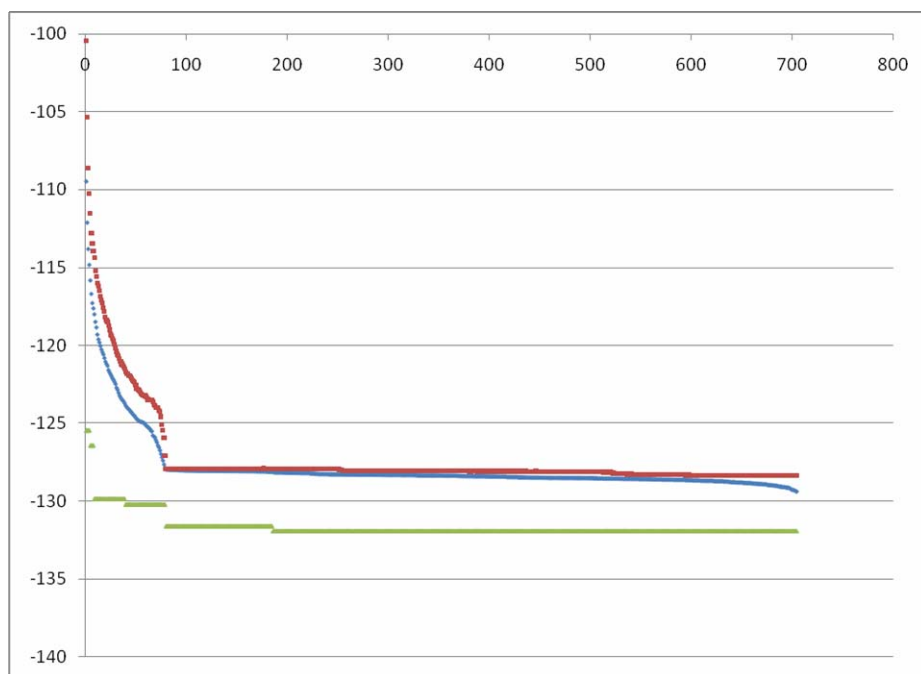
Fig. 2. Evolution of the best (green) average (blue) and worst (red) energy in the GA population for the P_{21c} run with P= 200 N= 1000

The second alternative that we pursued to intensify the search of the $P_{21/c}$ polymorphs was to perform a large number of MGAC runs in the $P_{21/c}$ and P_{21} space groups. We performed 50 MGAC runs for each symmetry group, with the standard mutation rate, a population of 30 and for 75 generations. The results for the $P_{21/c}$ runs are depicted in Fig. 3. The figure show that different runs have very different rates of convergence, but with most of the runs reaching values within a few Kcal/mol of the
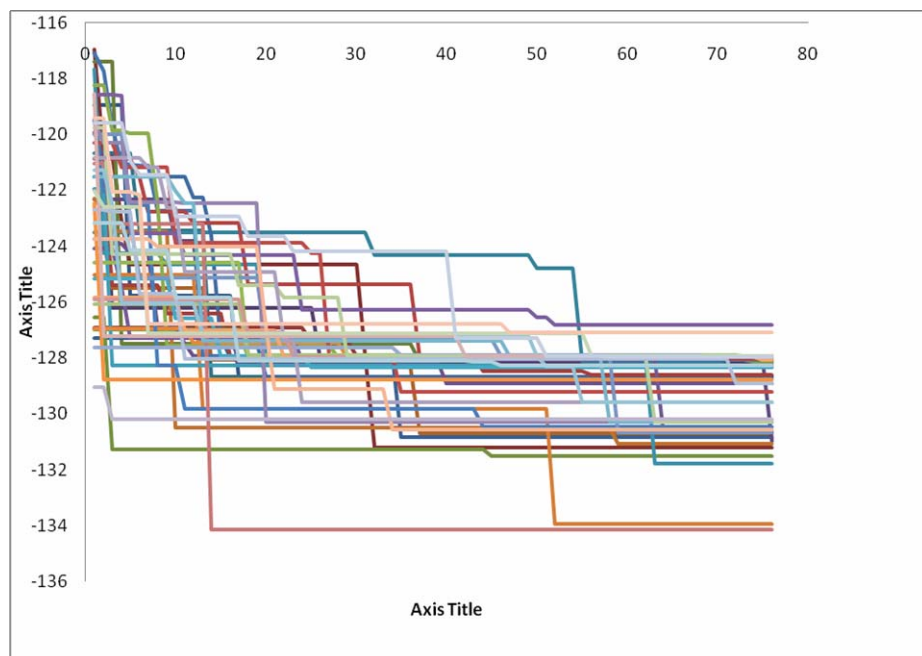
Fig. 3. Evolution of JAYCES P21/c P=30, G= 75. The figure shows the best energy for 50 different runs as a function of the generation.

experimentally known $P_{21/c}$ structure (-133.62 Kcal/mol). Unfortunately, the analysis of the structures with the 300 lowest energies from all these 50 runs does not contain any good match of the experimental one and contains many physically not plausible structures.

4   Conclusions

Our study shows that it is possible to find good matches between predicted and experimental crystal structures of flexible molecules using a standard force field. But it also shows that it is significant more challenging to find other higher energy polymorphs. The techniques pursued here to find the $P_{21/c}$ polymorph were not successful, and the results suggest that increasing the effectiveness of the GA search has the undesirable consequence of bringing more physically non-plausible structures into the population. In this sense the GA is performing extraordinary well in finding very unusual structures that according to the merit function, the crystal energy, are highly completive. Clearly a single merit function is not sufficient because the energy given by the empirical potential function selects physically non-plausible structures. We believe that to correct this problem it will be advantageous to use a multi-objective approach to the GA optimization. A careful analysis of the physically non-plausible structures shows that they have unit cell volumes much larger than those from the experimental crystals; this finding makes the unit cell volume a clear choice for a

second optimization target. This is quite appealing because even in absence of experimental data the unit cell volume can be estimated using the empirical rules given by Ammon. [40]

Acknowledgements

References

1. Schellhammer, P.F.: An Evaluation of Bicalutamide in the Treatment of Prostate Cancer. Expert Opinion on Pharmacotherapy 3, 1313–1328 (2002)
2. Muderris, I., Bayram, F., Ozçelik, B., Güven, M.: New Alternative Treatment in Hirsutism: Bicalutamide 25 mg/day. Gynecological Endocrinology 16, 63–66 (2002)
3. Vega, D.R., Polla, G., Martinez, A., Mendioroz, E., Reinoso, M.: Conformational Polymorphism in Bicalutamide. International Journal of Pharmaceutics 328, 112–118 (2007)
4. Hu, X.R., Gu, J.M.: N-[4-Cyano-3-(trifluoromethyl)phenyl]-3-(4-fluorophenylsulfonyl)-2-hydroxy-2-methylpropionamide. Acta Crystallographica Section E 61, 3897–3898 (2005)
5. Dunitz, J.D., Bernstein, J.: Disappearing Polymorphs. Acc. Chem. Res. 28, 193–200 (1995)
6. Threlfall, T.L.: Analysis of Organic Polymorphs. A Review. Analyst (Cambridge, United Kingdom) 120, 2435–2460 (1995)
7. Erk, P., Hengelsberg, H., Haddow, M.F., Gelder, R.V.: The Innovative Momentum of Crystal Engineering. CrstEngComm 6, 474 (2004)
8. Day, G.M., et al.: A Third Blind Test of Crystal Structure Prediction. Acta Crystallogr., Sect. B: Struct. Sci. 61, 511–527 (2005)
9. Lommerse, J.P.M., Motherwell, W.D.S., Ammon, H.L., Dunitz, J.D., Gavezzotti, A., Hofmann, D.W.M., Leusen, F.J.J., Mooij, W.T.M., Price, S.L., Schweizer, B., Schmidt, M.U., Eijck, B.P.V., Verwer, P., Williams, D.E.: A Test of Crystal Structure Prediction of Small Organic Molecules. Acta Cryst. B 56, 697 (2000)
10. Day, G.M., Motherwell, W.D.S., Ammon, H.L., Boerrigter, S.X.M., Della Valle, R.G., Venuti, E., Dzyabchenko, A., Dunitz, J.D., Schweizer, B., van Eijck, B.P., Erk, P., Facelli, J.C., Bazterra, V.E., Ferraro, M.B., Hofmann, D.W.M., Leusen, F.J.J., Liang, C., Pantelides, C.C., Karamertzanis, P.G., Price, S.L., Lewis, T.C., Nowell, H., Torrisi, A., Scheraga, H.A., Arnautova, Y.A., Schmidt, M.U., Verwer, P.: CSP workshop at Cambridge UK. Acta Cryst. B-STRUCTURAL SCIENCE 65, 107–125 (2009)
11. Goldberg, D.E.: Genetic Algorithms in Search, Optimization and Machine Learning. Addison-Wesley, New York (1989)
12. Man, K.F., Tang, K.S., Kwong, S.: Genetic Algorithms. Springer, Berlin (1999)
13. Niesse, J.A., Mayne, H.R.: Global Optimization of Atomic and Molecular Clusters Using the Space-fixed Modified Genetic Algorithm Method. J. Comput. Chem. 18, 1233 (1997)

14. Bazterra, V.E., Ferraro, M.B., Facelli, J.C.: Modified Genetic Algorithm to Model Crystal Structures. I. Benzene, Naphthalene and Nnthracene. J. Chem. Phys. 116, 5984–5991 (2002)
15. Bazterra, V.E., Ferraro, M.B., Facelli, J.C.: Modified Genetic Algorithm to Model Crystal Structures. II. Determination of a Polymorphic Structure of Benzene Using Enthalpy Minimization. J. Chem. Phys. 116, 5992–5995 (2002)
16. Bazterra, V.E., Ferraro, M.B., Facelli, J.C.: Modified Genetic Algorithm to Model Crystal Structures. III. Determination of Crystal Structures Allowing Simultaneous Molecular Geometry Relaxation. Int. J. Quantum Chem. 96, 312–320 (2004)
17. Bazterra, V.E., Thorley, M., Ferraro, M.B., Facelli, J.C.: A Distributed Computing Method for Crystal Structure Prediction of Flexible Molecules: An Application to N-(2-Dimethyl-4-5-dinitrophenyl) Acetamide. J. Chem. Theory and Comp. 3, 201–209 (2007)
18. Kim, S., Orendt, A.M., Ferraro, M.B., Facelli, J.C.: Crystal Structure Orediction of Flexible Molecules Using Parallel Geneic Algorithms with Standard Force Field. J. Comp. Chem. (in press, 2009)
19. Axel, D.B.: Density-functional Thermochemistry. III. The Role of Exact Exchange. The Journal of Chemical Physics 98, 5648–5652 (1993)
20. Kohn, W., Sham, L.J.: Self-Consistent Equations Including Exchange and Correlation Effects. Physical Review 140, A1133 (1965)
21. Lee, C., Yang, W., Parr, R.G.: Development of the Colle-Salvetti Correlation-energy Formula into a Functional of the Electron Density. Physical Review B: Condensed Matter and Materials Physics 37, 785–789 (1988)
22. Stephens, P.J., Devlin, F.J., Chabalowski, C.F., Frisch, M.J.: Ab Initio Calculation of Vibrational Absorption and Circular Dichroism Spectra Using Density Functional Force Fields. J. Phys. Chem. 98, 11623–11627 (1994)
23. Ziegler, T.: Approximate Density Functional Theory as a Practical Tool in Molecular Energetics and Dynamics. Chem. Rev. 91, 651–667 (1991)
24. Besler, B.H., Merz, K.M., Kollman, P.A.: Atomic Charges Derived from Semiempirical Methods. Journal of Computational Chemistry 11, 431–439 (1990)
25. Coombes, D.S., Price, S.L., Willock, D.J., Leslie, M.: Role of Electrostatic Interactions in Determining the Crystal Structures of Polar Organic Molecules. A Distributed Multipole Study. J. Phys. Chem. 100, 7352–7360 (1996)
26. Stone, A.J., Alderton, M.: Distributed Multipole Analysis. Molecular Physics 56, 1047–1064 (1985)
27. Williams, D.E.: Representation of the Molecular Electrostatic Potential by Atomic multipole and Bond Dipole Models. Journal of Computational Chemistry 9, 745–763 (1988)
28. Brodersen, S., Wilke, S., Leusen, F.J.J., Engel, G.: A Study of Different Approaches to the Electrostatic Interaction in Force Field Methods for Organic Crystals. Physical Chemistry Chemical Physics 5, 4923–4931 (2003)
29. Neumann, M.A., Perrin, M.A.: Energy Ranking of Molecular Crystals Using Density Functional Theory Calculations and an Empirical van der Waals Correction. J. Phys. Chem. B 109, 15531–15541 (2005)
30. Neumann, M.A.: Crystal Structures of Moderately Complex Organic Molecules are Predictable. In: 24th European Christallographic Meeting, Micro Symposium 14, Advanced computational methods in structural chemistry, Marrakech, Morocco, pp. 11H00–11H20 (2007)
31. Neumann, M.A., Leusen, F.J.J., Kendrick, J.: A Major Advance in Crystal Structure Prediction. Angew. Chem. Int. Ed. 47, 2427–2430 (2008)

32. Misquitta, A.J., Welch, G.W.A., Stone, A.J., Price, S.L.: A First Principles Prediction of the Crystal Structure of C6Br2ClFH2. Chem. Phys. Lett. 456, 105–109 (2008)

33. Price, S.L.: Quantifying Intermolecular Interactions and Their Use in Computational Crystal Structure Prediction. Cryst. Eng. Comm. 6, 344–353 (2004)

34. Wang, J., Wolf, R.M., Caldwell, J.W., Kollman, P.A., Case, D.A.: Development and Testing of a General Amber Force Field. J. Comput. Chem. 25, 1157 (2004)

35. Brooks, B.R., Bruccoleri, R.E., Olafson, B.D., States, D.J., Swaminathan, S., Karplus, M.: CHARMM: A Program for Macromolecular Energy, Minimization, and Dynamics Calculations. J. Comp. Chem. 4, 187–217 (1983)

36. MacKerell, A.D., Brooks, J.B., Brooks III, C.L., Nilsson, L., Roux, B., Won, Y., Karplus, M.: CHARMM: The Energy Function and Its Parameterization with an Overview of the Program. In: Schleyer (ed.) The Encyclopedia of Computational Chemistry, pp. 271–277. John Wiley & Sons, Chichester (1998)

37. Bayly, C.I., Cieplak, P., Cornell, W., Kollman, P.A.: A Well-behaved Electrostatic Potential Based Method Using Charge Restraints for Deriving Atomic Charges: the RESP Model. J. Phys. Chem. 97, 10269–10280 (1993)

38. Cornell, W.D., Cieplak, P., Bayly, C.I., Kollman, P.A.: Application of RESP Charges to Calculate Conformational Energies, Hydrogen Bond Energies, and Free Energies of Solvation. J. Am. Chem. Soc. 115, 9620–9631 (1993)

39. Chisholm, J.A., Motherwell, S.: COMPACK: a Program for Identifying Crystal Structure Similarity Using Distances. Journal of Applied Crystallography 38, 228–231 (2005)

40. Ammon, H.L.: Updated Atom/Functional Group and Atom_Code Volume Additivity Parameters for the Calculation of Crystal Densities of Single Molecules, Organic Salts, and Multi-Fragment Materials Containing H, C, B, N, O, F, S, P, Cl, Br, and I. Propellants, Explosives, Pyrotechnics 33, 92–102 (2008)

CAPS Genomic Subtyping on *Orthomyxoviridae*

Sheng-Lung Peng[1],[*], Yu-Wei Tsay[1], Chich-Sheng Lin[2], and Chuan Yi Tang[3]

[1] Department of Computer Science and Information Engineering
National Dong Hwa University, Hualien 974, Taiwan
slpeng@mail.ndhu.edu.tw
[2] Department of Biological Science and Technology
National Chiao Tung University, Hsinchu 300, Taiwan
[3] Department of Computer Science
National Tsing Hua University, Hsinchu 300, Taiwan

Abstract. The *Orthomyxoviridae* is a family of single strained RNA viruses including five genera: Influenza virus A, Influenza virus B, Influenza virus C, Thogotovirus, and Isavirus. Usually, Influenza viruses are identified by antigenic differences in their nucleoprotein and matrix protein. In this paper, we propose an algorithm to determine a set of suitable restriction enzymes for producing recognizable restriction maps on *Orthomyxoviridae*. Our method is applied to viral strains of highly pathogenic avian influenza (HPAI), containing potentially homozygous, heterozygous, and various genetic variations. In the analysis of CAPS (Cleaved Amplified Polymorphic Sequence) subtyping, our method outperforms the RNA coding of representative and epidemiologically significant human wild-type viruses, including H3N8, H5N1, H5N9, H7N1, H7N7, and H9N2. These isolates are analyzed by CAPS with enzymes AgeI, EciI, KpnI, and XbaI. The HPAI strains show a different RFLP (Restriction Fragment Length Polymorphism) profile by comparing with other low pathogenic avian influenza (LPAI) strains. We provide a rapid, specific, and reproducible identification of the genotypes on *Orthomyxoviridae*. It permits us to quickly confirm subtypes of *Orthomyxoviridae*.

1 Introduction

Molecular techniques are major tools for the analysis of microorganisms from biological substances. The functions of these techniques are amplification, separation, detection, and expression. Such techniques provide ways to screen for a broad range of microorganisms in a single test. Analytical laboratories use these techniques for rapid differentiation of species, strain identification, and annotation of related strains from infected samples. Molecular methods vary due to discriminatory power, reproducibility, ease to use, and competitive price.

The CAPS technique, known as PCR-RFLP markers, provides a way to utilize the DNA sequences of mapped RFLP markers and eliminate tedious blotting.

---

[*] Corresponding author.

The CAPS is performed by digesting locus-specific PCR amplification with restriction enzymes, and then separating the digested DNA on agarose or polyacrylamide gels [1]. Comparing to RFLP, CAPS markers have an extra advantage that is particularly useful to detect and classify microorganisms when clinical samples are insufficient.

Despite of the improved procedure in CAPS markers, there are still some conditions needed to be noticed in using CAPS markers. First, CAPS requires developed PCR primers as genetic markers deriving from target sequences. The PCR products are corresponding to the amplified bands produced by specific primers. Then the amplified bands should be distinguishable after cleavage with restriction enzymes. Second, that the nucleotide changing affects restriction sites is essential for the detection of DNA polymorphism, including DNA mutations disrupted or a restriction enzyme recognition site determined. Third, after the digestion of several restriction enzymes, the resulting fragments are separated in gel electrophoresis of agarose gel or polyacrylamide gel. From the above statements of CAPS markers, we may infer a practical procedure of CAPS technique in the following criteria:

– **A well-designed primer pair** - The selectivity in discrimination of each product is significant. Regardless of mutation or single nucleotide polymorphism (SNP) on particular digestion site, the polymorphic nucleotides are still available by restriction enzymes. It is also called *derived cleaved amplified polymorphic sequence* (dCAPS) [2].
– **A well-designed restriction enzyme set** - In this case, it is only acquired a proper template DNA ranged below 100k base pairs in the procedure of enzyme digestion. The requisite DNA sequence is first amplified by PCR undoubtedly. With the help of suitable restriction enzyme set, the presence or absence of gel analysis maps facilitates to differentiate allelic divergence.

In this paper, we consider the selection of restriction enzymes in the procedure of CAPS. Due to the bias of genotype frequencies, population heterozygosities, degree of inbreeding, population subdivision, and degree of individual relatedness, these issues are often challenged for the validity of CAPS markers. Therefore, the proposed method will be applied to viral strains, containing potentially homozygous, heterozygous, and various genetic variations. For the differentiation of all alleles with specific enzymes, a measure of genetic diversity is given to evaluate the utility for subtyping of CAPS markers.

According to the molecular data, each fragment obtained by enzymes is transferred to a boolean vector (0,1). Then the measurement of similar set is analyzed qualitatively. Suppose X and Y are two relative species requiring to quantify genetic relationships, Π_X and Π_Y measure the nucleotide diversity among several sequences in a given region of genome within a population, V_{XY} measures population divergence [3] based on the degree of sequence variation. Then, the measure of genetic diversity gives a criterion of our proposed rule in enzyme selection.

2   Methods

In this section, we propose an algorithm for differentiating two populations by using a particular enzyme set. Our problem and method can be stated in the following subsections.

2.1   Problem Formation

Let $G = \{g_1, g_2, \ldots, g_n\}$ be a set of considered genomes and $Z = \{z_1, z_2, \ldots, z_k\}$ be the set of possible enzymes. Let M be the matrix where M_{ij} records the information of the digestion sites of enzyme z_i on genome g_j. Let $\{X, Y\}$ be a partition of G. That is, $X \cup Y = G$ and $X \cap Y = \emptyset$. For convenience, we use $G = X + Y$ to denote that $\{X, Y\}$ is a partition of G. Given $G = X + Y$, Z, M, and a threshold α, the *CAPS subtyping problem* is to find a minimum subset $U \subseteq Z$ such that X and Y can be distinguished above a requested nucleotide diversity α according to their digestion maps by using enzymes of U.

Table 1. Example of CAPS digestion pattern M

$Z \setminus G$	H1N1	H2N3	H6N6	H9N1	H5N1	H7N1	H9N2
size	13,579	13,622	13,604	13,598	13,000	13,315	13,597
AatI	557,+	581,+	740,+	740,+	715,+	756,+	740
AgeI	416	2145	442,+	1523,+	361,+	371,+	681
BglII	1033,+	51,+	51,+	51,+	429,+	469,+	136,+
BspDI	1178,+	947	1250	661	0	1374	541
PctI	82,++	111,+	111,+	209,++	177,+	219,+	111,+
SalI	176	0	443	0	0	0	0

In this table, 'size' indicates the length of a genome g, and each entry (z_i, g_j) is a linked list that stores the binding sites of z_i on g_j. For simplicity, we use '+' to denote if the number of digestion sites is between 2 and 5, and '++' if it is more than 5.

For example, let $G = \{$H1N1, H2N3, H5N1, H6N6, H7N1, H9N2, H9N4$\}$ and $Z = \{$AatI, AgeI, BglII, BspDI, PctI, SalI$\}$. Let $X = \{$H1N1, H2N3, H6N6, H9N4$\}$ and $Y = \{$H5N1, H7N1, H9N2$\}$ be a partition of G. Table 1 shows an example of the matrix M. In this table, enzyme AgeI can digest H1N1 on the location 1541. In the problem of genomic subtyping, an enzyme set U is expected to generate recognizable CAPS patterns from group X to group Y. However, such enzyme set may not easily be found in general case for CAPS subtyping, especially for those pathogenetic relative strains. For practice, we use a heuristic for selecting enzymes. Concerning to genome set X, if enzyme z can digest a genome g in X but no genome in Y, then it will be preferred. For this purpose, we use an array $N_X[k]$ (respectively, $N_Y[k]$) derived from matrix M such that for each i $N_X[i]$ (respectively, $N_Y[i]$) stores the number of digestion sites of enzyme z_i on genomes of X (respectively, Y). For example, N_X may be obtained as [149, 124, 85, 63, 39, 28] and [5, 56, 37, 10, 0, 5] for N_Y. In this case, enzymes that cannot digest genomes of X will not be considered. On the

other hand, if we consider Y first, then enzymes that cannot digest genomes of Y will not be considered. In such a manner, difference of N_X and N_Y enlarges their variance of digestions.

Once each possible enzyme set has been found, it is required to verify the similarity between X and Y. Therefore, a criterion is needed to determine which enzyme is applicable for simplified CAPS maps.

2.2   CAPS Rating Criteria

To determine which enzyme set is suitable to distinguish the target set is a time-consuming work, even gel electrophoresis with a naked eye. Each gel map is produced due to the lengths of fragments after digested by an enzyme set. Hence, a quantitative criterion is needed to determine which enzyme is applicable for simplified gel maps. In brief, we need to normalize each gel map and check their similarity by given enzyme set. The rating criterion is given as follows.

Let V be a set of CAPS fragment lengths on some genome g generated by enzyme set U. Let *CAPS Pattern Vector of V with respect to U* (abbreviated as C-Vector(U,V)) be a boolean vector of CAPS pattern V by enzyme set U. Let L be the degree of fineness according to CAPS fragments.

From the point of view of macrography, CAPS gel patterns must be centralized at some specific areas, or simply distributed at few locations. As the fineness L goes smaller, the size of CAPS pattern vector may raise. For example, suppose that 10k bp is the maximum possible length. Let $L = 0.2$. Then we can obtain a C-Vector of size 5 with the first position considering fragments of length smaller or equal to 2k bp, and so on. For example, if strain g can be digested into fragment set V with fragment lengths of 0.3k, 0.7k, 2k, 3k, and 4k by U. Then C-Vector$(U,V) = [1,1,0,0,0]$. For the consideration of fineness L, we simply intersect the CAPS Pattern Vector of V and ignore those vector values are equal to 0. In this manner, the distance between two CAPS patterns can be measured by some distance function, *e.g.*, Euclidean distance, one of the most common used distance function. The shorter distance of C-Vector, the similar gel maps they might have.

2.3   Measurement of Selection Criterion

Nucleotide diversity is a concept in molecular genetics measuring the degree of polymorphism within a population [4]. It estimates the number of nucleotide variations per digestion site between two isolates. In [3], a refined formula is proposed for this purpose. In this paper, we use the formula to determine whether the selected enzyme is good or not.

Intrapopulation nucleotide diversity. Π_X measures the variations among several strains, defined as follows:

$$\Pi_X = \frac{1}{r}lnH,\ H = F(3 - 2H^o)^{\frac{1}{4}},\ \text{and}\ F = \frac{\sum X_i(X_i \cdot n - 1)}{\sum X_i(n - 1)}$$

where r is the number of recognition nucleotides of restriction enzyme, $\ln H$ is a natural logrithm of the probability that was no substitution in the digestion site, F is propotion of shared fragments, H^o is double radical of F, n is the number of halploid genotypes in population X, and X_i estimates frequency of ith fragment in the population.

This measure calculates the diversity for each restriction site, varying from 0 to 1. Mostly, the equation can be used with haploid, mDNA, cDNA, or haplotypes.

Interpopulation nucleotide diversity V_{XY} calculates the divergence among populations X and Y based on restriction data [3], defined as follows:

$$V_{XY} = d_{XY} - \frac{(\Pi_X + \Pi_Y)}{2}, \; d_{XY} = -\frac{2}{r}ln(H_{XY}) \; H_{XY} = F_{XY}(3 - 2H^o_{XY})^{\frac{1}{4}},$$

and $F_{XY} = \frac{2\sum X_{iX}X_{iY}}{\sum (X_{iX}+X_{iY})}$

where d_{XY} is the fragment diversity among two populations, H^o is double radical of F, and X_{iX} estimates frequency of the ith fragment in population X.

In our CAPS simulation, the value of r is replaced by the most enzyme length, *i.e.*, 6. Finally, by using these formula, we propose an algorithm for CAPS enzyme selection problem for genomic subtyping. The detail of our algorithm is shown in **Algorithm 1**. For simplicity, in the algorithm, we use $S + s$ to denote $S \cup \{s\}$ for a set S and a singleton s.

Algorithm 1. A Heuristic Algorithm for CAPS Subtyping

1. **Input:** two genome sets X and Y, a threshold L, a preprocessing enzyme set Z and corresponding arrays N_X and N_Y obtained from matrix M
2. **Output:** an enzyme set U
3. $K = \emptyset, U = \emptyset, U_1 = \emptyset, T = \emptyset, \alpha = 0, \alpha_1 = 0, loop = 2;$
4. **while** $loop \neq 0$ **do**
5.   **for** $i = 1$ to $|N_X|$ **do**
6.   **if** $N_X[i] - N_Y[i] > L$ **then**
7. $K = K + z_i;$
8.   **end if**
9.   **end for**
10. $U = \{k_1\}$
11. $\alpha = V_{XY}(X,Y,U)$
12.   **for** $i = 2$ to $|K|$ **do**
13.   **if** $V_{XY}(X,Y,U + k_i) > \alpha$ **then**
14. $U = U + k_i;$
15.   **end if**
16.   **end for**
17. $loop = loop - 1, T = X, X = Y, Y = T, K = \emptyset, U_1 = U, U = \emptyset, \alpha_1 = \alpha$
18. **end while**
19. **if** $\alpha < \alpha_1$ **then**
20. $U = U_1$
21. **end if**
22. **return** U

3 Results

According to the OIE (World Organisation for Animal Health) [5], avian influenza viruses that are studied in order to control poultry disease include the following subtype: H5N1, H7N3, H13N6, H5N9, H3N8, H9N2, H7N7, H7N1, and others. Recently, H5N1 avian influenza viruses have been found to cross host barriers and cause human infections [6]. The *Orthomyxoviridae* genome is about 12k-15k nucleotides, containing 7 to 8 segments of linear negative-sense single stranded RNA. Due to the limitation of space, some of strains in *Orthomyxoviridae* are not selected as our material. In the experiment results, four of specific enzymes, namely, KpnI:GGTACC, EciI:GGCGGA, AgeI:ACCGGT, and XbaI:TCTAGA, are proposed to differentiate H5N1 from other influenza viruses. That is, we let $X = \{H5N1\}$ and Y be the set of LPAI. After finding a suitable enzyme set U, we then do a comparison between H5N1 and each interested strain of *Orthomyxoviridae*. Partial result is shown in Table 2. In the table, three enzyme sets demonstrate each population diversity among H5N1 (X) and an Influenza A (Y). That is, the value V_{XY}^a obtained from our proposed enzyme set shows a frequent variation between H5N1 and *Orthomyxoviridae*. In the diversity measurement, proportion of shared alleles cannot be found. This reason leads to the effect of high divergence among H5N1 and *Orthomyxoviridae*.

For a comparison with other frequently used enzymes, we try another two sets of restriction enzymes. Table 2 shows that these two enzyme sets fail to identify H5N1 from *Orthomyxoviridae*. It shows that our proposed restriction enzymes have high specificity. Obviously, the nucleotide differentiation based on V_{XY}^b (GsuI, TspEI, BtrI, and ChaI) gives an unacceptable outcome of their nucleotide diversity, even in V_{XY}^c (HphI, RsaI, BseRI, and CfoI). In V_{XY}^a, result of population diversity is incompatible with V_{XY}^b and V_{XY}^c, due to the nucleotide differentiation on variant restriction sites. In this case of *Orthomyxoviridae*
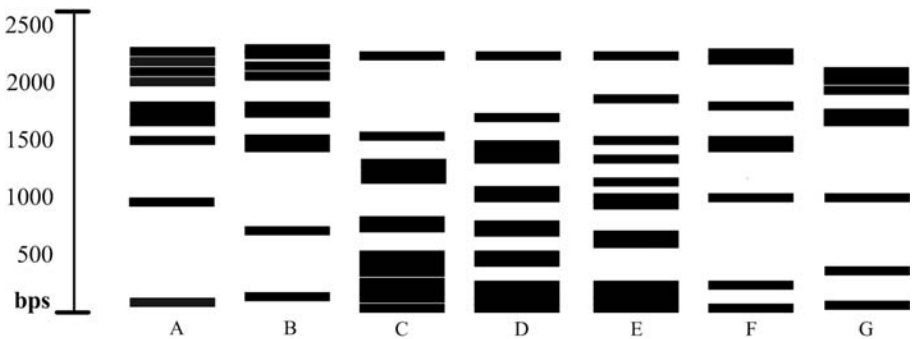
Fig. 1. The simulated gel maps of our proposed method on HPAI and LPAI with enzyme set {AgeI, EciI, KpnI, XbaI}. Each column indicates specific *Orthomyxoviridae*: **A**-H1N5, **B**-H3N3, **C**-H5N1, **D**-H7N7, and **E**-H9N2 are type-A Influenza. **F**-Taiwan/1484/01 is type-B Influenza. **G**-Ann Arbor/1/50 is type-C Influenza.

Table 2. H5N1 subtyping by different enzyme sets

Strains	Locus	Year	V_{XY}^a	Π_Y^a	V_{XY}^b	Π_Y^b	V_{XY}^c	Π_Y^c
Influenza A								
H1N1	New York	2009	0.705	0.358	0.377	0.241	0.381	0.191
H1N5	New York	1978	0.968	0.353	0.356	0.181	0.361	0.125
H1N9	Ohio	1987	0.937	0.391	0.376	0.177	0.353	0.129
H3N3	Nanchang	1993	1.149	0.349	0.369	0.201	0.398	0.181
H3N6	Nanchang	2000	0.562	0.292	0.330	0.126	0.350	0.144
H3N8	Manitoba	2005	0.584	0.298	0.424	0.235	0.345	0.129
H4N3	Alberta	1990	1.175	0.350	0.376	0.191	0.418	0.181
H4N7	California	2007	0.992	0.326	0.377	0.241	0.381	0.191
H5N1	Shantou	2002	0.445	0.290	0.302	0.293	0.328	0.145
H5N3	Italy	2004	0.992	0.326	0.366	0.109	0.359	0.153
H5N9	Italy	1998	0.553	0.302	0.323	0.182	0.333	0.127
H6N1	Hong Kong	1977	0.955	0.353	0.374	0.199	0.415	0.154
H6N9	Alberta	1979	0.873	0.354	0.349	0.193	0.349	0.147
H7N1	Mongolia	2002	0.552	0.340	0.416	0.218	0.351	0.155
H7N2	New York	2005	0.861	0.355	0.406	0.276	0.355	0.156
H7N7	Victoria	1985	0.556	0.329	0.397	0.231	0.393	0.137
H9N1	Alberta	1983	0.839	0.387	0.434	0.139	0.356	0.161
H9N2	Hubei	2007	0.529	0.355	0.368	0.164	0.364	0.168
H10N1	Alberta	1995	0.980	0.325	0.340	0.147	0.345	0.135
H10N9	Hong Kong	1979	0.599	0.329	0.349	0.200	0.399	0.212
H11N3	Ohio	1993	0.591	0.380	0.401	0.221	0.341	0.107
H13N6	Maryland	1977	0.590	0.358	0.359	0.250	0.418	0.182
Influenza B								
Bangkok/143/94			0.730	0.387	0.383	0.241	0.402	0.156
Hong Kong/02/93			0.703	0.358	0.397	0.216	0.443	0.233
Taiwan/71523/07			0.774	0.383	0.431	0.296	0.400	0.191
Influenza C								
Ann Arbor/1/50			0.829	0.293	0.355	0.150	0.409	0.180
Johannesburg/1/66			0.953	0.189	0.300	0.060	0.375	0.104
JJ/50			∞	0.191	0.327	0.071	0.362	0.146

The column *Strains* indicates the taxonomy of full length RNA in each *Orthomyxoviridae*. Columns *Year* and *Locus* show the outbreak time and location of specific typing, respectively. V_{XY}^a demonstrates each nucleotide diversity and individual intrapopulation Π_Y^a with enzyme set {AgeI, EciI, KpnI, XbaI} measuring the diversity between H5N1 and each interested strain of HPAI (gray) and LPAI. Similarly, V_{XY}^b (with Π_Y^b) is for randomly selected enzyme set {GsuI, TspEI, BtrI, ChaI} and V_{XY}^c (with Π_Y^c) is for a most used enzyme set {HphI, RsaI, BseRI, CfoI}. The larger value is better for differentiation. Our result shows that our algorithm selects a better enzyme set for distinguishing H5N1 from *Orthomyxoviridae*.

subtyping, the nucleotide diversity of Π_Y displays similar results, revealing similar strains among influenza viruses. In summary, the proposed CAPS marker selection gives an appealing issue on H5N1 subtyping. A randomly selected enzyme set (for V_{XY}^b) and a frequently used enzyme set (for V_{XY}^c) contrast sharply with our proposed rule. The proposed CAPS subtyping method does not simply simplify the patterns in both input strains; instead, the genomic strains in one of the groups are fragmentary when V_{XY} is raising. Fig.1 shows some CAPS maps of various types of full-length RNA *Orthomyxoviridae* digested by our enzyme set. The major difference on gel maps of CAPS patterns is plenty of mass fragments with length below 500 bp. These significant patterns can be a recognizable sign between HPAI and LPAI. It also needs to be noticed that H10N9 (0.599) reveals a comparatively low variation in restriction distance. These subtypes may belong to highly pathogenic avian influenza virus.

4   Discussion

By studying several researches on the limitation of PCR-based markers, a method for CAPS enzyme selection is proposed to improve genomic subtyping. Due to the SNP or numerous reasons of nucleotide variations, we give a simple procedure for generating a minimum set of CAPS marker. Parameters of CAPS enzyme selection can be adjusted by demand, including number of genome sets, category of enzyme set, agarose gel, polyacrylamide gel, or pulsed field gel electrophoresis (PFGE) simulation. In the subtyping of HPAI and *Orthomyxoviridae*, results also give an encouraged validation for proposed method. In the future, the concept of CAPS enzyme selection may be extended to several ways on DNA fingerprint. In particular, we expect our method can be used in genetics, plant discrimination, forensics, and other fields.

References

1. Michaels, S.D., Amasino, R.M.: A robust method for detecting single-nucleotide changes as polymorphic markers by pcr. Plant J. 14(3), 381–385 (1998)
2. Neff, M.M., Neff, J.D., Chory, J., Pepper, A.E.: dcaps, a simple technique for the genetic analysis of single nucleotide polymorphisms: experimental applications in arabidopsis thaliana genetics. Plant J. 14(3), 387–392 (1998)
3. Karp, A., Isaac, P.G.: Molecular Tools for Screening Biodiversity: Plants and Animals. Springer, University of Bristol Long Ashton, UK (1998)
4. Nei, M., Li, W.H.: Mathematical model for studying genetic variation in terms of restriction endonucleases. Proc. Natl. Acad. Sci. U S A 76(10), 5269–5273 (1979)
5. WHO: Avian influenza a(h5n1)- update 31: Situation (poultry) in asia (epidemic and pandemic alert and response), http://www.who.int/csr/don/2004_03_02/en/
6. Subbarao, K., Klimov: Characterization of an Avian Influenza A (H5N1) Virus Isolated from a Child with a Fatal Respiratory Illness. Science 279, 393–396 (1998)

Verification of Pathotyping by Quasispecies Model

Sheng-Lung Peng* and Yu-Wei Tsay

Department of Computer Science and Information Engineering
National Dong Hwa University, Hualien 974, Taiwan
slpeng@mail.ndhu.edu.tw

Abstract. Discrimination using genetic diversity provides a significant support in genetic research and applications. Mostly, DNA markers indicate a process of determining the genotype presented at specific locations along the DNA molecule. Some developed DNA marker methods are RFLP, RAPD, AP-PCR, DAF, and AFLP. For these systems, enzymes play an important role. In this paper, we propose a mechanism to verify the enzyme efficacy for pathotyping. A procedure is given to inspect the validation on cleavage pattern by restriction enzymes, adapting the concept of genetic algorithm to quasispecies model – a genetic evolutionary processes of self-replicating macromolecules. The proposed mechanism is applied to viral strains of HPV (*Papillomaviridae*), including mutated strains from quasispecies model of homozygous, heterozygous, and various genetic variations. In the analysis of full length DNA strain PCR-RFLP subtyping, results showed that if digested patterns of HPV can be discriminated by specific enzyme set from non-high-risk and other papillomavirus, then it is also can be discriminated by the same enzyme set, under the condition of mutated simulation with quasispecies model. In addition, a measure of genetic diversity also evaluates the utility for PCR-RFLP markers in pathotyping, depending on the degree of digestion variation. We provide a specific and valid mechanism of examination on PCR-based pathotyping. Our approach offers a practical and verifiable direction for genomic pathotyping.

1 Introduction

Molecular techniques are major tools for the analysis of microorganisms from biological substances. Such techniques provide ways to screen for a broad range of microorganisms in a single test. Analytical laboratories have taken up the method for rapid differentiation of species, strain identification, and annotation of related strains from specific samples. Molecular methods vary according to discriminatory power, reproducibility, ease to use, and competitive price. DNA markers have now become a popular method for identification and authentication of plants, animals, bacteria, and viruses. Some of identification methods based on variations in genomic sequences are proposed, *e.g.*, RFLP (Restriction Fragment

---

* Corresponding author.

Length Polymorphism) [1], RAPD (Random Amplified Polymorphic DNA) [2], and AFLP (Amplified Fragment Length Polymorphism) [3].

These proposed methods provide a rapid tool for the detection of new taxa and mutations. For decades, there are a variety applications in pathogenic strains of subtyping. Four restriction enzymes (PstI, HaeIII, DdeI, RsaI) can be used to discriminate all individual mucosal HPV types in a single infection, and even detection of multiple infections [4]. In a study of [5], a new rapid method for pathotyping of H5 subtype avian influenza viruses is described which is based on RT-PCR and restriction enzyme cleavage pattern assay using the restriction enzyme MboII. As deduced from here, we may learn that the DNA profiling techniques in pathotyping analysis are practical.

However, due to the bias of genotype frequencies, population heterozygosities, degree of inbreeding, population subdivision, and degree of individual relatedness, one species may consist of several types of polymorphism, especially in properties of pathogenic and non-pathogenic. These issues are often challenged for the validity of PCR-RFLP markers. Therefore, we attempt to propose a mechanism to verify validation in PCR-RFLP efficacy. Our procedure contains two components: (1) genetic diversity evaluation – it measures any variation in nucleotides of organisms quantitatively for various DNA markers; (2) quasispecies model – a design that provides a qualitative understanding in the evolutionary process for self-replicating macromolecules in bacteria or viruses.

2   Methods

In the analysis of DNA markers, the validity of PCR-RFLP markers is constantly affected by various biases. Here, we give a process of verifier to confirm those enzymes applying in pathotyping. The verifier combines two components, namely, nucleotide diversity measure and quasispecies model.

2.1   Problem Formation

First of all, the digested patterns need to be quantified to a boolean vector, taking only binary values as its characters. Let $S = \{s_1, s_2, \ldots, s_n\}$ be a set of considered genomes, Z be the set of possible enzymes, and V be a set of PCR-RFLP fragment lengths on some genome s generated by enzyme set $U \subseteq Z$. Let PCR-RFLP pattern vector of V with respect to U (abbreviated as C-Vector(U, V)) be a boolean vector of digested pattern V by enzyme set U. And let L be the degree of fineness according to CAPS (Cleaved Amplified Polymorphic Sequence) fragments. As the fineness L goes smaller, the size of PCR-RFLP pattern vector may raise. For example, suppose that 10k bp is the maximum possible length. Let $L = 0.2$. Then we can obtain a C-Vector of size 5. For example, if strain g can be digested into fragment set V with fragment lengths of 0.3k, 0.7k, 2k, 3k, and 4k by U, then C-Vector$(U, V) = [1, 1, 0, 0, 0]$. Therefore, the distance between two PCR-RFLP patterns can be measured by some distance function, *e.g.*, Euclidean distance.

2.2   Nucleotide Diversity Measure

Nucleotide diversity is a concept in molecular genetics, measuring the degree of polymorphism within a population [6]. It estimates the number of nucleotide variations per digestion site between two isolates. In this paper, we use this formula to determine whether a selected enzyme set is good or not. In the following, we briefly introduce the formula.

Intrapopulation nucleotide diversity. Π_X measures variations among several strains, defined as follows:

$$\Pi_X = \frac{1}{r}lnH, H = F(3 - 2H^o)^{\frac{1}{4}}, \text{ and } F = \frac{\sum X_i(X_i \cdot n - 1)}{\sum X_i(n - 1)}$$

where F is proportion of shared fragments, H^o is double radical of F, n is the number of halploid genotypes in population X, and X_i estimates frequency of ith fragment in the population.

Interpopulation nucleotide diversity. V_{XY} calculates the divergence among populations X and Y based on restriction data, defined as follows:

$$V_{XY} = d_{XY} - \frac{(\Pi_X + \Pi_Y)}{2}, d_{XY} = -\frac{2}{r}ln(H_{XY}), H_{XY} = F_{XY}(3 - 2H_{XY}^o)^{\frac{1}{4}},$$

and $F_{XY} = \frac{2\sum X_{iX}X_{iY}}{\sum(X_{iX}+X_{iY})}$ where d_{XY} is the fragment diversity among two populations, H^o is double radical of F, and X_{iX} estimates frequency of the ith fragment in population X.

2.3   Quasispecies Model

The most doubtful part of pathotyping by DNA markers is that nucleotide changing affects restriction sites, especially on fast evolving strain. Viruses are easy to mutate when facing successful variants of their genes, and then lead to the host population becoming infected. Therefore, it is required to reproduce an evolutionary processes of self-replicating entity within the existent viral strains, inspecting the reliability of specific DNA markers [7]. Quasispecies model [8] describes a simple sequence evolution in terms of strain length, population size, mutation rate, and selection intensities. It is usually used to characterize roughly the hypothetical polynucleotide sequence of biological evolution.

The evolved population is obtained from ancestor of S by evolution process including Functions *Select*() and *Mutate*(). Function *Select*() attempts to pick k strains iteratively selected from a population, simulating the process by favorable heritable traits in successive generations of a population. Furthermore, from the definition of natural selection by Darwin's theory, it should be noted that we need to define favorable and unfavorable heritable traits in Function *Select*(), called *fitness function*. It evaluates their offspring that should be selected before reproduction. Preferable strains will be chosen and they are more likely to survive and duplicate than those with less favorable strains. Here, sequence similarity

computed by BLAST [9] is applied to evaluate generation's selection. First, all sequences in S are aligned based on an all-against-all comparison. In this manner, the average of each $s_i \in S$ against all strains can be assigned as its heritage value, written as the following equation:

$$Heritage[s_i] = \left(\frac{\sum_{j \neq i} (M_{Score}[s_i, s_j] * M_{Identity}[s_i, s_j])}{n-1} \right)^2$$

M_{Score} stores the maximal score of the BLAST alignment of s_i and s_j for each pair i, j. Similarly, $M_{Identity}$ stores the maximal identity of s_i and s_j in their alignment. Since each heritage value has been calculated, selection method rates the fitness of each solution and preferentially selects a best solution by its probability. This helps the diversity of the population large, preventing premature convergence on poor solutions. The value of heritage can be normalized by its proportion if an expected number is not apparent. After the $Select()$ function choosing in existing population, the next step is to generate a second generation population of solutions from those selected strains. By the function $Mutate()$, each strain is modified on its position p randomly with mutation rate q, which is estimated by the following formula:

$$Mutation[s_i] = 1 - \frac{1}{n-1} \left(\sum_{j \neq i} M_{Coverage}[s_i, s_j] * M_{Identity}[s_i, s_j] \right)$$

For each reproduced offspring to be born, parents are selected for breeding from the pool previously. $M_{Coverage}$ defines the maximal coverage of s_i to s_j. This evolutionary formula computes the dissimilarity counts in selection population. The Jukes-Cantor substitution model is applied to our quasispecies model, assuming any nucleotide is equally likely to be replaced by another nucleotide at any site at any time, $i.e.$, $\alpha = 0.25$. Besides, a parameter of deletion rate is estimated from ancestors by standard deviation. In Table 1, we give an example to demonstrate the result of Function $Select()$. In this case, a population S consists four strains ($n = 4$). After the computing of fitness function, each heritage value gives individual probability of being selected that directly proportionate to their fitness in $F(x)$, $F(1) = (92.16/303.04) = 30.41\%$. So the expected counts of population 1 will be $(92.16/75.76) = 1.216$. This could be a basis of individuals in the real selection. In the process of mutation, Function $Mutate()$ determines a possible mutation rate $M(x)$ within the selection pool. After that, selected individual is mutated by specific mutation rate $M(x)$ and substitution model on random positions. And $N(x)$ is the mutated numbers of nucleotide in each strain.

3   Results

In the analysis of PCR-RFLP markers, the validity of PCR-RFLP markers is constantly affected by various biased genotype frequencies, population heterozygosities, degree of inbreeding, population subdivision, and degree of individual relatedness. Results give an encouraged validation for our proposed mechanism.

Table 1. Example of Quasispecies Model

Strain No.	Initial Population	Herit. Value	F(x)	Expected Count	Real Count
1	ATATTAGACCT	92.16	30.41%	1.216	1
2	ATATTAGTTTA	105.47	34.80%	1.392	2
3	ATATTAGTCTA	105.47	34.80%	1.392	1
4	AGGAAACAACC	0	0.00%	0.000	0
AVG	-	75.76	25.00%	1	1
SUM	-	303.04	100.00%	4	4

Strain No.	Selected Population	Mutated Population	M(x)	N(x)
1	ATATTAGACCT	ATTTTTGAGCA	37%	4
2	ATATTAGTTTA	ATATTAGTTTC	21%	2
2	ATATTAGTTTA	ATCTTTGTTTA	21%	2
3	ATATTAGTCTA	ATTTTAGTGTA	31%	3

This table gives an example of how an initial population can reproduce its offspring once by Functions *Select*() and *Mutate*().

3.1 Effect of Nucleotide Diversity

Genetic distance between populations X and Y is denoted as V_{XY}, evaluating respective frequencies of variations for specific enzymes. However, the quantitative result of PCR-RFLP enzyme evaluation needs to determine its validity. For example, suppose V_{XY} is estimated by populations X and Y as 0.9891 using enzyme set U. Is it practical to use enzyme set U in real experiment to discriminate X and Y? In this experiment, let populations $X = \{$HPV-16, HPV-18$\}$ and Y be other HPV. We perform eight randomly selected enzyme sets on X and Y. They are $U_1 = \{$SpeI, BtrI, HaeIII$\}$, $U_2 = \{$HhaI, FinI, CviAII$\}$, $U_3 = \{$BspMI, AsuII, Bce83I$\}$, $U_4 = \{$ScaI, BsePI, SimI$\}$, $U_5 = \{$SnaBI, BtrI, PmII$\}$, $U_6 = \{$SacI, AgeI, BspTI$\}$, $U_7 = \{$NaeI, SacI, BspTI$\}$, and $U_8 = \{$Bce83I, AfeI, SacII$\}$. Fig. 1(a) and Fig. 1(b) examine the relation of quantitative results V_{XY} and their gel maps. In fact, the bands of electrophoresis are recognizable when V_{XY} is large enough. For example, in the case of $V_{XY}^1 = 0.402$, it is hard to distinguish X and Y. While V_{XY} is getting larger, *e.g.*, $V_{XY}^2 = 0.464, \ldots, V_{XY}^7 = 1.563$, and $V_{XY}^8 = 1.967$, the target genomic DNA of X and Y are more recognizable in 2D gel maps. This implies the consistency between genetic variation and PCR-RFLP gel patterns. So far, the accuracy of nucleotide diversity and PCR-RFLP gel maps has been proved by its cross validation.

3.2 Effect of Genetic Variations

The Human papilloma virus (HPV) contains more than 100 strains or types. Some HPV types cause benign skin warts, or papillomas, associating with the

development of common warts are transmitted environmentally or by casual skin-to-skin contact. However, the identification of high-risk HPV and low-risk HPV is still in progress. In this silicon experiment, a method from our previous result gives an appropriate enzyme set (NaeI:GCCGGC, SacI:GAGCTC, and AsuII:TTCGAA) to group HPV-16, HPV-18 (set X) and other HPV (set Y) into recognizable PCR-RFLP maps. In Table 2, V_{XY}^a measures the genomic diversity between high-risk HPV (HPV-16 and HPV-18, totally 55 isolates) and others papillomavirus. Fig. 1(c) and Fig. 1(d) shows some simulated patterns of PCR-RFLP full length genomic simulation on HPV. With the help of quasispecies model, we reproduce all of papillomas to another mutated populations, simulating the reasonable propagation for enzyme set validation. Obviously, the mutated populations in Fig. 1(e) and Fig. 1(f) also reveal similar results in $V_{XY}^{\hat{a}}$ and $\Pi_Y^{\hat{a}}$ due to the low mutation rate, digested by the same enzyme set (NaeI, SacI, and AsuII). Besides, another control group of V_{XY}^b is given, which is generated by randomly selected restriction enzyme set. The more larger of diversity V_{XY}, the more variant of populations X and Y should be.

However, it needs to specify that Human papillomavirus-26 (0.3754) and Human papillomavirus-6 (0.3882) reveal a low variation in restriction distance. Some study supports that Human papillomavirus-26 (HPV-26, HPV-51,

Table 2. *Papillomaviridae* subtyping by specific enzyme sets

Species	Nums	V_{XY}^a	Π_Y^a	$V_{XY}^{\hat{a}}$	$\Pi_Y^{\hat{a}}$	V_{XY}^b	Π_Y^b
Human papillomavirus-10	4	∞	0.2389	∞	0.2499	0.3303	0.0514
Human papillomavirus-2	9	1.2176	0.0828	1.1776	0.1460	0.6966	0.0630
Human papillomavirus-6	8	0.3882	0.2369	0.3709	0.2703	0.4755	0.1117
Human papillomavirus-26	3	0.3754	0.1115	0.3190	0.0368	0.3308	0.0662
Human papillomavirus-5	9	1.9782	0.2350	∞	0.2565	0.4000	0.0864
Human papillomavirus-53	3	1.4119	0.1732	∞	0.2142	0.4125	0.0826
Unclassified HPV	10	∞	0.2557	∞	0.2642	0.3702	0.0909
Non High-Risk HPV	136	0.9891	0.6663	0.8957	0.6195	0.4167	0.1314
Bovine	20	∞	0.0540	∞	0.2650	0.4506	0.1182
Phocoena spinipinnis	2	0.2373	0.0000	0.2018	0.0000	0.3102	0.0000
Mastomys coucha	4	∞	0.1163	1.5793	0.0834	0.3136	0.0382
Others Papillomavirus	71	∞	0.2553	∞	0.2624	0.3702	0.0909

'**Species**' indicates the taxonomy of genomes in *Papillomaviridae*. '**Nums**' shows the number of strains in the group. V_{XY}^a - initial strains and suggested enzyme set. $V_{XY}^{\hat{a}}$ - mutated strains and suggested enzyme set. V_{XY}^b - initial strains and random enzyme set. Due to the limitation of space, several subtypes of papillomavirus are not listed.
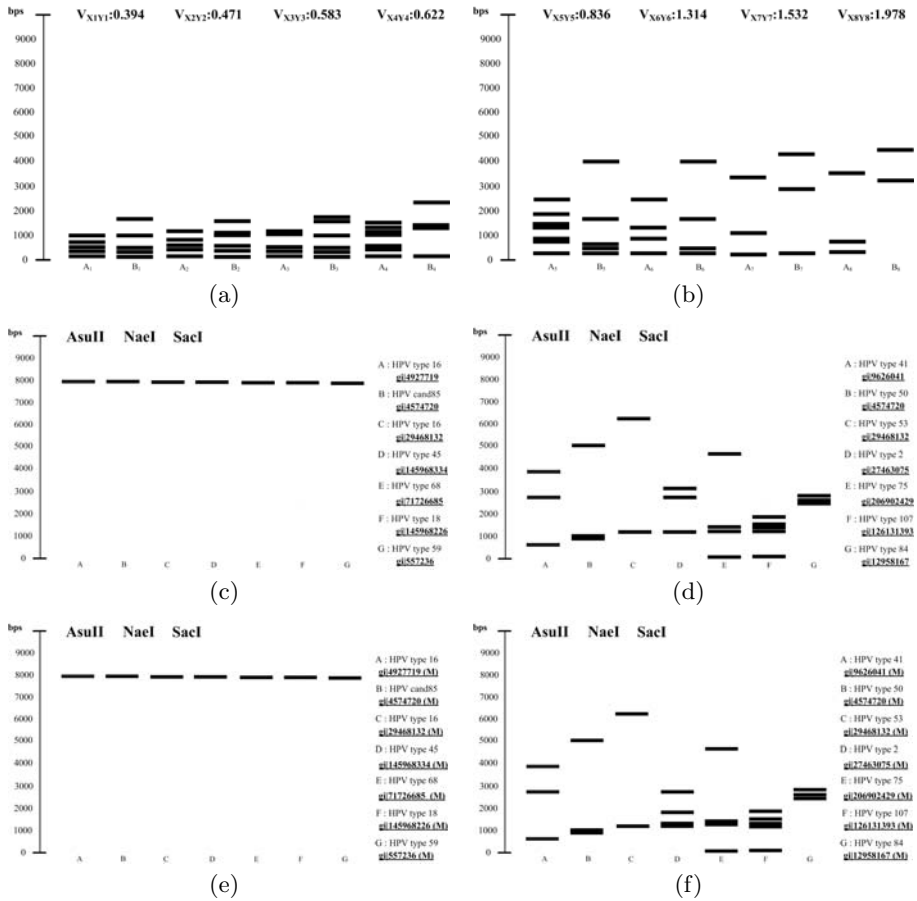
Fig. 1. Qualitative comparisons: a relation of genetic diversity and its gel map

HPV-69, HPV-82) and Human papillomavirus-6 (HPV-11, HPV-13, HPV-44, HPV-6, HPV-6a, HPV-6b, HPV-6c, HPV-6e, HPV-6vc, HPV-74) involved with genital warts [10]. In addition, a distantly related strain Phocoena spinipinnis papillomavirus (PsPV, 0.2373) also shares seldom genetic variations among high-risk HPV. It is interesting because PsPV even not exist in homo sapines, it is a Omikronpapillomavirus and the other is Alphapapillomavirus (HPV-16 and HPV-18). The ORF of PsPV also shows high percentage of nucleotide identity (54-55%) with HPV. This supports the hypothesis that PV evolution may not be monophyletic across all genes [11]. In conclusion, for the problem of patho-typing, whether a given enzyme set can be used to distinguish two populations can be determined by simulated in silico using our proposed method.

4   Discussion

Due to the SNP or numerous reasons of nucleotide variations, we propose a simple method to inspect PCR-based markers in genomic pathotyping. The mutated population obtained from quasispecies model simulates fast evolving strains, giving an encouraged cross validation in *Papillomaviridae*. In addition, the relation between nucleotide diversity and CAPS gel electrophoresis has been proved its validation. In the future, the concept of PCR-based marker verification may be extended to several ways on DNA fingerprint. In particular, we expect our method may give a referral to biologists working in genetics, plant discrimination, forensics, and other related fields.

References

1. Wyman, A.R., White, R.: A highly polymorphic locus in human DNA. Proc. Natl. Acad. Sci. U S A 77(11), 6754–6758 (1980)
2. Williams, J.G., Kubelik, A.R., Livak, K.J.: DNA polymorphisms amplified by arbitrary primers are useful as genetic markers. Nucleic Acids Res. 18(22), 6531–6535 (1990)
3. Neff, M.M., Neff, J.D., Chory, J., Pepper, A.E.: DCAPS, a simple technique for the genetic analysis of single nucleotide polymorphisms: experimental applications in arabidopsis thaliana genetics. Plant J. 14(3), 387–392 (1998)
4. Nobre, R.J., de Almeida, L.P., Martins, T.C.: Complete genotyping of mucosal human papillomavirus using a restriction fragment length polymorphism analysis and an original typing algorithm. Journal of Clinical Virology 42, 13–21 (2008)
5. Fereidouni, S.R., Harder, T.C., Starick, E.: Rapid pathotyping of recent h5n1 highly pathogenic avian influenza viruses and of h5 viruses with low pathogenicity by rt-pcr and restriction enzyme cleavage pattern (recp). Journal of Virological Methods 154, 14–19 (2008)
6. Nei, M., Li, W.H.: Mathematical model for studying genetic variation in terms of restriction endonucleases. Proc. Natl. Acad. Sci. U S A 76(10), 5269–5273 (1979)
7. Vladimir, R., Yuri, T.: Estimation of the evolution speed for the quasispecies model: Arbitrary alphabet case. In: Rutkowski, L., Tadeusiewicz, R., Zadeh, L.A., Żurada, J.M. (eds.) ICAISC 2006. LNCS (LNAI), vol. 4029, pp. 460–469. Springer, Heidelberg (2006)
8. Eigen, M., McCaskill, J., Schuster, P.: The molecular quasi-species. Advances in Chemical Physics 75, 149–263 (1989)
9. Altschul, S., Gish, W., Miller, W.: Basic local alignment search tool. J. Mol. Biol. 215, 403–410 (1990)
10. Gissmann, L., Wolnik, L., Ikenberg, H.: Human papillomavirus types 6 and 11 DNA sequences in genital and laryngeal papillomas and in some cervical cancers. Proc. Natl. Acad. Sci. U S A 80(2), 560–563 (1983)
11. Van Bressem, M.F., Cassonnet, P., Rector, A.: Genital warts in burmeister's porpoises: characterization of phocoena spinipinnis papillomavirus type 1 (pspv-1) and evidence for a second, distantly related pspv. Journal of General Virology 88(7), 1928–1933 (2007)

The EDML Format to Exchange Energy Profiles of Protein Molecular Structures[*]

Dariusz Mrozek[1], Bożena Małysiak-Mrozek[1], Stanisław Kozielski[1],
and Sylwia Górczyńska-Kosiorz[2]

[1] Institute of Informatics, Silesian University of Technology
Akademicka 16, 44-100 Gliwice, Poland
{Dariusz.Mrozek,Bozena.Malysiak,Stanislaw.Kozielski}@polsl.pl
[2] Department of Internal Diseases, Diabetology and Nephrology
Silesian Medical University, 3-go Maja 13-15, 41-800 Zabrze, Poland
sgorczynska-kosiorz@sum.edu.pl

Abstract. The Energy Distribution Markup Language (EDML) is an XML-based format that we have designed and developed to exchange protein structure energy profiles between many computers and users. Energy profiles that are distributions of various potential energies over consecutive atoms in protein molecular structures can be downloaded from the Energy Distribution Data Bank (EDB, *http://edb.aei.polsl.pl*) in the EDML format. In the paper, we describe the purpose of the EDML, a possible use of energy profiles, and internal structure of documents created in the EDML format.

Keywords: bioinformatics, protein structure, energy profiles, XML, databases.

1 Introduction

The Energy Distribution Data Bank (EDB, *http://edb.aei.polsl.pl*) is a worldwide repository for energy profiles of protein molecular structures [1]. A single energy profile is a set of distributions of various types of energy over protein 3D structure. Energy profiles stored in the EDB repository are obtained in the computational process. To this purpose, we use molecular mechanics techniques and force field methods, which produce a set of energy properties for particular molecular structures [2]. The computations base on locations of atoms described by the Cartesian coordinates (x, y, z). Protein structures are taken from the well-known macromolecular structure repository Protein Data Bank [3].

The molecular data and energy profiles stored in the EDB can be used to support various research and studies in the area of biochemistry, molecular biology and bioinformatics. In our research, we use energy profiles deposited in the EDB in the searching of particular structural regions of proteins, e.g. active sites of enzymes, or energetically favorable places in protein structures. We have developed the EAST method for the

---

approximate similarity searching of protein substructures on the basis of their energy properties [4]. The Energy Distribution Data Bank gives access to energy profiles, which can be perceived as energy templates for structure prediction, just like structural templates. We use energy profiles in our work as referential templates in the structure prediction processes with the NPF algorithm [5]. Energy profiles allow observations of small changes in protein conformations at the level of protein geometry and at the energy level in 3D space or after the projection to 2D grids [6]. This can be a part of the conformational analysis, which is very important for the analysis of protein activity in cellular reactions. Users can use data stored in the EDB in the similarity searching carried at the level of structure and energy, simultaneously. Finally, energy breakdowns can be useful for the investigation of protein reactivity, and therefore, very important for protein docking and modeling of cellular reactions.

Users can easily access the EDB data through the EDB website, which provides a simple query mechanism to retrieve records from the EDB database. As of Tuesday, March 24, 2009 there are 34 372 energy profiles in the EDB. The access to the information gathered in the Energy Distribution Data Bank is free to public. Since the information stored in the EDB covers many aspects of protein construction and several molecular properties, it is difficult to exchange the data between many users without appropriate data format. For this reason, we decided to design and develop a special exchange language, which bases on the XML technology. The EDB users have now the possibility to download any energy profile from the EDB website in the EDML format (Energy Distribution Markup Language). This paper gives a brief description of the EDML exchange language. A short overview of molecular and energy features stored in the EDB is presented in section 2. In section 3 we describe the most important parts of the proposed EDML format and possible alternative forms related to our current and future works that are reported in sections 4 and 5.

2   Theoretical Background of Protein Molecular Properties

The Energy Distribution Data Bank joined the wide group of repositories that store biological information related to molecules of life, such as DNA/RNA and proteins. These repositories usually concentrate on chosen description levels of biological existence of living organisms. Therefore, they store information specific to the description level. In the construction of proteins we can distinguish four description (or representation) levels: primary structure, secondary structure, tertiary structure and quaternary structure [7].

The primary structure is defined by the amino acid sequence in protein linear chain [8]. Databases, like: UniProt/SwissProt [9], Protein Information Resource (PIR) [10], Protein Research Foundation (PRF) [11] or NCBI Protein [12], store data concerning amino acid sequences of proteins. Example of a sequence of the human hemoglobin molecule (4HHB, chain A) is presented in Fig. 1. Each letter in a sequence corresponds to one amino acid in the protein chain. Proteins can be built up to several amino acid chains.

Secondary, tertiary and quaternary structures define the protein conformation or protein spatial structure [8], [13], [14]. The protein spatial structure is frequently described by the (x, y, z) Cartesian coordinates of atoms composing the structure and

covalent bonds connecting particular atoms. In Fig. 2 we can observe an example of the molecular structure of the Human Topoisomerase I in different representations. Popular repositories, like Protein Data Bank (PDB) [3], EBI Macromolecular Structure Database (MSD) [15] or Molecular Modeling DataBase (MMDB) [16], collect data related to protein structures. These databases make their data accessible for the broad community of users by using exchange formats, like: mmCIF [17], ASN.1 [18] or PDBML [19].

The Energy Distribution Data Bank consists both, structural data in the form of Cartesian coordinates of particular atoms in a protein structure and energy profiles including free energy breakdown over each atom of the structure. The computations of energy profiles were done with the use of molecular mechanics methods and the Amber94 force field [20]. We use the following functional form for the force field, which includes bonded and non-bonded interactions between atoms:

$$E_T\left(r^N\right) = \sum_{i=1}^{bonds} \frac{k_i^s}{2}\left(d_i - d_i^0\right)^2 + \sum_{i=1}^{angles} \frac{k_i^b}{2}\left(\theta_i - \theta_i^0\right)^2 + \sum_{i=1}^{torsions} \frac{V_n}{2}\left(1 + \cos(n\omega - \gamma)\right)$$

$$+ \sum_{i=1}^{N}\sum_{j=i+1}^{N}\left(4\varepsilon_{ij}\left[\left(\frac{\sigma_{ij}}{r_{ij}}\right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}}\right)^6\right] + \frac{q_i q_j}{4\pi\varepsilon_0 r_{ij}}\right),$$
(1)

where $E_T(r^N)$ denotes the total potential energy, which is a function of atomic positions [2]. There are different types of contributing energies that are calculated for the structure r^N (N is a number of atoms): bond stretching (first term of the eq. 1), angle bending (second term), torsional angle (third term), van der Waals (fourth term), and

```
>4HHB:A|PDBID|CHAIN|SEQUENCE
VLSPADKTNVKAAWGKVGAHAGEYGAEALERMFLSFPTTKTYFPHFDLSHGSAQVKGHGKKVADALTNAVAHVDDMP
NALSALSDLHAHKLRVDPVNFKLLSHCLLVTLAAHLPAEFTPAVHASLDKFLASVSTVLTSKYR
```

Fig. 1. Amino acid sequence of the human hemoglobin (chain A only) in the FASTA format

a)   b)   c)
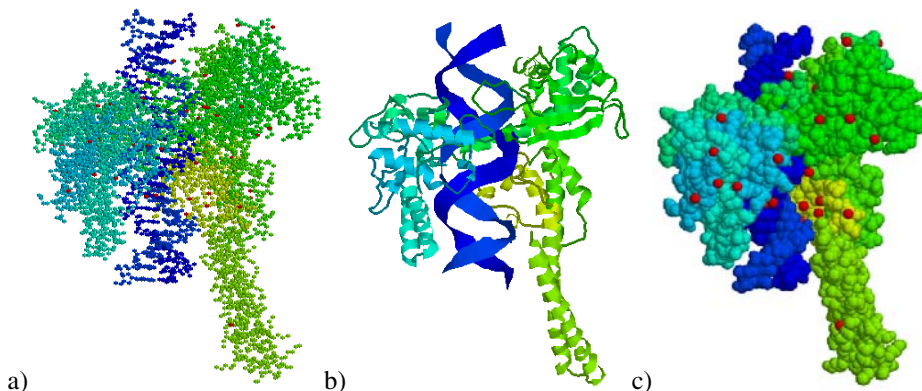
Fig. 2. Different representations of the spatial structure of the human topoisomerase (1A36) from the PDB: a) atomic, b) ribbon (visible characteristic secondary structure elements, e.g. spiral α-helices), c) space fill

electrostatic (fifth term). The total potential energy $E_T(r^N)$ is a sum of all contributing energies. These contributing energies summarize charges coming from atomic inter- actions. The functional form (1) is frequently used in the modeling of protein struc- tures. In our research, we cast calculated charges to particular atoms, so they are distributed over all atoms in the structure r^N. We call these distributions as energy profiles. Distribution of the van der Waals potential can be observed in Fig. 3 as po- tential surfaces surrounding atoms of the structure. Since energy profiles are calcu- lated for protein spatial structures, they describe molecular structures in terms of energy properties. For our calculations we use protein structures received from the well-known Protein Data Bank (PDB). The Energy Distribution Data Bank consists of energy profiles for many structures from the PDB (34 372). These profiles can be now exchanged with the use of the EDML format.

3   Description of the EDML Format

The Energy Distribution Markup Language (EDML) is a way of storing and exchang- ing energy profiles determined for protein molecular structures. The EDML format was developed on the basis of the W3C XML recommendation [21]. It benefits from all advantages the XML technology provides for data transfer and data storage. In Fig. 4 we present a simplified framework of a sample EDML file describing energy profile for the 2HBS molecule (Human Deoxyhemoglobin S).

The document presented in Fig. 4 begins with the *EnergyProfile* XML element. This root element consists of one *Header* subelement and one or many *Energy* subelements. The *Header* element contains general description of the molecule. *Energy* elements hold distribution of different charges for consecutive chains of the transferred molecule. All these elements are described in more details in this section.

Well-formed and valid EDML file should begin with the *EnergyProfile* root ele- ment. Obligatory *forcefield* attribute of the element indicates the force field used in the computation of the energy profile (Fig. 5). The following, required attribute *pdbid*

a)   b)
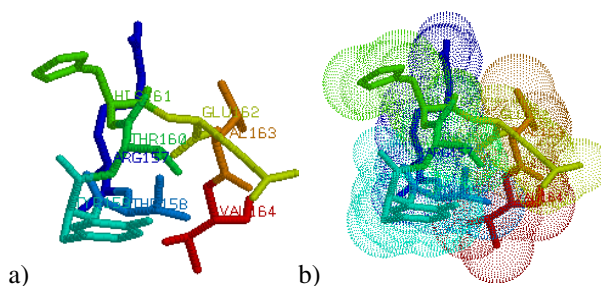
Fig. 3. Parts of molecular structures of the Human Cyclin Kinase 2 (molecule 1B38 in the PDB): a) plain sticks representation, b) sticks representation with the van der Waals potential surfaces

```
<edb:EnergyProfile xmlns:edb="http://edb.aei.polsl.pl"
   xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
   xsi:schemaLocation="http://edb.aei.polsl.pl/edml edml.xsd"
   forcefield="Amber94" pdbid="2HBS">
   <edb:Header>
      ...
   </edb:Header>
   <edb:Energy Chain="A">
      ...
   </edb:Energy>
   <edb:Energy Chain="B">
      ...
   </edb:Energy>

   ...
   <edb:Energy Chain="H">
      ...
   </edb:Energy>
</edb:EnergyProfile>
```

Fig. 4. General construction of the sample EDML file for the molecule 2HBS

consists of the PDB ID identifier of the molecule, for which the profile is computed. The identifier is compliant to Protein Data Bank identifiers (PDB IDs). Single EDML file can store the energy profile of only one, single molecule, and holds energy distributions computed with the use of only one force field. Other attributes of the *EnergyProfile* element are strictly associated with XML technology – they define the *edb* namespace and the location of the XML Schema file validating all EDML documents.

The *Header* element (under the *EnergyProfile* element) contains descriptor of the molecule and primary structure information for all chains (Fig. 6). In the presented case, the 2HBS molecule is made up of eight amino acid chains: A, B, C, D, E, F, G, H. The descriptor contains also the following important information: the identifier of the molecule (*PDBID* element), name of the molecule (*Name* element), short description (*Description* element), class of the molecule (*ProteinClass* element), source organism of the molecule (*Source* element and its subelements) – scientific and common organism name (*OrganismScience* element and *OrganismCommon* element) and optionally: tissue (*Tissue* element), cell type (*Cell* element) and location inside the cell (*CellularLocation* element). All these data were extracted from the PDB files during the energy profile calculation.

```
<edb:EnergyProfile xmlns:edb="http://edb.aei.polsl.pl"
   xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
   xsi:schemaLocation="http://edb.aei.polsl.pl/edml edml.xsd"
   forcefield="Amber94" pdbid="2HBS">
...
</edb:EnergyProfile>
```

Fig. 5. The overview of the root *EnergyProfile* element

```
<edb:Header>
    <edb:PDBID>2HBS</edb:PDBID>
    <edb:Isomer>A</edb:Isomer>
    <edb:Name>HEMOGLOBIN S</edb:Name>
    <edb:Description>THE HIGH RESOLUTION CRYSTAL STRUCTURE OF
            DEOXYHEMOGLOBIN S</edb:Description>
    <edb:ProteinClass>OXYGEN TRANSPORT</edb:ProteinClass>
    <edb:Source>
      <edb:OrganismScience>HOMO SAPIENS</edb:OrganismScience>
      <edb:OrganismCommon>HUMAN</edb:OrganismCommon>
      <edb:Tissue>BLOOD</edb:Tissue>
      <edb:Cell>RED BLOOD CELLS</edb:Cell>
      <edb:CellularLocation>CYTOPLASM</edb:CellularLocation>
    </edb:Source>
    <edb:Sequence Chain="A" code="3">VLSPADKTNVKAAWGKVGAHAGEYGAEALERMFLS
      FPTTKTYFPHFDLSHGSAQVKGHGKKVADALTNAVAHVDDMPNALSALSDLHAHKLRVDPVNFKLL
      SHCLLVTLAAHLPAEFTPAVHASLDKFLASVSTVLTSKYR</edb:Sequence>
    <edb:Sequence Chain="B" code="3">VHLTPVEKSAVTALWGKVNVDEVGGEALGRLLVVY
      PWTQRFFESFGDLSTPDAVMGNPKVKAHGKKVLGAFSDGLAHLDNLKGTFATLSELHCDKLHVDPE
      NFRLLGNVLVCVLAHHFGKEFTPPVQAAYQKVVAGVANALAHKYH</edb:Sequence>
      ...
    <edb:Sequence Chain="H" code="3">VHLTPVEKSAVTALWGKVNVDEVGGEALGRLLVVY
      PWTQRFFESFGDLSTPDAVMGNPKVKAHGKKVLGAFSDGLAHLDNLKGTFATLSELHCDKLHVDPE
      NFRLLGNVLVCVLAHHFGKEFTPPVQAAYQKVVAGVANALAHKYH</edb:Sequence>
</edb:Header>
```

Fig. 6. The *Header* section in the sample EDML file

Single *Sequence* element contains information about the primary structure of appropriate polypeptide chain of the molecule. Since proteins can be made up of several chains, the *Sequence* element can occur multiple times. Single polypeptide chain is identified by its one letter name stored as a value of the required *Chain* attribute. The *code* attribute decides how the primary structure is recorded – with 1-letter or 3-letter code. In the example shown in Fig. 7 the 1-letter code was used.

```
<Sequence chain="A" code="1">VLSPADKTNVKAAWGKVGAHAGEYGAEALERMFLSFPTTKT
      YFPHFDLSHGSAQVKGHGKKVADALTNAVAHVDDMPNALSALSDLHAHKLRVDPVNFKLLSHCLL
      VTLAAHLPAEFTPAVHASLDKFLASVSTVLTSKYR
</Sequence>
```

Fig. 7. Primary structure recorded in the 1-letter code inside the *Sequence* element

The same sequence written with the 3-letter code is presented in Fig. 8.

```
<edb:Sequence Chain="A" code="3">VAL LEU SER PRO ALA ASP LYS THR ASN VAL
    LYS ALA ALA TRP GLY LYS VAL GLY ALA HIS ALA GLY GLU TYR GLY ALA
    GLU ALA LEU GLU ARG MET PHE LEU SER PHE PRO THR THR LYS THR TYR
    PHE PRO HIS PHE ASP LEU SER HIS GLY SER ALA GLN VAL LYS GLY HIS
    GLY LYS LYS VAL ALA ASP ALA LEU THR ASN ALA VAL ALA HIS VAL ASP
    ASP MET PRO ASN ALA LEU SER ALA LEU SER ASP LEU HIS ALA HIS LYS
    LEU ARG VAL ASP PRO VAL ASN PHE LYS LEU LEU SER HIS CYS LEU LEU
    VAL THR LEU ALA ALA HIS LEU PRO ALA GLU PHE THR PRO ALA VAL HIS
    ALA SER LEU ASP LYS PHE LEU ALA SER VAL SER THR VAL LEU THR SER
    LYS TYR ARG
</edb:Sequence>
```

Fig. 8. Primary structure recorded in the 3-letter code inside the *Sequence* element

The *Energy* elements, which are located under the root element in the EDML document hierarchy, store data concerning energy profile of the molecule. The *Energy* element can occur many times, according to the number of chains the molecule is made up of. The chain, for which energy distribution is currently described, is identified by a value of the *Chain* attribute in the *Energy* element (Fig. 9). *Peptide* subelements represent consecutive amino acids in the protein polypeptide chain. Each peptide is described by the *aminoacid* attribute, which consists of the type of the amino acid in the 3-letter code, e.g. *VAL* for *Valine*, and *residueno* attribute, which is a number of the residue in the polypeptide chain. The order of *Peptide* elements is the same as the order of amino acids in the primary structure. The *xsi:type* attribute describes the name of the force field used in the computation of the energy profile. The attribute also determines the structure of elements inside the *Atom* element.

The *Atom* element represents a single atom in the protein structure and an energy point calculated for the atom. The required attribute *name* determines the chemical element, e.g. *N* stands for *Nitrogen*, *C* stands for *Carbon* (*CA* is a characteristic α-carbon), etc., and the *atomno* attribute holds the number of the atom in the protein structure. Cartesian coordinates of an atom are described by the *Coordinates* element and *Cartn_x*, *Cartn_y*, and *Cartn_z* subelements (Fig. 10).

```xml
<edb:Energy Chain="A">
    <edb:Peptide aminoacid="VAL" residueno="1" xsi:type="AmberType">
        <edb:Atom name="N" atomno="1">
            ...
        </edb:Atom>
        <edb:Atom name="CA" atomno="2">
            ...
        </edb:Atom>
        ...
    </edb:Peptide>
    <edb:Peptide aminoacid="LEU" residueno="2" xsi:type="AmberType">
        <edb:Atom name="N" atomno="19">
            ...
        </edb:Atom>
        <edb:Atom name="CA" atomno="20">
            ...
        </edb:Atom>
        ...
    </edb:Peptide>
    ...
</edb:Energy>
```

Fig. 9. A simplified structure of the *Energy* element

```xml
<edb:Atom name="N" atomno="19">
    <edb:Coordinates>
        <edb:Cartn_x>35,2169990539551</edb:Cartn_x>
        <edb:Cartn_y>57,8769989013672</edb:Cartn_y>
        <edb:Cartn_z>34,5320014953613</edb:Cartn_z>
    </edb:Coordinates>
    <edb:ETotalEnergy>4,63280016962381</edb:ETotalEnergy>
    <edb:EBondStretch>0,0461000017821789</edb:EBondStretch>
    <edb:EAngleBend>9,99999974737875E-05</edb:EAngleBend>
    <edb:ETorisonalAngle>9,999999747375E-05</edb:ETorisonalAngle>
    <edb:EVanDerWaals>0,498600006103516</edb:EVanDerWaals>
    <edb:EChargeCharge>4,08790016174316</edb:EChargeCharge>
    <edb:EImproperTorsion>0</edb:EImproperTorsion>
</edb:Atom>
```

Fig. 10. Structure of a single *Atom* element

Values of particular charges are given in the following elements under each *Atom* element: *ETotalEnergy*, *EBondStretch*, *EAngleBend*, *ETorsionalAngle*, *EVanDerWaals*, *EChargeCharge*, *EImproperTorsion*. Each atom in the protein molecular structure is described in the same way – the *Atom* elements always store the information regarding location and energy properties for each type of potential energy. Since proteins are built up of thousands of atoms, there are thousands of *Atom* elements in a single EDML file.

The structure of the energy breakdown can differ for various force fields. Therefore, the number of elements at the list can change and alternative elements can appear on it. Alternative constructions of the *Atom* element depends on the value of the *xsi:type* attribute in the *Peptide* element. However, inside a single EDML file the construction of *Atom* element is always constant, since one EDML file consists of energy profiles calculated with the use of one, chosen force field. A sample construction of the *Atom* element for the *Amoeba-Protein* force field [22] is shown in Fig. 11.

The abbreviated EDML file for the sample 2HBS molecule (Human Deoxy-hemoglobin S) is presented in Fig. 12.

4 Querying EDML Documents

One of the main advantages of exchanging data in the EDML format is the possibility of further processing of these data according to the current needs and goals. Moreover, there is a set of existing tools and technologies that support or cooperate with the XML technology. The XPath [23] and XQuery [24] languages allow to submit queries against EDML documents, find appropriate information, navigate through XML elements and attributes, and even create new documents by reprocessing the input EDML files. Since XPath and XQuery are W3C Recommendations, similarly to the XML, all the major database engines support these query languages. This allows to incorporate and work with XML data in relational database systems.

The sample query presented below shows how we retrieve the primary structure for specified chain A from the following EDML file from Fig. 12.

```xml
<edb:Atom name="N" atomno="19">
   <edb:Coordinates>
      <edb:Cartn_x>35,2169990539551</edb:Cartn_x>
      <edb:Cartn_y>57,8769989013672</edb:Cartn_y>
      <edb:Cartn_z>34,5320014953613</edb:Cartn_z>
   </edb:Coordinates>
   <edb:ETotalEnergy>-17,33000</edb:ETotalEnergy>
   <edb:EBondStretch>0,2382</edb:EBondStretch>
   <edb:EAngleBend>0,0880</edb:EAngleBend>
   <edb:ETorisonalAngle>-0,1006</edb:ETorisonalAngle>
   <edb:EVanDerWaals>1,5878</edb:EVanDerWaals>
   <edb:EChargeCharge>0,0000</edb:EChargeCharge>
   <edb:EStretch-Bend>0,0047</edb:EStretch-Bend>
   <edb:EPolarization>0,5713</edb:EPolarization>
   <edb:EOut-of-PlaneBend>0,3207</edb:EOut-of-PlaneBend>
   <edb:EPi-OrbitalTorsion>0,0682</edb:EPi-OrbitalTorsion>
   <edb:ETorsion-Torsion>-0,1690</edb:ETorsion-Torsion>
   <edb:EAtomicMultipoles>-19,9393</edb:EAtomicMultipoles>
</edb:Atom>
```

Fig. 11. The structure of the *Atom* element for the Amoeba-Protein force field

```
  for $s in /edb:EnergyProfile/edb:Header/edb:Sequence
  where $s/@Chain="A"
  return data($s)

<edb:EnergyProfile xmlns:edb="http://edb.aei.polsl.pl"
    xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
    xsi:schemaLocation="http://edb.aei.polsl.pl/edml edml.xsd"
    forcefield="Amber94" pdbid="2HBS">
    <edb:Header>
        <edb:PDBID>2HBS</edb:PDBID>
        <edb:Isomer>A</edb:Isomer>
        <edb:Name>HEMOGLOBIN S</edb:Name>
        <edb:Description>CRYSTAL STRUCTURE OF DEOXYHEMOGLOBIN S</edb:Description>
        <edb:ProteinClass>OXYGEN TRANSPORT</edb:ProteinClass>
        <edb:Source>
            <edb:OrganismScience>HOMO SAPIENS</edb:OrganismScience>
            <edb:OrganismCommon>HUMAN</edb:OrganismCommon>
            <edb:Tissue>BLOOD</edb:Tissue>
            <edb:Cell>RED BLOOD CELLS</edb:Cell>
            <edb:CellularLocation>CYTOPLASM</edb:CellularLocation>
        </edb:Source>
        <edb:Sequence Chain="A" code="3">VLSPADKTNVKAAWGKVGAHAGEYGAEALERMFLS
            FPTTKTYFPHFDLSHGSAQVKGHGKKVADALTNAVAHVDDMPNALSALSDLHAHKLRVDPVNFKL
            LSHCLLVTLAAHLPAEFTPAVHASLDKFLASVSTVLTSKYR</edb:Sequence>
        <edb:Sequence Chain="B" code="3">VHLTPVEKSAVTALWGKVNVDEVGGEALGRLLVVY
            PWTQRFFESFGDLSTPDAVMGNPKVKAHGKKVLGAFSDGLAHLDNLKGTFATLSELHCDKLHVDP
            ENFRLLGNVLVCVLAHHFGKEFTPPVQAAYQKVVAGVANALAHKYH</edb:Sequence>
        ...
        <edb:Sequence Chain="H" code="3">VHLTPVEKSAVTALWGKVNVDEVGGEALGRLLVVY
            PWTQRFFESFGDLSTPDAVMGNPKVKAHGKKVLGAFSDGLAHLDNLKGTFATLSELHCDKLHVDP
            ENFRLLGNVLVCVLAHHFGKEFTPPVQAAYQKVVAGVANALAHKYH</edb:Sequence>
    </edb:Header>

    <edb:Energy Chain="A">
        <edb:Peptide aminoacid="VAL" residueno="1" xsi:type="AmberType">
            <edb:Atom name="N" atomno="1">
                <edb:Coordinates>
                    <edb:Cartn_x>10,3900003433228</edb:Cartn_x>
                    <edb:Cartn_y>20,4270000457764</edb:Cartn_y>
                    <edb:Cartn_z>13,210000038147</edb:Cartn_z>
                </edb:Coordinates>
                <edb:ETotalEnergy>4,22569996111997</edb:ETotalEnergy>
                <edb:EBondStretch>0,116700001060963</edb:EBondStretch>
                <edb:EAngleBend>9,99999974737875E-05</edb:EAngleBend>
                <edb:ETorisonalAngle>0,0020999999251216</edb:ETorisonalAngle>
                <edb:EVanDerWaals>0,658100008964539</edb:EVanDerWaals>
                <edb:EChargeCharge>3,44869995117188</edb:EChargeCharge>
                <edb:EImproperTorsion>0</edb:EImproperTorsion>
            </edb:Atom>
            <edb:Atom name="CA" atomno="2">
                <edb:Coordinates>
                    <edb:Cartn_x>10,5900001525879</edb:Cartn_x>
                    <edb:Cartn_y>18,9479999542236</edb:Cartn_y>
                    <edb:Cartn_z>13,1040000915527</edb:Cartn_z>
                </edb:Coordinates>
                <edb:ETotalEnergy>0,722300007939339</edb:ETotalEnergy>
                <edb:EBondStretch>0,267199993133545</edb:EBondStretch>
                <edb:EAngleBend>0,213300004601479</edb:EAngleBend>
                <edb:ETorisonalAngle>0,19760000705719</edb:ETorisonalAngle>
                <edb:EVanDerWaals>-0,269400005722205</edb:EVanDerWaals>
                <edb:EChargeCharge>0,31360000371933</edb:EChargeCharge>
                <edb:EImproperTorsion>0</edb:EImproperTorsion>
            </edb:Atom>
            ...
        </edb:Peptide>
```

Fig. 12. The abbreviated EDML file for the sample 2HBS molecule

```
<edb:Peptide aminoacid="LEU" residueno="2" xsi:type="AmberType">
    <edb:Atom name="N" atomno="19">
        <edb:Coordinates>
            <edb:Cartn_x>35,2169990539551</edb:Cartn_x>
            <edb:Cartn_y>57,8769989013672</edb:Cartn_y>
            <edb:Cartn_z>34,5320014953613</edb:Cartn_z>
        </edb:Coordinates>
        <edb:ETotalEnergy>4,63280016962381</edb:ETotalEnergy>
        <edb:EBondStretch>0,0461000017821789</edb:EBondStretch>
        <edb:EAngleBend>9,99999974737875E-05</edb:EAngleBend>
        <edb:ETorisonalAngle>9,999999747375E-05</edb:ETorisonalAngle>
        <edb:EVanDerWaals>0,498600006103516</edb:EVanDerWaals>
        <edb:EChargeCharge>4,08790016174316</edb:EChargeCharge>
        <edb:EImproperTorsion>0</edb:EImproperTorsion>
    </edb:Atom>
    <edb:Atom name="CA" atomno="20">
        <edb:Coordinates>
            <edb:Cartn_x>34,1689987182617</edb:Cartn_x>
            <edb:Cartn_y>57,0750007629395</edb:Cartn_y>
            <edb:Cartn_z>35,2169990539551</edb:Cartn_z>
        </edb:Coordinates>
        <edb:ETotalEnergy>0,683500021696091</edb:ETotalEnergy>
        <edb:EBondStretch>0,201499998569489</edb:EBondStretch>
        <edb:EAngleBend>0,26010000705719</edb:EAngleBend>
        <edb:ETorisonalAngle>0,142100006341934</edb:ETorisonalAngle>
        <edb:EVanDerWaals>-0,256199985742569</edb:EVanDerWaals>
        <edb:EChargeCharge>0,335999995470047</edb:EChargeCharge>
        <edb:EImproperTorsion>0</edb:EImproperTorsion>
    </edb:Atom>
    ...
</edb:Peptide>
    ...
</edb:Energy>
    ...
</edb:EnergyProfile>
```

Fig. 12. (*continued*)

The query is written in the XQuery language and returns the sequence in the form presented in Fig. 13.

```
VLSPADKTNVKAAWGKVGAHAGEYGAEALERMFLSFPTTKTYFPHFDLSHGSAQVKGHGKKVADALTNAVA
HVDDMPNALSALSDLHAHKLRVDPVNFKLLSHCLLVTLAAHLPAEFTPAVHASLDKFLASVSTVLTSKYR
```

Fig. 13. Amino acid sequence returned by the sample XQuery query

The following query in the XQuery language retrieves residue numbers and calculates the total electrostatic energy for each residue (peptide) on the basis of atomic potentials (*Atom/EChargeCharge* elements, Fig. 12).

```
for $e in /edb:EnergyProfile/edb:Energy,
    $p in $e/edb:Peptide
where $e/@Chain="A"
return
    <Peptide number="{$p/@residueno}"
        ecc="{ sum(data($p//edb:EChargeCharge)) }" />
```

Each value of the summed potential is returned in the *ecc* attribute with the residue number (*number* attribute) inside the *Peptide* element of newly created document:

```
<Peptide number="1" ecc="56.0598964095117" />
<Peptide number="2" ecc="-27.6810989379884" />
<Peptide number="3" ecc="-5.21520256996158" />
<Peptide number="4" ecc="-7.90960001945499" />
<Peptide number="5" ecc="-5.34040427207943" />
...
```

With the use of the above result set we are able to generate cumulated energy characteristics for chosen type of potential energy. E.g. in Fig. 14 we present a sample cumulated electrostatic energy characteristic for the molecule 2HBS. These characteristics can be used e.g. to analyze conformational deformations of protein structures based on the comparison of energy patterns and verification of energy discrepancies. The electrostatic energy characteristic presented in Fig. 14 was generated at the EDB web site (*http://edb.aei.polsl.pl*).
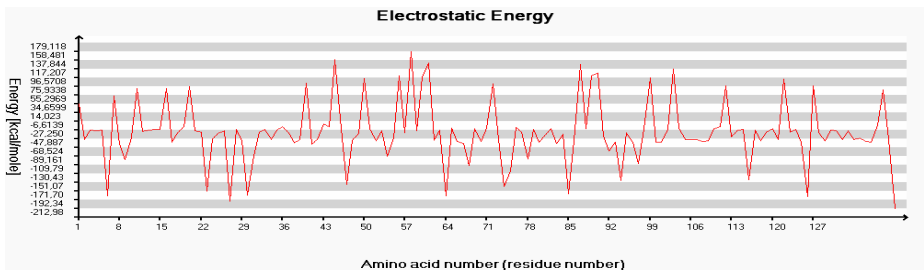
Fig. 14. Cumulated electrostatic energy characteristic for molecule 2HBS

5   Concluding Remarks

The general purpose of the Energy Distribution Markup Language that we have developed is exchanging data related to energy profiles for protein molecular structures. Calculated energy profiles are permanently stored in the Energy Distribution Data Bank (EDB) and can be transferred between user's computers with the use of proposed EDML format. In our resent research on protein similarities, the EDML format has been used to transfer energy profiles of chosen set of molecules between software agents comparing distributions of energy properties over molecular structures. The work is supported by Grant No. N N516 265835: *Protein Structure Similarity Searching in Distributed Multi Agent System.*

Since energy profiles include distributions of various energies over all atoms in a protein structure, the EDML files are sizeable (usually several MB). This is also caused by description tags typical for the XML technology. However, this inconvenience is rewarded by the possibility of flexible data reprocessing and the wealth of ready-to-use tools in the area of the XML technology.

In the future, we plan to extend the EDML format in order to store energy profiles calculated using more complex force field parameter sets, like *Amoeba-Protein*. The preliminary structure of the EDML file for the force field was proposed in section 3.

References

1. Mrozek, D., Małysiak-Mrozek, B., et al.: The Energy Distribution Data Bank: Collecting energy features of protein molecular structures. In: IEEE International Conference on Bioinformatics and Bioengineering, BIBE 2009, Taiwan, pp. 1–6 (2009)

2. Leach, A.: Molecular Modelling: Principles and Applications, 2nd edn. Pearson Education EMA, UK (2001)

3. Berman, H.M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T.N., Weissig, H., et al.: The Protein Data Bank. Nucleic Acids Res. 28, 235–242 (2000)

4. Mrozek, D., et al.: Searching for Strong Structural Protein Similarities with EAST. Journal of Computer Assisted Mechanics and Engineering Sciences 14, 681–693 (2007)

5. Znamirowski, A.W., Znamirowski, L.: Two-Phase Simulation of Nascent Protein Folding. In: Proc. of the 4th IASTED Inter. Conference on Modelling, Simulation, and Optimization 2004, Kauai, Hawaii, pp. 293–298. ACTA Press (2004)

6. Mrozek, D., Małysiak, B., Kozielski, S.: Energy Profiles in Detection of Protein Structure Modifications. In: IEEE International Conference on Computing and Informatics, Kuala Lumpur, pp. 1–6 (2006)

7. Branden, C., Tooze, J.: Introduction to Protein Structure. Garland (1991)

8. Allen, J.P.: Biophysical Chemistry. Wiley-Blackwell (2008)

9. Apweiler, R., Bairoch, A., Wu, C.H., Barker, W.C., et al.: UniProt: the Universal Protein knowledgebase. Nucleic Acids Res. 32(Database issue), D115–D119 (2004)

10. Wu, C.H., Yeh, L.S.L., Huang, H., Arminski, L., et al.: The Protein Information Resource. Nucleic Acids Research 31, 345–347 (2003)

11. Peptide/Protein Sequence Database (PRF/SEQDB), The Protein Research Foundation, `http://www4.prf.or.jp/en/pn.html`

12. Wheeler, D.L., Chappey, C., Lash, A.E., Leipe, D.D., et al.: Database resources of the National Center for Biotechnology Information. Nucleic Acids Res. 28(1), 10–14 (2000)

13. Lodish, H., Berk, A., Zipursky, S.L., et al.: Molecular Cell Biology, 4th edn. W. H. Freeman and Company, New York (2001)

14. Creighton, T.E.: Proteins: Structures and molecular properties, 2nd edn. Freeman, San Francisco (1993)

15. Boutselakis, H., Copeland, J., et al.: E-MSD: the European Bioinformatics Institute Macromolecular Structure Database. Nucleic Acids Res. 31, 458–462 (2003)

16. Marchler-Bauer, A., Addess, K.J., Chappey, C., et al.: MMDB: Entrez's 3D structure database. Nucleic Acids Res. 27(1), 240–243 (1999)

17. Bourne, P.E., Berman, H.M., Watenpaugh, K., et al.: The macromolecular Crystallographic Information File (mmCIF). Methods Enzymol. 277, 571–590 (1997)

18. Ohkawa, H., Ostell, J., Bryant, S.: MMDB: an ASN.1 specification for macromolecular structure. Proc. Int. Conf. Intell. Syst. Mol. Biol. 3, 259–267 (1995)

19. Wesbrook, J., Ito, N., Nakamura, H., et al.: PDBML: the representation of archival macromolecular structure data in XML. Bioinformatics 21(7), 988–992 (2005)

20. Cornell, W.D., et al.: A Second Generation Force Field for the Simulation of Proteins, Nucleic Acids, and Organic Molecules. J. Am. Chem. Soc. 117, 5179–5197 (1995)

21. Bray, T., Paoli, J., et al.: Extensible Markup Language (XML) 1.0 (5th edn.). W3C Recommendation (November 26, 2008), `http://www.w3.org/XML` (visited 2009)

22. Ponder, J.W., Case, D.A.: Force Fields for Protein Simulation. Adv. Prot. Chem. 66, 27–85 (2003)

23. Berglund, A., Boag, S., Chamberlin, D., et al.: XML Path Language (XPath) 2.0. W3C Recommendation (January 23, 2007), `http://www.w3.org/TR/xpath20/` (visited 2009)

24. Boag, S., Chamberlin, D., Fernández, M.F., et al.: XQuery 1.0: An XML Query Language. W3C Recommendation (January 23, 2007), `http://www.w3.org/TR/xquery/` (visited 2009)

Automatic Music Transcription Based on Wavelet Transform

Amir Azizi[1], Karim Faez[2], Amin Rezaeian Delui[3], and Saeid Rahati[1]

[1] Islamic Azad University Mashhad Branch
Amirazizi_b@Yahoo.com, Rahati@mshdiau.ac.ir
[2] Amir Kabir University of Technology
Kfaez@aut.ac.ir
[3] Toos Institute of Higher Education
Rezaeian@Toos.ac.ir

Abstract. In this paper, we introduce a method which uses a note model and signal post processing for a musical instrument to make a piece of music .one of the important issues in note transcription is extraction of multiple pitches. Most of the examined methods face error in joint harmonics and frequencies. A good model for note of a specified musical instrument can help us identify a note better. The presented method is based on wavelet transform, onset detection, note model and conformity reduction error algorithm or regression and post-processing for improved result. The results obtained show that detecting musical notes in a piece played on the guitar is, in comparison with similar methods, of higher detection accuracy and even in the case of noisy sound signals, the results are more acceptable.

Keywords: Note Transcription-Music-Signal Processing-Wavelet Transform-Onset Detection.

1 Introduction

Music Transcription is by definition listening to music and writing its note in a symbolic format (Such as midi or music line).it takes less space than sound files and also needs less processing for music re-compositions.

This technique is used for:

- Online Music Learning
- Music recording by midi format which has less space than other formats and is used in cell phones.
- Query by humming: use of music searcher for finding music on database.
- Creating music for a song by use of singer's voice.

Music transcription is classified into:

- Monophonic: Music is created by one instrument .this instrument plays one note at a time .many researchers have been done in this field.
- Polyphonic: Music is played by one instrument, but several notes can be played at the same time.
- Multi instrument: many instruments create the music at the same time.

For the latter, there are many fields to do research .there is no certain conceptual mechanism for separating the notes and instrument, because of lack of computational model for assessing this process.

2 Concepts

- Note*:* For all of the musical notes, there are only seven names. Some countries use syllabic nomination and some use alphabetical nomination.
  - o *Sylabic:*Do Re Mi Fa Sol La Si
  - o *Alphabetical :*C D E F G A B
- Octave*:* There are seven music notes, but the range of sound contains over 60 notes. Therefore repetition of the notes is used after the end of the last note, it is started from the first note (after Si, comes Do and…).
- Frequency relationships*:* As mentioned before, every instrument has several octaves, and every octave in Bach Pitch (Western-Music) has 12 halftones. Fig 1 shows the notes with their frequency in Piano.
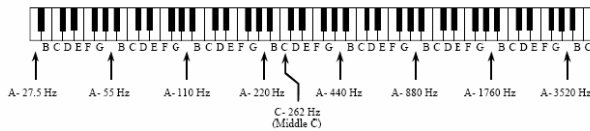
Fig. 1. Notes in Piano

The distance between each note and the lower halftone is $\sqrt[12]{2}$. Thus; the frequency of the notes of each octave is twice frequency of the same note in the lower octave.

- Onset: The sound that is created while the note is played. For example in Trumpet the sound is created when fingers are on a certain note and we below into the Trumpet and in Guitar, the sound is created in the first moment of playing.
- Offset: The sound which exists after playing the note. For example in guitar, after the stroke hits the string until the string stops trembling.
- Duration: When a note is played and has the same time value. In this paper, duration means the time between two onsets which covers the general meaning too.

3 Previous Works

Many researches have been done in the field the first works were only limited to monophonic.Marolt & Privosnik [1, 2, 3]: suggested a system that uses a combination of onset detection and note recognition.Klapuri [4]: suggested the other system is based on algorithm of repetition of smoothing spectrum and reduction. This method is used for the estimation of multiple pitches of simultaneous notes .this is one of the

best algorithms. Kawahara & Alain [5]: used philosophical balance principles of Yin-Yang to nominate it .its authors attempt to create balance between correlation and elimination in algorithm. The problem with correlation techniques was that the peaks occurred under the harmonics and sometimes it was difficult to recognize whether the peak was in basic frequency or in harmonics or details.Cepstrum [6]: is one of the most important methods. Theory of this method emphasizes this fact that transformation of Fourier of a signal with pitch has usually some regular peaks which shows signal harmonic spectrum. Harmonic Product Spectrum which was presented by Noll in 1969[7] is the easiest method to implement and it is used a lot.ML (Maximum Likelihood [7]): is another method that searches for the possible ideal spectrum and select the one which has the best conformity with the entry spectrum. Ideal spectrum is: A blow with frequency ω which is multiplied by window spectrum of a signal.

4   Proposed Method

The proposed method, in this paper is based on wavelet transform .in Fourier analysis, a constant length of window was considered and this window was the same for all frequencies. The gained information from this analysis for many signals, whose frequency changes should be recorded accurately, is not very significant. If we consider the time information of signal and frequency information, considering the length of variable windows, analysis of the signal will be executed more accurately. Wavelet analysis has considered this step and tries to solve the deficiency of Fourier [7].

4.1   The Relationship of Scale in Wavelet and Note Frequency

The distance between the frequency of scale wavelet and note frequency of two adjacent note in an octave is $\sqrt[12]{2}$. It means a note with higher frequency, has a frequency equal to the frequency of the lower note multiplied by $\sqrt[12]{2}$. We can see this relation in the analysis of wavelet of the notes in an instrument [8]. This relation is used for solving the problem of note recognition.

4.2   Recognition of Onset by the Use of Wavelet

By the use of wavelet transform, we can obtain onset for plucked instruments with recognition of the accepted threshold. First we obtain DWT transform from signal in a certain level (this level was obtained in different experiments and signal samples and the best amount was 24).

- Second, by use of Wavelet, we denoise the signal.
- In the next step, we detect the onset amounts through threshold.

5   Experimental Results

The proposed method is divided onto two parts training part and validation part:

5.1 Training Step

In this part, First of all, the notes will be played orderly with a certain approximate interval.Then by using the onset detection we detect the notes, and then we gain the note feature. Fig 2 shows the chart of training system.

We can receive the sound on line through the microphone or by a sound file .for the first part, we should consider the sampling rate sufficiently in order to cover all of the notes .for the second part we should know the file structure. Mostly, the windows synthesizers are used to test the method. First midi file is written by certain software, and then they are transformed into wav file. The sampling rate from signal is considered 22050.the maximum duration of each note is 250 mili-second .it means in every second; four notes are played at the most. In experiments, two groups of scale are considered. The first group is scales with 1 distance of pitch and has 80 scales: 1, 2, 3... 80. The second group is scales with distance of pitch $\sqrt[12]{2}$ this starts from 2.8904 and has 65 scales in order to cover all the changes in coefficients of the notes.
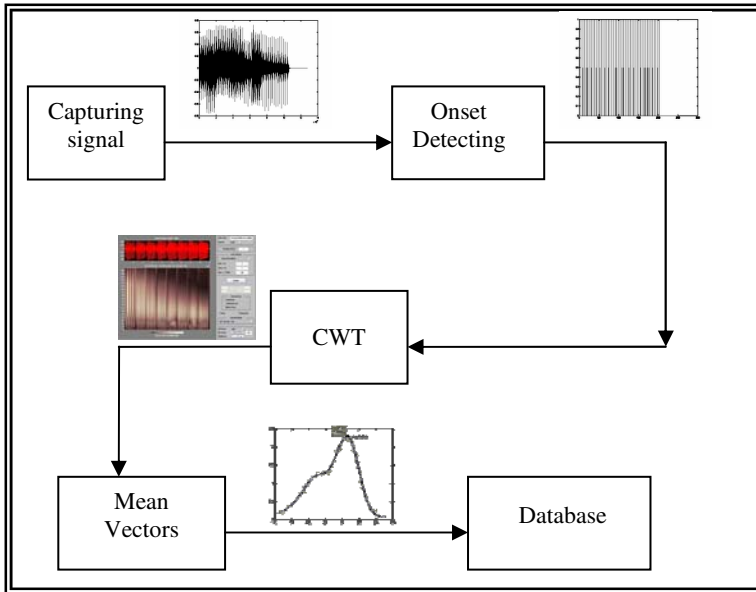
Fig. 2. Training Step

5.2 Validation Step

In this part, the notes are played and the system must detect them. Fig 3 shows this step. This step is very similar to the Training step, the other transformation which is observed here is the limitation that we motioned before .it means in every second, and only four notes are played.
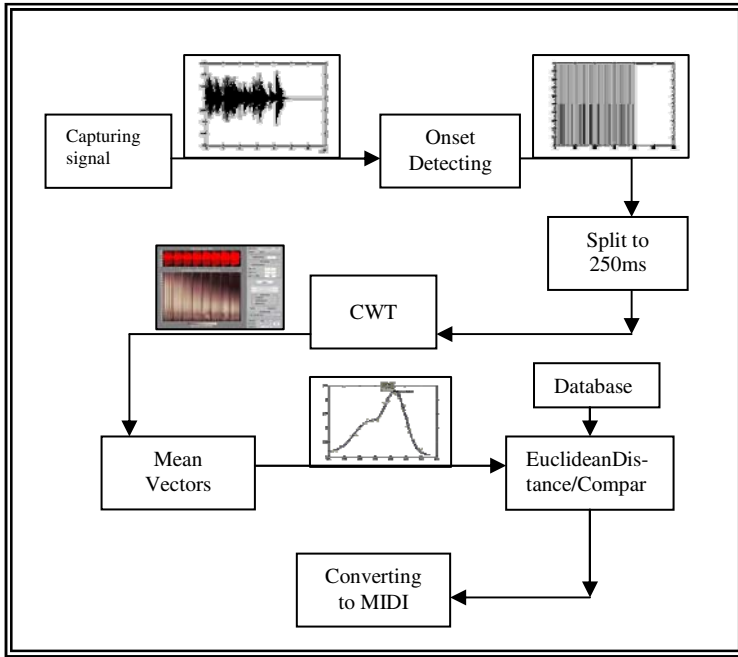
Fig. 3. Validation Step

Therefore, to analyze CWT faster with less memory space, before using CWT we divide the entering wav file into pieces (250 millisecond).The sampling rate is 22050, therefore 250 millisecond is equal to 22050/4=5512 samples from the signal. To detect the note in these 250 milliseconds, first we transform CWT of this piece of signal with the same signal (65 or 80:65 with distance of pitch at $\sqrt[12]{2}$ and 80 with distance of pitch 1). Then we obtain the mean vector of the coefficients in length of this vector of scale. We use the existing vectors in the bank from the lowest frequency, to gain the amounts. If the Euclidean distance of this vector and other vectors in the bank were less than threshold, the note of that vector is played. Then, we reduce the ratio of the mean vector in the validation step divided by mean of the detected for the entire coefficient in all scales for this piece of CWT matrix. The results related to onset detection for the stimulated instrument (Classic Guitar) with the same limitation is about 95%. We obtain error as follows:

$$E = \frac{Substituation + Deletion + Insertion}{Notes} \tag{1}$$

The proposed method is used for instrument whose played note is reduced in duration of time. The error rates for the following three files in two groups of scale are as follows:

Table 1. Result of Experiments

TestFile1		TestFile2		TestFile3	
80 scale	65 scale	80 scale	65 scale	80 scale	65 scale
0.32	0.21	0.45	0.42	0.22	0.12

6   Conclusion

In this paper, an initiative method is used by the use of note model for an instrument. This method is based on wavelet and uses onset detection in order to increase the frequency accuracy for every duration of the accuracy of multiple pitch estimation. To use note model, it is necessary that the user play the notes one by one in duration of time, with a recognized proportion .then the played note will enter the system. As for future research, we can also use other methods of transform such as FFT which have more resistance against the harmonic notes.

References

1. Marolt, M.: A connectionist approach to automatic transcription of polyphonic piano music, University of Ljubljana (2003)
2. Marlot, M., Divjak, S.: On detecting repeated notes in piano music University of Ljubljana, IRCAM (2002)
3. Marlot, M.: Networks of adaptive oscillators for partial tracking and transcription of music recording, University of Ljubljana (2003)
4. Klapuru, A.: Virtanen.T, Holm.J.M.: Robust Multipitch Estimation for the Analysis and Manipulation of Polyphonic Musical Signals. In: COST-G6 Conference on Digital Audio Effects, December 7–9 (2000)
5. Yin, J., Sim, T., Wang, Y.: Music transcription using an instrument model, National University of Singapore, ICASSP (2005)
6. Correa, J.P.B., Mary, Q.: Towards the automated analysis of simple polyphonic music: A Knowledge-based Approach, University of London, Thesis for the degree of Doctor of Philosophy (2003)s
7. Grimaldi, C.M., Kokaram, P.A.: A wavelet packet representation of audio signals for music genre classification using different ensemble and feature selection techniques, Trinity College Dublin, MIR (2003)
8. Fitch, J., Shabana, W.: A wavelet-based pitch detector for musical signals, Department of Mathematical Sciences, University of Bath

Appendix A: Exprimental Result for Classic Guitar
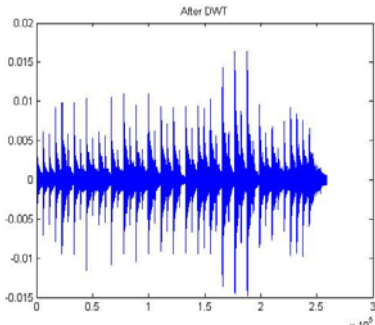
Music signal after DWT
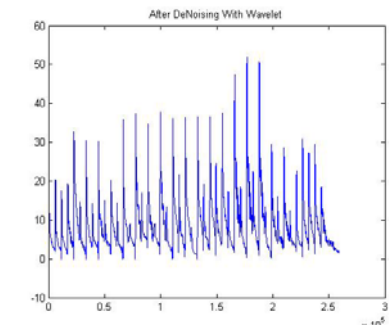
Denoising Music signal with wavelet (mexh)
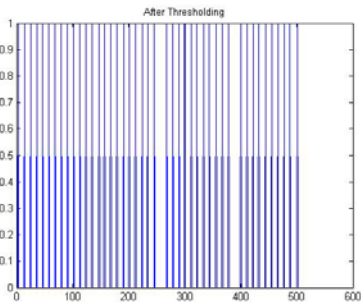
After Thresholding
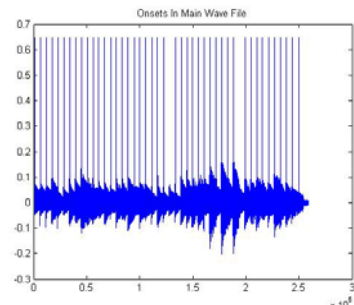
Onset Detection

Appendix B: Result of Classic Guitar Note Transcription in Piano Roll Form
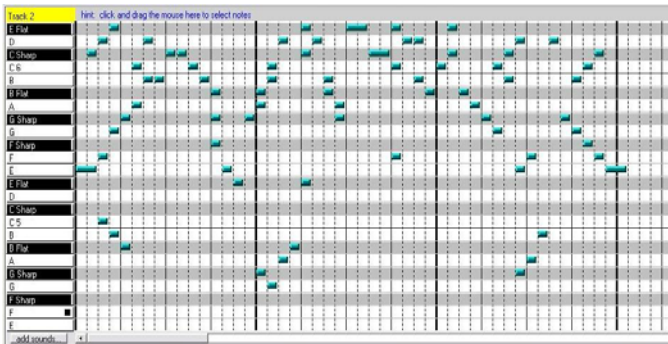
Played Music
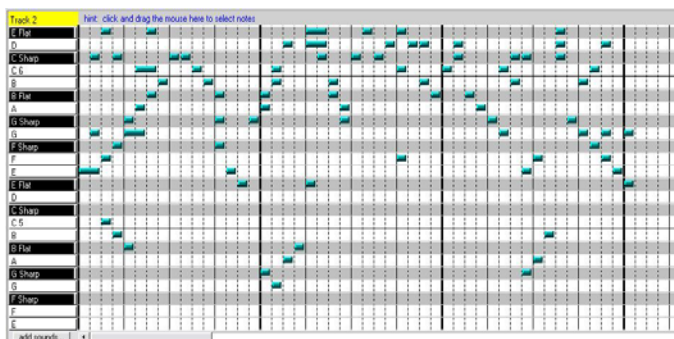
Result in 65 Scales
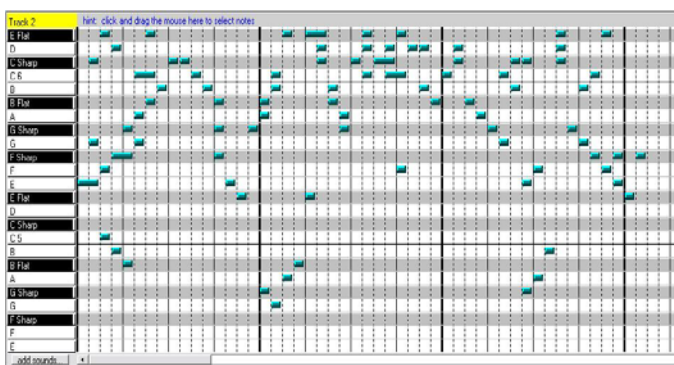
Result in 80 scales

Synthesis of Bowhead Whale Sound Using Modified Spectral Modeling

Pranab Kumar Dhar[1], Sangjin Cho[2], and Jong-Myon Kim[1,*]

[1] School of Computer Engineering and Information Technology
pranab_cse@yahoo.com, jongmyon.kim@gmail.com
[2] School of Electrical Engineering
sjcho75@ulsan.ac.kr

Abstract. Spectral modeling synthesis (SMS) considers a sound as a combination of a deterministic plus a stochastic component that makes possible for a synthesized sound to attain all the perceptual characteristics of the original sound. However, sometimes considerable phase variations occur in the deterministic component by using SMS since the addition of different frequency sinusoids in the overlap region causes amplitude distortion. As a result, subtraction between original and deterministic signal in time domain do not provide a good approximation of the residual signal. To overcome this problem, we propose a modified SMS that provides good approximation of the residual signal by calculating the complex residual spectrum in frequency domain. Analysis and simulation results for synthesizing bowhead whale sounds suggest that the proposed method is comparable to the SMS in both time and frequency domain. However, the proposed method outperforms the SMS in better spectrum matching because of the use of original phase information to synthesize the deterministic component as well as good approximation of the residual signal by subtracting the deterministic spectrum from the original spectrum and then utilizing spectral fitting.

Keywords: Spectral modeling synthesis, whale sound synthesis, phase variation, short time Fourier transform, additive synthesis.

1 Introduction

Bowhead whales are one of the most important marine mammals found in five separate populations in the Arctic Ocean, migrating north and south with the seasonal movement of the edge of the pack ice. They rely on sound for communication, navigation, or detection of predators while using sound to attract mates, repel rivals, communicate within a social group or between groups, navigate, or find food. They can produce distinct sounds, such as songs, moans etc. Bowhead whale sounds have been recorded from Canadian Beaufort Sea and these sounds consist of various low-frequency (25 to 900 Hz) moans and well defined sound sequences organized into songs (20-5000 Hz) [10].

---

[*] Corresponding author.

The spectral modeling synthesis (SMS) extracts the synthesis parameters out of real sounds using analysis procedures, being able to reproduce and modify actual sounds. This approach is based on modeling sounds as stable sinusoids (partials) plus noise (residual components) to analyze sounds and generate new sounds. The analysis procedure detects partials by utilizing the time-varying spectral characteristics of a sound, and represents them with time-varying sinusoids [5], [6], [8]. These partials are then subtracted from the original sound where the remaining residual is represented as a time-varying filtered white noise component. The synthesis procedure is a combination of additive synthesis for the sinusoidal part and subtractive synthesis for the noise part [1], [2]. However, sometimes phase variations occur in the deterministic component when we generate it using the SMS. This is because the addition of different frequency sinusoids in the overlap region can result in an amplitude distortion due to the phase cancellation [9]. As a result, subtracting the deterministic signal from the original signal cannot provide a good approximation of the residual signal.

To overcome this problem, we propose a modified SMS that utilizes phase information to synthesize the deterministic component as well as good approximation of the residual signal by subtracting the deterministic spectrum from the original spectrum and then using spectral fitting. The stochastic signal is generated by using an inverse short time fourier transform (STFT) on a series of magnitude spectrum envelopes that function as a time varying filter excited by white noise. We then add the deterministic and stochastic signal in time domain for each frame. In this paper, we synthesize bowhead whale sounds using the proposed method and compare the proposed method to the SMS technique. The result of the synthesis sound indicates that the proposed method are comparable to the SMS in both time and frequency domain. However, the proposed method outperforms the SMS in better spectrum matching with original spectrum because of the use of original phase to synthesize deterministic component and better approximation of the residual signal. Synthesis of whale sounds is a new idea in the field of whale sound modeling. Using the modified SMS technique we can efficiently generate the synthesized whale sound which resembles much more closely the original sound. The synthesized whale sound can be further used to create whale music.

The rest of this paper is organized as follows. Section 2 presents background information regarding the deterministic plus stochastic model and a general overview of the SMS analysis and synthesis process. Section 3 presents our proposed method for the higher quality of whale sound synthesis. Section 4 summarizes and discusses experimental results of the bowhead whale sound for both the SMS and the proposed method, and Section 5 concludes this paper.

2   Background Information

2.1   Deterministic Plus Stochastic Model

A sound model assumes certain characteristics of the sound waveform or the sound generation mechanism. Sounds produced by musical instruments, any physical system, or any human voice can be modeled as the sum of a set of sinusoids plus a noise residual. The sinusoidal or deterministic component normally corresponds to the main

modes of vibration of the system. The residual comprises the energy produced by not stationary vibrations plus any other energy component that is not sinusoidal in nature.

A deterministic signal is traditionally defined as anything that is not noise. A stochastic or noise signal is fully described by its power spectral density which gives the expected signal power versus frequency. When a signal is assumed stochastic, it is not necessary to preserve the instantaneous phase. This model considers a waveform signal $s(t)$ as the sum of a series of sinusoids plus a residual $e(t)$, which is defined as

$$s(t) = \sum_{r=1}^{R} A_r(t)\cos[\theta_r(t)] + e(t)$$

where R is the number of sinusoids, $A_r(t)$ and $\theta_r(t)$ is the instantaneous amplitude and phase of the rth sinusoid, respectively, and $e(t)$ is the noise component at time t (in seconds).

The model assumes that the sinusoids are stable partials of the sound, and each one has a slowly changing amplitude and frequency. The instantaneous phase is taken to be the integral part of the instantaneous frequency $\omega_r(t)$ and therefore satisfies

$$\theta_r(t) = \int_0^t \omega_r(\tau)d\tau$$

where $\omega_r(t)$ is the frequency in radians and r is the sinusoidal number.

By assuming that $e(t)$ is a stochastic signal, it can be described as a filtered white noise,

$$e(t) = \int_0^t h(t,\tau)u(\tau)d\tau$$

where $u(t)$ is the white noise and $h(t,\tau)$ is the response of a time varying filter to an impulse at time t. Thus, the residual signal is modeled by the convolution of white noise with time varying frequency-shaping filter [1], [3].

2.2   A General Overview of the SMS Analysis and Synthesis Process

The deterministic plus stochastic model supports many possible implementations. Both analysis and synthesis models are the frame-based process with the computation done one frame at a time. Figure 1 shows a block diagram for the SMS analysis process. We have analyzed the sound by multiplying it with an appropriate analysis window. Its spectrum is obtained by fast fourier transform (FFT) and then the prominent spectral peaks are detected and incorporated into the existing partial trajectories by the mean of a peak continuation algorithm. It detects the magnitude, frequency, and phase of the partials presented in the original sound (the deterministic components). When the sound is pseudo harmonic, a pitch detection step can improve the analysis by utilizing the fundamental frequency information in the peak continuation algorithm as well as by selecting the size of the analysis window [1], [2], [3].

The stochastic component of the current frame is calculated by generating the deterministic signal with additive synthesis and then subtracting it from the original
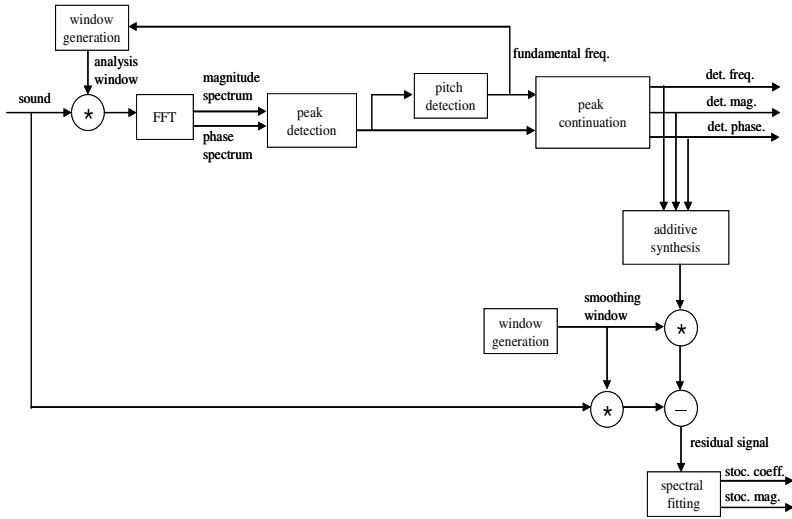
Fig. 1. Block diagram of the SMS analysis process
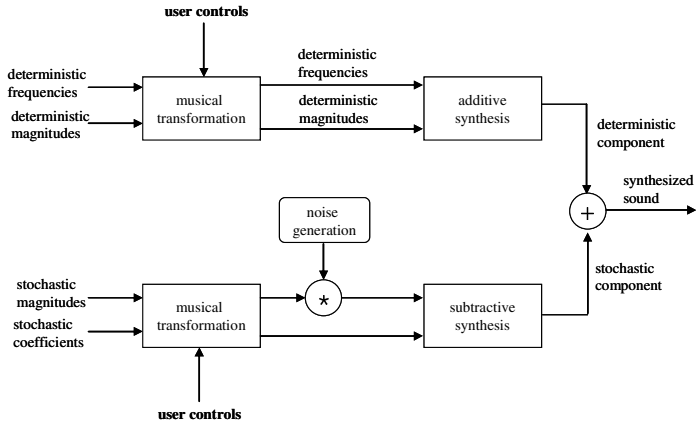
Fig. 2. Block diagram of the SMS synthesis process

waveform in time domain. The stochastic representation is then obtained by performing a spectral fitting of the residual signal.

Figure 2 shows a block diagram of the SMS synthesis process. The deterministic component (sinusoidal component) is calculated from the frequency and magnitude trajectories. The result of the synthesized stochastic signal is a noise signal by time varying spectral shape obtained in the analysis (i.e., subtractive synthesis). It can be implemented by a convolution in time domain or by a complex spectrum for every spectral envelope of the residual and an inverse-FFT in frequency domain.

3   Proposed Method

To provide better approximation of the residual signal, we propose a modified SMS that calculates the complex residual spectrum in frequency domain. We can obtain the stochastic representation of the residual signal by subtracting the deterministic spectrum from the original spectrum and utilizing spectral fitting (line segment approximation) of the magnitude spectrum.

In the synthesis process, the deterministic signal is calculated by a sine wave for each magnitude, frequency, and phase trajectory. The stochastic signal is calculated by a complex spectrum envelope of the residual and an inverse STFT. We then add the deterministic component with stochastic one using an overlap add method [4], [7] in time domain for each frame to obtain the synthesized bowhead whale sound. Figures 3 and 4 show the analysis and synthesis processes of the proposed method, respectively.

The success of the analysis process depends on the selection of the program parameters such as STFT window, window size, hop size, and the number of peaks to
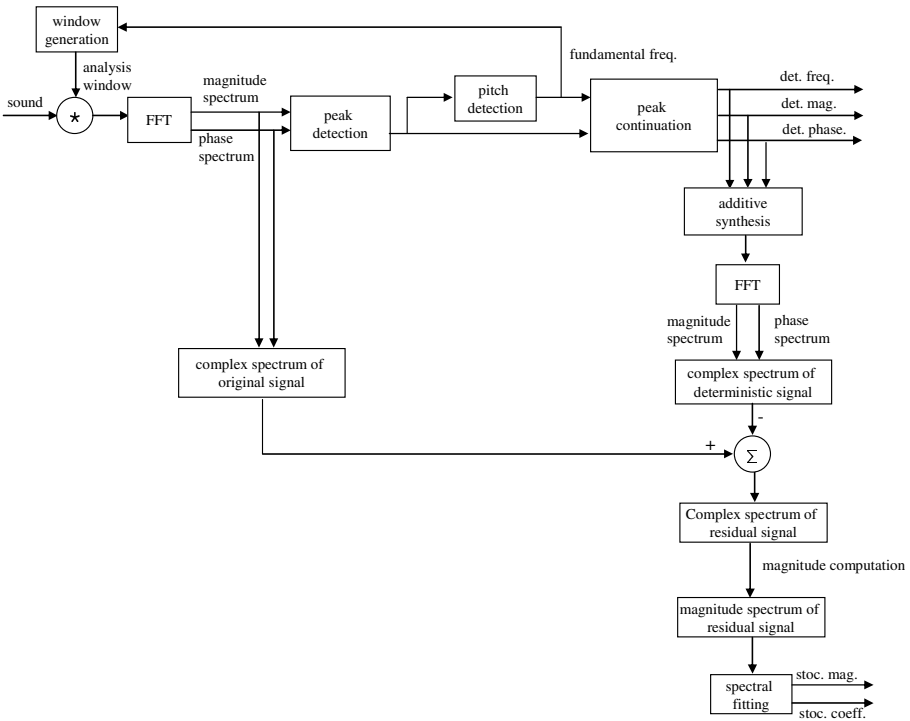
Fig. 3. Block diagram of the analysis process in the proposed approach
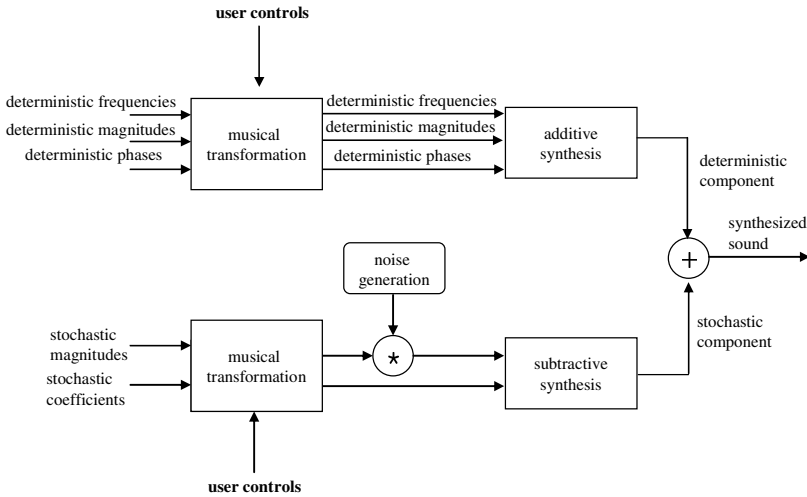
Fig. 4. Block diagram of the synthesis process in the proposed method

be detected. One among successful parameter set is the hanning window, window size of 512, hop size of 256, and the number of peaks are 80. These selected parameters provide a better result for the bowhead whale sound analysis.

4   Results and Discussion

In this section, we evaluate the performance of our proposed method to synthesize the bowhead whale sound, and compare the proposed method to the SMS. The metrics of time domain representation, frequency domain representation, spectrum matching, and listening of each case form the basis of the study comparison.

We observe that both the proposed method and the SMS generate a good synthesized sound which resembles much more closely the original sound. However, phase variations were occurred in the deterministic component when we generated it using the SMS. This is because the addition of different frequency sinusoids in the overlap region causes amplitude distortion due to the phase cancellation. This results in not providing a good spectrum matching and approximation of the residual signal. Figures 5, 6, 7 illustrate time domain, frequency domain, and spectrum matching of the original and synthesized bowhead whale sounds using the SMS, respectively.

The proposed method overcomes this problem by calculating the complex residual spectrum in frequency domain, providing better spectrum matching and approximation of the residual component. Figures 8, 9, 10 illustrate time domain, frequency domain, and spectrum matching of the original and synthesized bowhead whale sounds using the proposed method, respectively.
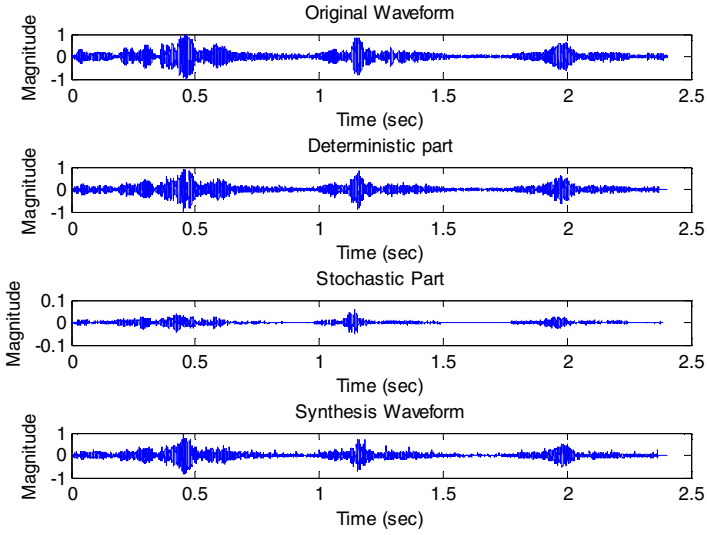
Fig. 5. Time domain representation of original, deterministic, stochastic and synthesized bowhead whale sound using SMS
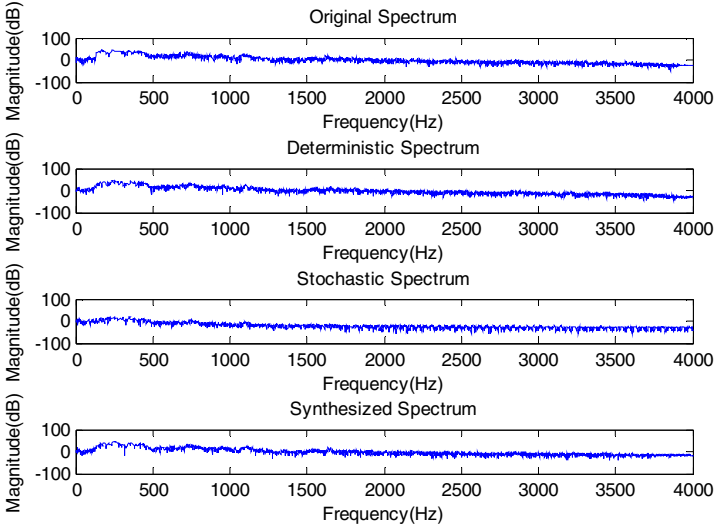
Fig. 6. Frequency domain representation of original, deterministic, stochastic and synthesized bowhead whale sound using SMS

Figures 11 and 12 represent the spectrum matching of the original and synthesized sounds for 0 to 1 KHz and 3 to 4 KHz using the SMS and the proposed method, respectively. The proposed method provides better results than the SMS in spectrum matching of original and synthesized sound.
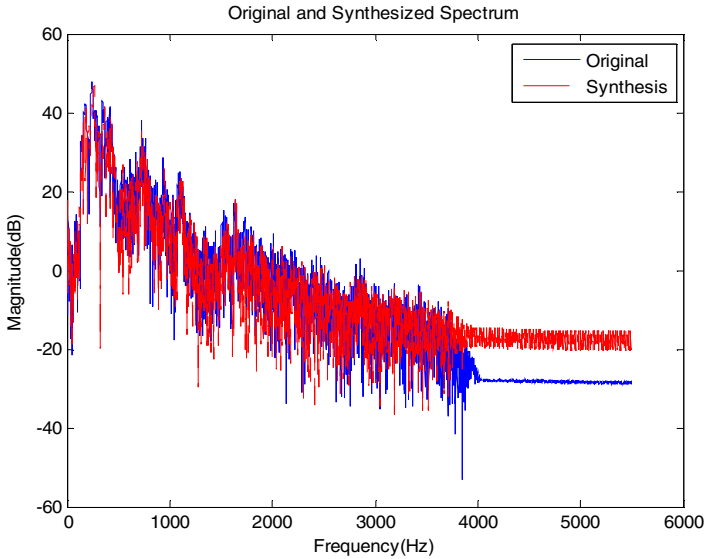
Fig. 7. Spectrum matching of original and synthesized bowhead whale sound using SMS
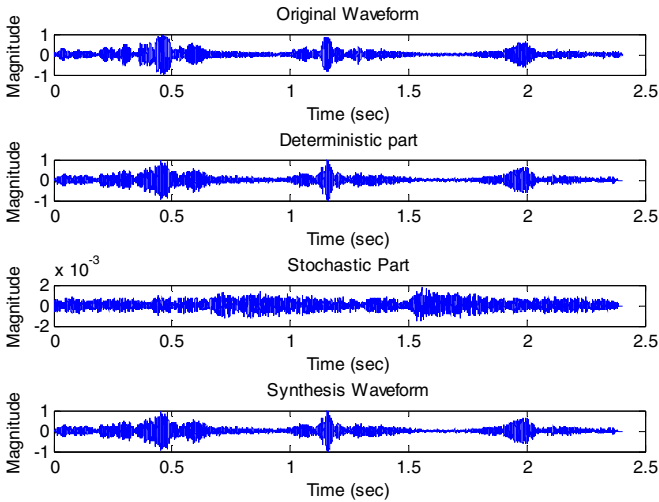
Fig. 8. Time domain representation of original, deterministic, stochastic and synthesized bowhead whale sound using the proposed method

In this implementation, the most important frequency range of the sound is 150 to 500 Hz because most of the energies are concentrated in this frequency range. The highest peak frequency of this sound is 245 Hz. From the figure 11(a), we observe that the SMS cannot detect this peak as a sinusoid component. This results in
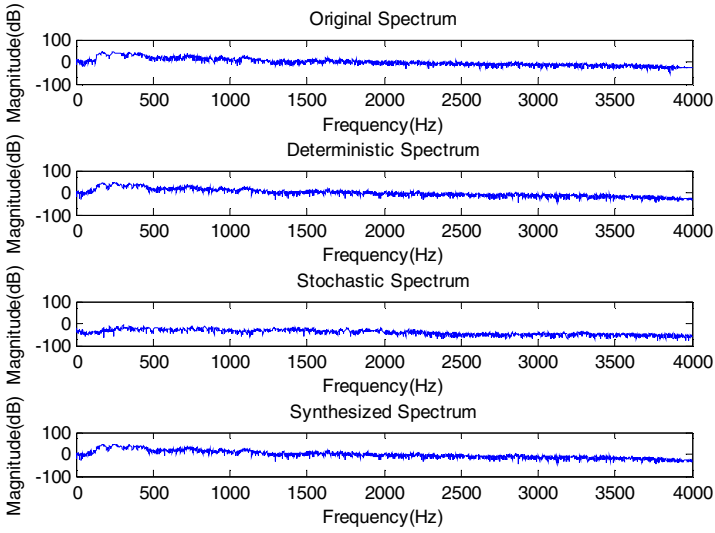
Fig. 9. Frequency domain representation of original, deterministic, stochastic and synthesized bowhead whale sound using the proposed method
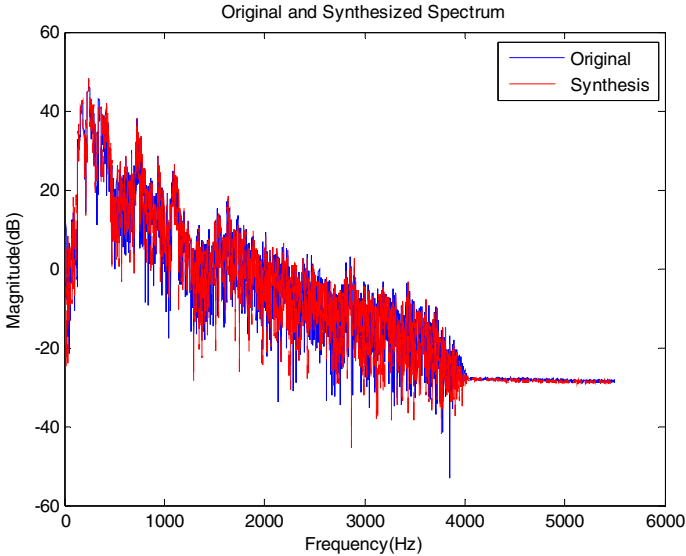
Fig. 10. Spectrum matching of original and synthesized bowhead whale sound using the proposed method

considering this peak as a residual component in the synthesized sound. In contrast, the proposed method correctly detects this peak as sinusoid as shown in Figure 12(a). We also observe that between 3 to 4 KHz, our proposed method provides better spectrum matching with the original spectrum over the SMS as shown in Figures 11(b)
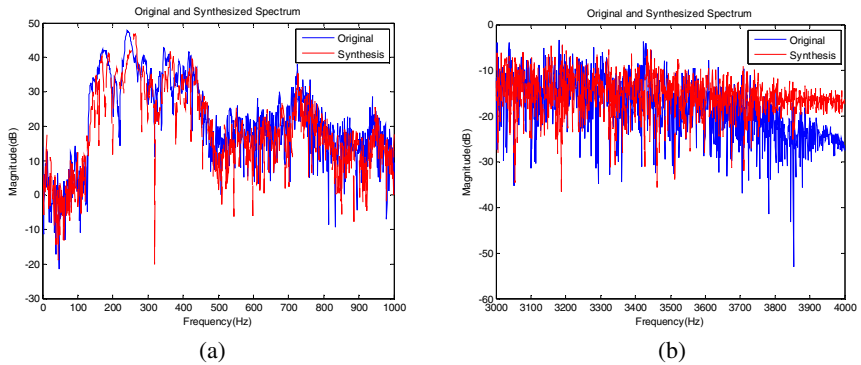
Fig. 11. Spectrum matching of the original and synthesized sound using SMS
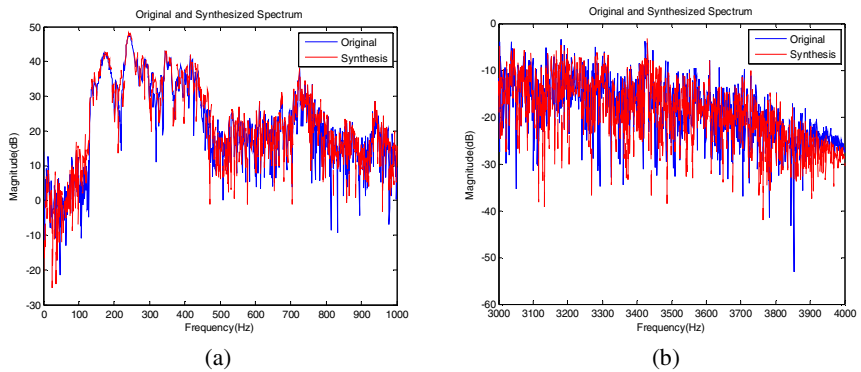
Fig. 12. Spectrum matching of the original and synthesized sound using the proposed method

and 12(b). This is because our proposed method utilizes original phase information to synthesize deterministic component and also provides a better approximation of residual signal. Overall, the proposed method outperforms the SMS in spectrum matching of the original and synthesized bowhead whale sound.

5 Conclusions

The spectral modeling synthesis (SMS) gives us a powerful starting point in studying bowhead whale sound modeling. However, the SMS has occurred considerable phase variations in the deterministic component because the addition of different frequency sinusoids in the overlap region causes amplitude distortion. So subtraction of the de-terministic signal from the original signal in time domain made some errors to approximate the residual component. To overcome this problem, we have proposed a modified SMS which calculates the complex residual spectrum in frequency domain. Analysis and simulation results for bowhead whale sound synthesis suggest that the proposed method outperforms the SMS in spectrum matching between the synthesized

spectrum and the original spectrum because the proposed method utilizes original phase information to synthesize the deterministic component and better approximation of the residual signal.

Acknowledgement

References

1. Serra, X.: Musical Sound Modeling with Sinusoid Plus Noise. In: Roads, C., Pope, S., Picialli, A., Poli, G.D. (eds.) Musical Sound Processing. Studies on New Music Research, pp. 91–122. Sweets and Zeitlinger Publishers (1997)
2. Serra, X., Smith, J.: Spectral Modeling Synthesis: A Sound Analysis/Synthesis System Based on a Deterministic Plus Stochastic Decomposition. Computer Music Journal 14(4), 12–24
3. Serra, X.: A System for Sound Analysis/Transformation/Synthesis based on a Deterministic plus Stochastic Decomposition. Ph.D Thesis, Stanford University (1989)
4. George, E.B., Smith, M.J.T.: Analysis-by-synthesis/Overlap-add Sinusoidal Modeling Applied to the Analysis and Synthesis of Musical Tones. Journal of Audio Engineering Society 40(6), 497–516 (1992)
5. Depalle, P., Garcia, G., Rodet, X.: Tracking of Partials for Additive Sound Synthesis Using Hidden Markov Models. In: Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), vol. 1, pp. 225–228 (1993)
6. Allen, J.B.: Short Term Spectral Analysis, Synthesis and Modification by Discrete Fourier Transform. IEEE transaction on Acoustics, Speech and Signal Processing ASSP-25, 235–238 (1977)
7. McAulay, R.J., Quatieri, T.F.: Speech Analysis/Synthesis Based on a Sinusoidal Representation. IEEE transaction on Acoustics, Speech and Signal Processing 34(4), 744–754 (1986)
8. Allen, J.B., Lawrence, R.: A Unified Approach to Short-Time Fourier Analysis and Synthesis. Proceedings of IEEE 65, 1556–1564 (1977)
9. Goodwin, M., Rodet, X.: Efficient Fourier Synthesis of Nonstationary Sinusoids. In: Proceedings of International Computer Music Conference (ICMC), pp. 333–334 (1994)
10. Cummings, W.C., Holliday, D.V.: Sounds and Source Levels from Bowhead Whales off Pt. Barrow, Alaska. Journal of Acoustical Society of America 82, 814–821 (1987)

Automatic Emphasis Labeling for Emotional Speech by Measuring Prosody Generation Error[*]

Jun Xu and Lian-Hong Cai

Key Laboratory of Pervasive Computing, Ministry of Education
Tsinghua National Laboratory for Information Science and Technology
Dept. of Computer Science & Technology, Tsinghua University, Beijing 100084, China
xujun00@mails.tsinghua.edu.cn, clh-dcs@tsinghua.edu.cn

Abstract. Emotion helps human to express their feelings and intentions clearly. And the emphasis labels of speeches are the key of speech emotion analysis and synthesis. In order to label the emotion emphasis of speech samples from a corpus with only phonetic and prosodic information, this paper introduces an automatic labeling algorithm by measuring the prosody generation error (PGE) of the result from a statistical synthesizer. Classification and Regression Tree (CART) and Maximum Entropy (ME) modeling are adopted for automatically labeling. Experiment shows that both models are helpful for labeling.

Keywords: Emotion emphasis, Prosody generation error, Speech synthesis.

1 Introduction

During daily speech communication of human beings, emotion can be easily perceived to help understanding their feeling and intentions. With recent advances in speech technologies, computers are able to recognize and synthesize natural speech with high intelligent and articulation. However, it is yet an important task for them to understand and express the implicated emotion, depending on the communication environment.

Previous studies [1] pointed out that there are acoustic correlates between emotion and emphasis. Since emphasis acts as an important component of emotion, it is also preferred to be labeled as a content of emotion [2]. Moreover, it is widely accepted that time-domain features such as duration and pitch contour are most relevant to emotional emphasis [3]. Thus many studies made effort to model and label emphasis with such features [4] and applied them to speech processing [5].

As a matter of fact, when training speech models using statistical speech processors, including recognizer and synthesizer such as Context Oriented Clustering (COC) [6], the acoustic features from those speech samples with identical contextual information will be statistically averaged. In other words, the acoustic features of speech

---

samples within the same context will be averaged during model training, regardless of whether they are emphasized or not. Thus the corresponding generated speech with this model will be smoothed to have a consistent quality, which leads to a mismatch between the natural emphasized speech samples and the averaged generated ones. Since time-domain features are easier to observe and process, we define the mismatch of time-domain acoustic features as **Prosody Generation Error (PGE)**, and use them to label the emphasized speech samples.

2   Prosody Generation Error of Statistical Speech Synthesis

2.1   HMM Based Speech Synthesis

Hidden Markov Model (HMM) based speech synthesis [7] is one of the state-of-art statistical parametric speech synthesis methods. All acoustic features of speeches including excitation and spectra parameters are modeled in a unified framework of HMM. Unlike unit selection based concatenative synthesizer, which always carries a large scale corpus at synthesis time, HMM synthesizer is able to generate speech with a small context-dependent model, and reproduce the original speaker's voice characteristics.

As shown in Fig. 1. At training stage, the system extracts excitation and spectra parameters from the corpus. F0 is used as the excitation parameter and spectra parameters are often represented by Mel-cepstrum, Mel-frequency cepstrum coefficient (MFCC), or Line spectral pairs (LSP). These parameters are then used to train corresponding HMM models. Related contextual information is used for tree clustering over all the HMM models to create a context-dependent model. At synthesis stage, contextual information is extracted from the input text by a front-end text analyzer, and then it is used to select appropriate models from the tree for HMM decoding to generate speech parameters. At last, a speech vocoder uses these speech parameters to synthesize final speech.
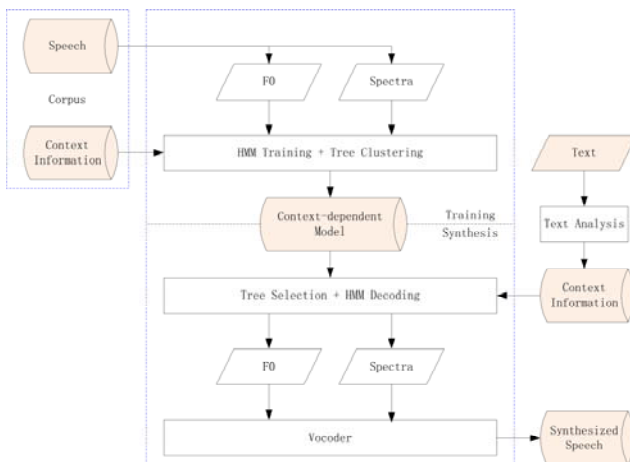
Fig. 1. Flowchart of HMM based speech synthesis

The contextual information we adopted for HMM speech synthesis contains the following aspects:

- Phonetic information: The initial, final and tone of current syllable;
- Phonetic environment: The phonetic information of neighboring syllables;
- Prosodic information: The prosodic hierarchy of Mandarin Chinese [8], such as prosody word and phrase count;
- Prosodic distance: The syllabic distance to prosodic word, phrase and utterance boundaries.

With above features, all samples from the corpus are extracted for modeling and reproduced to check the prosody generation error.

2.2   Prosody Generation Error

Since the statistical synthesizer generates the speech without emphasis information, the acoustic features of synthetic speech will be different from those are emphasized. In our experiment, the segment duration and pitch contour differences between generated parameter sequences and natural ones are computed as prosody generation error.
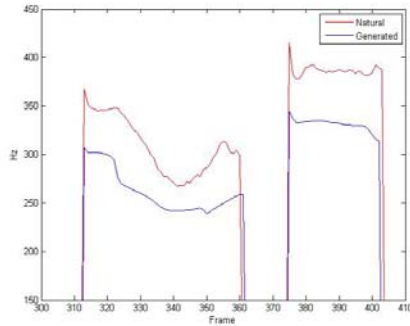
Fig. 2. Shift error of segment pitch sequences
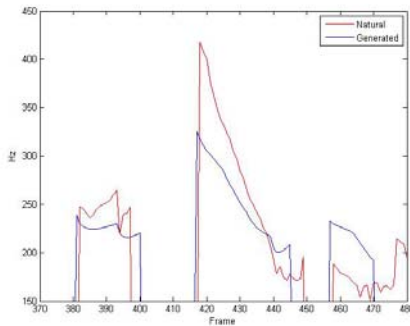
Fig. 3. Rotation error of segment pitch sequences
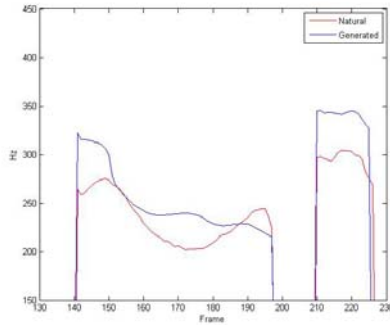
Fig. 4. Trajectory error of segment pitch sequences

   Duration is a scalar value for each segment, so frame count differences can be used to represent this mismatch.

   For pitch contours, we observe 3 types of generation error:

1. Shift error

The whole segment pitch contour is raised up or lowered down in some generated speech. As Fig. 2 shows, for this kind of error, the pitch data of each frame have a comparable incremental or decreased value. In other words, the differentials of two sequences keep almost the same.

2. Rotation error

The slope of the segment pitch sequences is different in some cases. Most of the time, the natural ones will have a steeper trajectory with a larger peak value. The derivative of the differential values keeps positive or negative all along the way in such situation. A sample is shown in Fig. 3.

3. Trajectory error

In fact, in most cases, the distortion of pitch contour cannot be easily classified. Sometimes the generated sequences have different peak or valley positions, and some of them even have different numbers of peaks or valleys. As illustrated in Fig. 4, the shape of pitch contour is totally different; however, we cannot simply say that the differences of pitch maximum, minimum, average and range follow some intuitive rules.

3   Automatic Labeling Based on CART and Max Entropy

Currently we already have contextual information a for each syllable, and more features b can be extracted from prosody generation error. And we aim to find out if a syllable sample is emphasized or not by these information.

   Consider a random process that produces an output value y based on these features x, with x and y being a member of a finite set X and Y respectively. In our

case, y is whether a syllable is emphasized, and x is the relative information about that syllable, which is the combination of a and b. Our task is to construct a stochastic model that accurately represents the behavior of the random process. In other words, it should give a reliable estimation of $p(y \mid x)$, which denotes the conditional probability that, given a context x, the process will output y.

For this purpose, we observe the behavior of the random process for some time, collecting N samples $(x_1, y_1), (x_2, y_2), \cdots, (x_N, y_N)$, and then use two models: CART and Max Entropy to represent such a random process.

1. CART Modeling

Suppose the feature vector x contains m features $v_1, v_2, \cdots, v_m$. The idea of CART tree building is to choose a best split value v_s among all of the features to make the resulting child nodes have the largest purity. In our case, obviously a tree node contains samples of which all are emphasized or un-emphasized has the largest purity.

If v_s is a categorical variable of K categories, there are $2^{K-1} - 1$ possible splits. If v_s is a continuous variable with K different values, there are $K - 1$ possible splits. Then the tree is grown starting from the root node by splitting a node into two child nodes repeatedly.A node will not be split when any of following conditions is satisfied:

- All the samples (x, y) belong to the same class $y = j$, which means they are totally pure and do not need to be classified.
- All the samples (x, y) have identical x, which means they have identical information that none of the predicators is able to split this node.
- User defined stopping conditions, such as tree depth, minimum size of tree node or minimum decrease of impurity, are met.

When all the nodes cannot be split, the tree is ready for use. However, an unreasonable big tree will introduce the over-fitting problem. Several validation and pruning technique are proposed to make a right sized tree.

2. Max Entropy Modeling

To express these facts of observed training samples using Max Entropy model, suppose the expected value of each feature f_i with respect to the statistics of training samples is $\bar{p}(f_i)$ and the the expected value of f_i with respect to the unknown model $p(y \mid x)$ is $p(f_i)$, then for each f_i

$$p(f_i) = \bar{p}(f_i) . \tag{1}$$

Requirement (1) is called a constraint equation or simply a constraint [10].

Suppose we have n features, then all the probability distribution that satisfy the constraints exerted by these features constitute a set C:

$$C \equiv \left\{ p(y \mid x) \mid p(f_i) = \overline{p}(f_i) \quad \text{for } i \in \{1, 2, \cdots, n\} \right\} . \tag{2}$$

Among all the models p in C, the maximum entropy philosophy dictates that we select the one with maximum conditional entropy:

$$H(p) \equiv -\sum_{(x,y)} \overline{p}(x) p(y \mid x) \log p(y \mid x) . \tag{3}$$

Considering $p^* = \arg \max H(p)$, it is a constrained optimization problem to find $p*$. The target maximum entropy model has the following form:

$$p^*(y \mid x) = \frac{1}{Z_\lambda(x)} \exp(\sum_i \lambda_i f_i(x, y)) , \tag{4}$$

where $Z_\lambda(x)$ is a normalizing constant and λ_i is a Lagrange multiplier which is commonly computed from the training set using GIS algorithm. Detailed steps are omitted here.

3.1   Feature Selection Strategy

To build an appropriate CART or ME model, it is important to locate most relevant information for our task. Besides the contextual features that we already utilized in statistical synthesis, Section 2.2 also introduced some observations of prosody generation error. The frame count difference can be directly used as the duration distortion feature, and we still need to find out what kinds of features are suitable for representing pitch distortion.

First of all, root mean square error (RMSE) is able to evaluate the differences between original and generated pitch contour. Without regard to the actual shape of trajectory, a larger RMSE often indicates a larger distortion. The RMSE of two sequences is defined as:

$$f_{RMSE}(S,T) = \sqrt{\sum_i (s_i - t_i)^2 / N} , \tag{5}$$

where S and T are two sequences with length N.

However, a large RMSE does not always stand for an emphasized sample. More pitch distortion features that are relevant to our specific application should be discovered. In our experiment, pitch features which are widely used in speech emotion analysis are considered to be important to represent the characteristics of emphasized samples. These features include maximum, minimum and average pitch value, as well as the range of pitch contour. The slope is also helpful for checking the rotation error of the sequence.

Besides the above features from prosody generation error, another type of features is also considered. It takes the prediction result of previous round into account, including the results of current and neighboring samples. The motivation of adding this information came from the observation that pitch contour is influenced by its

Table 1. Features used for building the models

Category	Features	Meaning & Values
Contextual features	PinW	Position inside a prosody word.
	PinP	Position inside a prosody phrase.
	PinS	Position in the utterance.
	Tone	The tone of current sample.
Prosody generation error based features	DurErr	Duration error in frame count difference.
	RMSE	RMSE of pitch sequences.
	MaxErr	Difference of maximum pitch.
	MinErr	Difference of minimum pitch.
	AvgErr	Difference of average pitch.
	RangeErr	Difference of the pitch range.
	SlopeErr	Difference of the pitch slope.
Prediction result based features	LastCur	Prediction result of current sample in last pass.
	LastLeft	Prediction result of left sample in last pass.
	LastRight	Prediction result of right sample in last pass.

neighboring samples. By contrast to those 'static' features that are fixed and known all along once parameters are generated, the features of this category remains unknown before a first round judgment is made. Table 1 lists all the features adopted in model building. Note that the selection of contextual features is based on our experiences on emotional emphasis, so they are much less than those used in speech synthesis in order to avoid over-fitting problem.

3.2 Resample and Two-Pass Prediction

It is obvious that the emphasized sample count is much smaller than those not emphasized ones. This introduces a severe problem that the model will prefer to refuse recognizing the sample to be emphasized. In an extreme situation, the model will judge all the samples to be un-emphasized, however the total precision considering all samples will be yet relatively high.
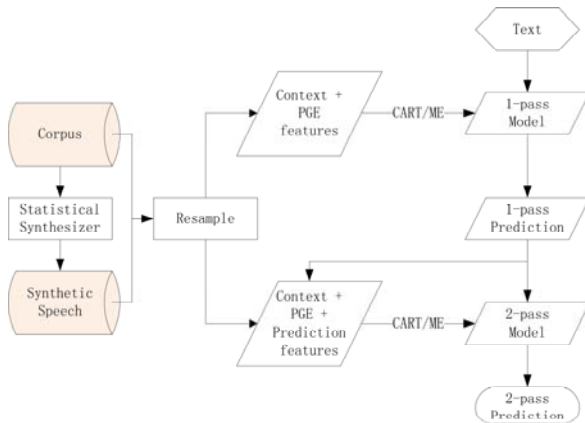
Fig. 5. Flowchart of two-pass prediction

In order to fix this bad tendency, we perform a resample strategy, which set a pre-defined weight for emphasized training samples and resample the training data.

Another observation is that an emphasized sample often leads to a high PGE of its neighboring samples, especially for pitch contour. It could be explained by the Pitch Target model's theory [9] that the pitch trajectory results not only from the underlying speech unit itself but also from the articulatory context that determine how this unit can be implemented. That is why we came up to the idea of two-pass labeling, of which the total procedure is shown in Fig. 5.

4   Experiment

Our corpus contains 4 sub sets with different emotions: angry, happy, sad and surprise. Each sub set has 220 utterances. The text is extracted from People's Daily 2000, and is labeled with phonetic identities (Tonal pinyin of Mandarin syllables) and prosody boundaries. These labels are automatic labeled with manual correction.

Emotion emphasis type is also labeled for the syllables but currently they are labeled manually. In a preliminary statistic the utterances contain 18 syllables on average, of which at most 4 syllables are emphasized in a single utterance but some utterances contain none emphasized syllables. That means the un-emphasized syllables are almost 10 times more than those are emphasized.

Among each sub set, we use 180 utterances for training and the remains for testing. Note that because emphasis rarely appears in sad emotion, we didn't train the model for sad.

The evaluation is based on an objective point of reference. Suppose we have s_1 emphasized samples and s_2 un-emphasized ones. c_1 and c_2 is the correctly labeled count for them respectively. Then we define **Precision** P, **Recall** R, **F-Score** F and **Total Precision** T as:

$$P = \frac{c_1}{c_1 + (s_2 - c_2)} \times 100\% \quad R = \frac{c_1}{s_1} \times 100\%$$
$$F = \frac{2PR}{P + R} \qquad T = \frac{c_1 + c_2}{s_1 + s_2} \quad . \tag{6}$$

Setting the resample weight as 4, we get the result listed in Table 2.

Table 2. Performance of automatic labeling (resample weight = 4)

Model	Precision	Recall	F-Score	Total Precision
CART 1-pass	66.67%	62.50%	64.52%	92.57%
CART 2-pass	73.33%	68.75%	70.97%	93.92%
ME 1-pass	71.43%	62.50%	66.67%	93.24%
ME 2-pass	84.62%	68.75%	75.86%	95.27%

When the resample weight varies from 1 to 10, we take ME 2-pass model as example and show its evaluation result in Fig. 6.
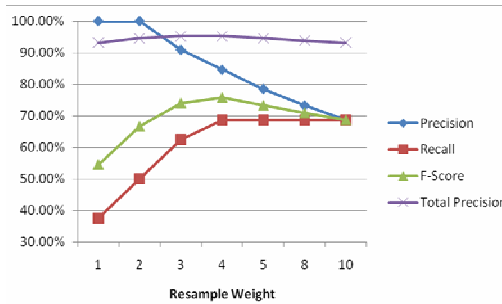
Fig. 6. Objective evaluation of ME 2-pass model with different resample weight

In general, the performance was acceptable, but the recall of emphasized samples was always lower than precision. We tried to explain this by going through the prosody generation error again and found that some emphasized samples were already synthesized quite well without emphasis label. This probably because that all the samples with identical contextual features in the corpus are emphasized, which results in a small generation error for them. That is also why the 2-pass model also preferred to improve the precision result. Max Entropy model acted better than CART in our application. Perhaps the data is still insufficient for covering all these features for CART to find a good split.

From Fig. 6, we can find that the resample weight has a big influence over the performance. As the weight increasing, the recall and F-score are improved but more and more un-emphasized samples will be misclassified. Considering the final performance, it is the best choice to choose 4 as our weight. This value is related to the count ratio of emphasized samples over un-emphasized ones, but the detail is not carefully studied yet.

5   Conclusion

In this paper, we stated the problem of labeling emotional emphasis for a corpus that already has phonetic and prosodic annotation. The prosody generation error from a statistical synthesizer trained from this corpus is examined and features are hence extracted to build up an automatic labeling model using CART and ME. During model training, samples are re-sampled to help fix the data ratio problem of emphasized sample over un-emphasized ones. And a two-pass procedure is also proposed for improving the performance. The final result is proved to be greatly enhanced with above efforts.

Further work involves preparing more available recordings and high-level syntactic information such as Part-of-Speech tagging can be introduced to achieve a better prediction result.

References

1. Erickson, D., Fujimura, O., Pardo, B.: Articulatory Correlates of Prosodic Control: Emotion and Emphasis. J. Language and Speech 41(3-4), 399–417 (1998)
2. Dioubina, O.I.: Annotation of Expressive Speech. In: ISCA Workshop on Voice Quality, Geneva, Switzerland, August 27–29 (2003)
3. Lee, S., Yildirim, S., Kazemzadeh, A., Narayanan, S.: An Articulatory Study of Emotional Speech Production. In: INTERSPEECH 2005, Lisbon, Portugal, September 4–8 (2005)
4. Kennedy, L.S., Ellis, D.P.W.: Pitch-based Emphasis Detection for Characterization of Meeting Recordings, Automatic Speech Recognition and Understanding. In: IEEE Workshop on Automatic Speech Recognition and Understanding, Saint Thomas, US Virgin Islands, November 30–December 3 (2003)
5. Raux, A., Black, A.W.: A Unit Selection Approach to F0 Modeling and Its Application to Emphasis. In: IEEE Workshop on Automatic Speech Recognition and Understanding, Saint Thomas, US Virgin Islands, November 30–December 3 (2003)
6. Nakajima, S., Hamada, H.: Automatic Generation of Synthesis Units based on Context Oriented Clustering. In: International Conference on Speech and Signal Processing 1988, New York, USA (April 1988)
7. Tokuda, K., Yoshimura, T., Masuko, T., Kobayashi, T., Kitamura, T.: Speech Parameter Generation Algorithms for HMM-based Speech Synthesis. In: International Conference on Speech and Signal Processing 2000, Istanbul, Turkey (June 2000)
8. Zhang, X., Xu, J., Cai, L.: Prosodic Boundary Prediction Based on Maximum Entropy Model with Error-Driven Modification. In: Huo, Q., Ma, B., Chng, E.-S., Li, H. (eds.) ISCSLP 2006. LNCS (LNAI), vol. 4274, pp. 149–160. Springer, Heidelberg (2006)
9. Xu, Y., Wang, Q.E.: Pitch Targets and Their Realization: Evidence from Mandarin Chinese. J. Speech Communication 33(4), 319–337 (2001)

A Method of Image Feature Extraction Using Wavelet Transforms

Minrong Zhao[1], Qiao Chai[1], and Shanwen Zhang[2]

[1] The Science Institute, Air-Force Engineering University, Sanyuan, 713800, P.R. China
[2] Faculty of Science, Zhongyuan University of Technology, Zhengzhou, 450007, P.R. China
zhaominrong1965@sina.com, wjdw716@163.com

Abstract. Image feature extraction is crucial in image target recognition. This paper presents a method of image feature extraction by combining wavelet decomposition. The image is first decomposed by wavelet transforms, and the decomposed coefficients are reconstructed to form a new time series, from which some energy vector can be extracted by time-frequency domain analysis. By calculating correlation coefficients, it is possible to recognize whether target signal is involved or not in gained image. The effectiveness of the method is verified by a real image with additive simulated noise signal, especially under the condition of low SNR.

Keywords: Wavelet Decomposition; Target Detection; Image; Feature Extraction.

1 Introduction

Image recognition is a difficult yet crucial technology in the domain of image management and computer vision, with the image feature extraction as its key point. As far as the task of target recognition is concerned, the quality of the extracted image feature, to a certain degree, not only affects the result of image recognition, but also has considerable effect on both the following examination and the quantity of recognition operation. The most commonly-used tool in feature extraction is time-frequency transform. With the characteristics of multi-differentiation and the capacity representing signal's regional feature in the realm of time-frequency, wavelet shift, a method of time-frequency regional analysis which can shift freely in both time and frequency, has been widely used in image feature extraction. Among them, Gabor wavelet transforms perfectly illustrates the sense perception of visual nerve cells biologically. Through the regulation of the sampled image feature in space and frequency according to particular requirements, we can obtain the features of the interested targets. Therefore, the research of Gabor filter and its application in image classification and computer vision has drawn an extensive attention. But, as for frequency domain transform, whether or not the target object's geometric characteristics can be properly pictured in the frequency range available is still a crucial yet hard problem to deal with at all times. However, geometric features of the target are the most visible and most powerful eigenvector used to illustrate the target objects.

Many methods have been proposed to represent contours. The chain-code and the improved chain-code approximate a contour with a sequence of directional vectors.

The B-spline descriptors describe a contour using a set of piece low-order polynomials and the Fourier descriptor describes a contour with coefficients via Fourier analysis. In recent years, the wavelet transform became an active area in multi-resolution signal and image analysis. The main idea of Chuang and Kuo is that a coordinate-based planar curve is firstly converted to wavelet coefficient domain using wavelet transform, and then wavelet coefficients of the curve are normalized using a procedure to be invariant to translation, rotation, and scaling so that curve comparison can be performed by comparing normalized wavelet coefficients. It is successfully used in many applications such as shape recognition and contour-based matching. In this paper we adopt this wavelet transforms to extract feature from image, in that:

(1)   Feature derived from wavelet transforms is very compact and can be normalized to be invariant to translation, rotation, and scaling.
(2)   Contours can be matched from the coarsest scale to finer one so that noise up to some levels can be handled and high computation speed can be achieved.

The rest of the paper is organized as follows. Section 2 introduces wavelet transforms. Section 3 proposes a method of image feature extraction. Experiments are illustrated in Section 4. Section 5 gives a brief summary of this paper.

2   Wavelet Transforms

A brief review of this wavelet-based curve descriptor is as follows: we use $\phi(t)$ and $\varphi(t)$ to denote a scaling function and corresponding wavelet function and both of them satisfy the dilation equations, with $\phi_n^m(t)$ and $\varphi_n^m(t)$ being their dilations and translations respectively.

$$
\begin{aligned}
\phi_n^m(t) &= 2^{-m/2}\phi(2^{-m/2}t - n), n \in Z, \\
\varphi_n^m(t) &= 2^{-m/2}\varphi(2^{-m/2}t - n), m \in Z, n \in Z
\end{aligned}
\tag{1}
$$

Let us denote a clockwise-oriented plane curve with parametric coordinates:

$$
C(t) = \begin{bmatrix} x(t) \\ y(t) \end{bmatrix}, t(l) = l/L, 0 \le l \le L
\tag{2}
$$

where t is normalized arc length. l and L are arc lengths along curve from a certain starting point t_0 and total arc length, respectively. By applying the wavelet transforms to the parameterized coordinates, we can obtain:

$$
\begin{bmatrix} x(t) \\ y(t) \end{bmatrix} = \begin{bmatrix} x_a^M(t) \\ y_a^M(t) \end{bmatrix} + \sum_{m=M-m0}^{M} \begin{bmatrix} x_d^m(t) \\ y_d^m(t) \end{bmatrix}
\tag{3}
$$

where

$$
x_a^M(t) = \sum_n a_n^M \phi_n^M(t), \quad y_a^M(t) = \sum_n c_n^M \phi_n^M(t)
\tag{4}
$$

are called the approximation signals at scale m and

$$x_d^m(t) = \sum_n r_n^m \varphi_n^m(t), \quad y_d^m(t) = \sum_n d_n^m \varphi_n^m(t) \qquad (5)$$

are called the detailed signals at scale m. Because noise is generally included in de-tailed signals, which are high frequency, we use the approximation coefficients a_n^M, c_n^m defined above as a planar curve descriptor for matching. Because of 2D rigid transformation (translation, rotation, and scaling) between the DEM and REM, the wavelet coefficients and approximation ones of the same contour may be differ-ent. However, in matching or recognition applications, it is known that the features selected as descriptors should be as insensitive as possible to the variation in 2D rigid transformation. Fortunately, it is well proved that normalized wavelet coefficients and normalized approximation ones are invariant to 2D rigid transformation. We can nor-malize a_n^M, c_n^m for individual curve to get normalized approximation coefficients. The normalization procedure can be referred to for detailed specification. We use a_n^M, c_n^m to measure the similarity of two contours and use vector (a_n^M, c_n^m) to get image feature, which can be used in target classification.

3   A Method of Image Feature Extraction

The steps of image feature extraction can be explained as follows:

Preprocessing: First, normalize the known infrared images $f(x_1, x_2)$; second, sub-tract image's average value (i.e. to generate zero value image) from the normalization. In this way, energy can be focused greatly, and accordingly we can enlarge the image prime numbers by the specific problems.

Multi-sized wavelet disassembly of the obtained images (if required, wavelet pack-age disassembly should be conducted). This can be done by extracting the respective signal features on the scale from low frequency to high frequency.

Reconstruct wavelet decomposition coefficients to extract signals of each fre-quency. We can use $s_0, s_1, ..., s_M$ to express decomposed reconstructed signals from low frequency coefficients and high frequency coefficients. (if needed, a_n^M, c_n^m can be dealt with by adopting the method of threshold value) Then signal s can be denoted as

$$s = s_0 + s_1 + ... + s_M \qquad (6)$$

On the assumption that the frequency of the target signal has been dealt with through normalization, say, the lowest frequency is 0, and the highest frequency is 1, then the frequency range $s_j (j = 0, 1, 2, ..., M)$ extracted can be denoted in the fol-lowing table.

Table 1. Frequency scope of decomposed signals

Signals	s_0	s_1	s_2	s_3
Frequency scope	0～0.125	0.125～0.25	0.25～0.5	0.5～1.0

Extraction of the total energy of all signal frequency scope. Suppose that E_j is the signal $s_0, s_1, ..., s_M$ corresponding energy, let's define, $s_0, s_1, ..., s_M$

$$E_j = \frac{1}{n} \sum_{k=1}^{n} \left[s_j(k) - \frac{1}{n} \sum_{k=1}^{n} s_j(k) \right]^2 \tag{7}$$

where $s_j(k)$, $(k = 1, 2, ..., n)$ denotes the n th scattering point of the reconstructed signals.

Construction of eigenvector E_j. By using energy E_j as the element, the eigenvector Z_z can be constructed as follows:

$$Z = (E_0, E_1, ..., E_M) \tag{8}$$

Image detection: Figure out the correlation coefficient d_k between unknown target eigenvector Z and eigenvector Z_k of the Kth target mode in the database.

$$d_k = \frac{1}{L} \cdot \left\{ \sum_{n} \sum_{i=0}^{N-|n|-1} \left| Z(i) * Z_k(i+n) \right| \right\} \bigg/ \left\{ \left[\sum_{i} |Z(i)|^2 \cdot \sum_{j} |Z_k(j)|^2 \right]^{\frac{1}{2}} \right\} \tag{9}$$

where L is the maximum value of vector Z and length Z_k. For all k, the corresponding template of the biggest value d_k is the recognized result of the target.

4 Experiment

To illustrate the superiority of this arithmetic in the field of eigenvector extraction, we undertake the experiment of image examination by adopting the above mentioned image classification arithmetic. The selected images are the most representative ones, including F-16、J-11、Su 27、transportion-8C、FC-1、annihination-10 and other target images, which are the extracted ones used as training samples. The following are the 6 silenced pictures. (SNR is 1.0 dB)

During the process of examination, we find that a single target is of as many as 30 images taken from the different angles respectively. Then we can identify whether the extracted images belong to the wanted aircraft, what type it is. This can be done only with the help of examining the extracted images of the randomly-sampled target samples, and then put the extracted numbers into the superior classification function.
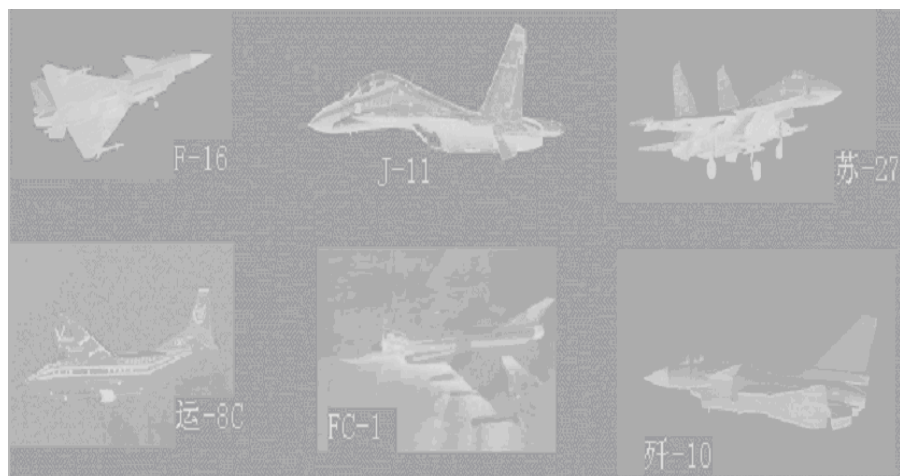
Fig. 1. The extracted images of six aircrafts
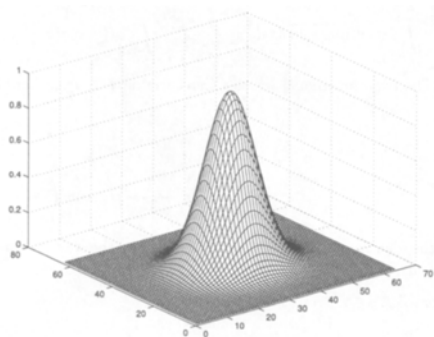
Fig. 2. Gaussian function
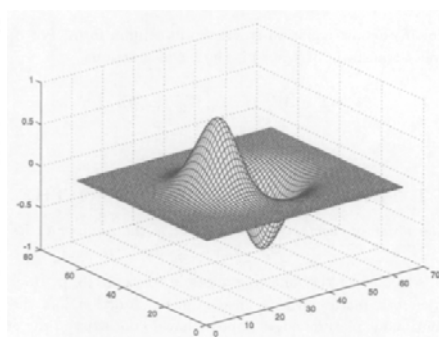
Fig. 3. Wavelet mother function

With the consideration of the affect of noises, we can add 0∼2 white noise of Gauss to the source /original images. From the result of the experiment, we find that within the threshold value and the noise level below or no more than 1.0, there exists no error, and the rate of recognition is up to 100% if we operate according to the above mentioned method. When the noise level is /equals 2.0, we can obtain a higher recognition rate, say 95%. Therefore, the discussed method is of high differentiation, perfect anti-noise capacity.

In many applications for image processing the smoothing function, selected $s(x, y)$ is the Gaussian function and it is illustrated in Figure2. The echo data of recognition target are denoted in radar echo with the combination of white noise of Gauss. The radar target 1-D range profile, wavelet decomposition and maximum-coefficients are showed in Fig3.

5   Conclusion

This paper applies the idea of wavelet transform to image recognition and classification. By using the wavelet transform, the source images can be decomposed into a sequence of different spatial resolution images. Image features' extraction is mainly based on wavelet coefficients. Lots of experiments on this aspect have been conducted and the results show that the proposed method is quite effective.

References

1. Mallat, S.: Multifrequency channel decomposition of images and wavelet models. IEEE Trans.on ASSP, ASSP 37(12), 2091–2110 (1989)
2. Daubechies: The wavelet transforms, time-frequency location and signal analysis. IEEE Trans. on IT 36, 961–1005 (1990)
3. Cetin, E., Ansari, R.: Signal Recovery from Wavelet Transform Maxima. IEEE Trans. on signal processing 42(1), 194–196 (1994)
4. Yinggang, D., Jinhui, L.: Target identification based on the optimal base number. In: Proceedings of ICSP 1998, pp. 271–275 (1998)
5. Heng, Y.J., Jing, C.: Robust Filter Algorithm Based on Wavelet. Journal of Electronics and Information Technology 24(3), 413–417 (2002)
6. Moghaddamzadeh, A., Bourbakis, N.: A fuzzy region growing approach for segmentation of color images. Pattern Recognition 30(6), 867–881 (1997)
7. Zugaj, D., Lattuati, V.: A new approach of color images segmentation based on fusing region and edge segmentations outputs. Pattern Recognition 31(2), 105–113 (1998)
8. Maeda, J., et al.: Segmentation of natural images using anisotropic diffusion and linking of boundary edges. Pattern Recognition 3(12), 1993–1999 (1998)
9. Lim, J.S.: Two-Dimensional Signal and Image Processing. Prentice Hall, New Jersey (1990)
10. Tang, Y.Y., Yang, L.H., Liu, J., Ma, H.: Wavelet Theory and its Application to Pattern Recognition. World Scientific, New Jersey (2000)

Study on Minimum Zone Evaluation of Flatness Errors Based on a Hybrid Chaos Optimization Algorithm

Ke Zhang

School of Mechanical and Automation Engineering,
Shanghai Institute of Technology 120 Caobao Road, 200235 Shanghai, China
zkwy@hotmail.com

Abstract. In this paper, according to characteristics of flatness error evaluation, a hybrid evaluation method to evaluate the minimum zone error is provided. The evolutional optimum model and the calculation process are introduced in detail. The hybrid optimization algorithm is based upon chaos optimization algorithm (COA) and Powell search. Compared with conventional optimum methods such as simplex search and Powell method, it can find the global optimal solution, and the precision of calculating result is very good. Moreover, the efficiency of COA is much higher than some stochastic algorithms such as simulated anneal algorithm and genetic algorithm (GA) when COA is used to a kind of continuous problems. The hybrid optimization algorithm can improve the efficiency of searching in the whole field by gradually shrinking the area of optimization variable. Finally, the control experiment results evaluated by different method such as the least square, simplex search, Powell optimum methods and GA, indicate that the proposed method does provide better accuracy on flatness error evaluation, and it has fast convergent speed as well as using computer expediently and popularizing application easily.

Keywords: Flatness, Evaluation, Chaos Optimization Algorithm, Minimum Zone.

1 Introduction

The minimum zone deviation, as defined in the ISO/R1101 standard, is generally accepted for specifying the form errors of geometric features [1,2]. However, no specific methods are recommended for finding minimum zone deviation in ISO/R1101. At present, the evaluation methods of form and position error are the least square, minimum zone method and so on. Although the least-squares method, because of its simplicity in computation and uniqueness of the solution provided, is most widely used in industry for determining form and position error, it provides only an approximate solution that does not guarantee the minimum zone value [3]. The results of minimum zone method not only verge on ideal error value, but also accord with ISO standard. Therefore, much research has been devoted to finding the minimum zone solutions for flatness error and other form errors using a variety of methods. Some

researchers applied the numerical methods of linear programming [3-5], such as the Monte Carlo method, the simplex search, spiral search, and the minimax approximation algorithm etc. Another approach has been to find the enclosing polygon for the minimum zone solution, such as the eigen-polyhedral method, the convex polygon method, and the convex hall theory etc. For example, computational geometry based techniques were applied to form error evaluation, and have shown to be efficient tools for calculating the minimum zone tolerance of various geometric features [6-9]. However, this algorithm also does not guarantee the minimum zone solution since only the facet-vertex antipodal pairs are taken into account.

The optimization algorithms are commonly used to approach the minima of flatness error objective function through iteration when a microcomputer is applied to assess flatness errors by minimum zone method. The essential prerequisite for convergence of any optimization algorithm is that the objective function to be solved has only one minimum in its definition domain, that is, it is a single valley one. If an objective function has more local minima in its definition domain, it's solution searched for by an optimization algorithms may not be its global minimum which is the wanted flatness error. Therefore, the mathematical models and algorithms for flatness error evaluation may be influenced in their solutions' reliability and practical values.

The traditional optimization methods are employed to refine the least-squares solution further. However, these traditional optimization methods have drawbacks in finding the global optimal solution, because it is so easy for these traditional methods to trap in local minimum points [10].

Chaos Optimization Algorithm (COA) is a stochastic optimization technique. By the use of the properties of ergodicity, stochastic property, and "regularity" of chaos variables, COA is prone to leap over local minimum points, and that continuity and differentiability of objective function are not required. The efficiency of COA is much higher than some stochastic algorithms such as simulant anneal algorithm (SAA) and genetic algorithm（GA）when COA is used to a kind of continuous problems. The chaos optimization algorithm can improve the efficiency of searching in the whole field by gradually shrinking the area of optimization variable. But when the search domain is larger or the local minimum is nearly same as the global, a longer computation time is required by this kind of approach in order to reach the global minimum. Considering advantages of Powell optimization method, such as the smaller calculation load, the faster optimization speed, the combination of COA and Powell optimization may greatly improve efficiency of algorithm.

In this paper, according to characteristics of flatness error evaluation and the minimum zone criterion, a hybrid algorithm by integrating COA and Powell search is presented to evaluate the minimum zone error of flatness error. The evolutional optimum model and the calculation process are introduced in detail. An example is given to test its validity herein.

The rest of this paper is organized as follows: The flatness error analysis is presented in Section 2. The hybrid optimization algorithm based upon chaos optimization algorithm (COA) and Powell search is given in section 3. An example is given to test its validity herein. Simulation results and the discussion of the results are presented in section 4. Finally, conclusions are given in Section 5.

2   Flatness Analysis

ISO Standards recommends a minimum zone evaluation of form and specifies that the ideal/reference features must be established from the actual measurement data such that the deviation between it and the actual feature concerned will be the least possible value.

According to ISO standard, the minimum zone flatness is defined by the minimum distance between two parallel planes that contain the real profiles, the two parallel planes are shown in Fig.1. Here, such two parallel planes can be represented by equation (1).

$$\begin{cases} z = Ax + By + C_1 \\ z = Ax + By + C_2 \end{cases},\tag{1}$$

where A, B, C_1 and C_2 are coefficients. If x-, y- and z-coordinates are obtained, both C_1 and C_2 become a function of A and B. define the measured data (x_i, y_j, z_{ij}), where $i = 1, 2, \ldots, n$; $j = 1, 2, \ldots, m$; n, m are number of data. Therefore, the distance t between such planes is

$$\begin{aligned} t &= \frac{|C_1 - C_2|}{(1 + A^2 + B^2)^{1/2}} \\ &= \frac{\max_{i,j}(z_{ij} - Ax_i - By_j) - \min_{i,j}(z_{ij} - Ax_i - By_j)}{(1 + A^2 + B^2)^{1/2}}, \end{aligned}\tag{2}$$

which is a function of A and B. The objective function, $f(A, B)$ in this problem is defined as follows:

$$f = \frac{\max_{i,j}(z_{ij} - Ax_i - By_j) - \min_{i,j}(z_{ij} - Ax_i - By_j)}{(1 + A^2 + B^2)^{1/2}}.\tag{3}$$
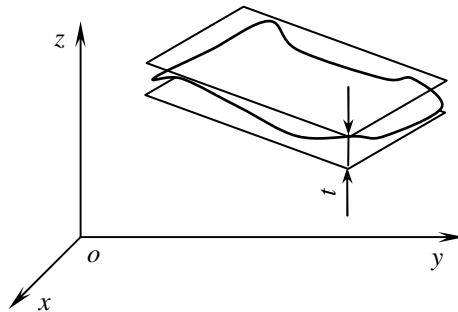
Fig. 1. Flatness by minimum zone method

Therefore, the minimum zone flatness is expressed as follows:

$$\min f(A,B). \tag{4}$$

Without losing the generality, we define optimized problems as the minimization problems. Using a hybrid optimization algorithm based on COA to evaluate flatness errors can be summarized as finding the set of variables (A,B) so that the objective function $f(A,B)$ is the minimum.

3 Hybrid Algorithm

3.1 Chaos Optimization Algorithm

We choose Logistic mapping formula [11] to generate chaos variables as follows:

$$\gamma(k+1) = \lambda\gamma(k)(1-\gamma(k)), \tag{5}$$

where λ is control parameter, $0 \leq \gamma(0) \leq 1$. This map is often used to model population dynamics. The parameter λ is the nonlinearity parameter; when $\lambda = 3.79$ the mapping is chaotic [12]. The logistic map is a one dimensional map. By the use of sensitivity of chaos motion for initial value, n chaos variables can be obtained by endowing n tiny difference initial values with (5).

Let optimization problem be minimum in the form as follows:

$$\begin{cases} min\, f(X) \\ s.t. a_i \leq x_i \leq b_i \end{cases}, \tag{6}$$

where $X = \{x_1, x_2, \ldots, x_n\}$, x_i are optimization variables, a_i and b_i are respectively upper limit and lower limit of x_i.

The procedure of COA is as follows:

Step 1: Choose maximum movement degree M of chaos variables and search number S.

Step 2: Generate $S \times n$ chaos variables $\gamma_j^i(0)$, $0 \leq \gamma_j^i(0) \leq 1$ (i=1,2,…, n; j=1,2,…, S).

Step 3: Let chaos variables $\gamma_j^i(0)$ map into the range of optimization variables x_i:

$$x_j^i(0) = a_i + \gamma_j^i(0)(b_i - a_i), \tag{7}$$

where $(i = 1,2,\ldots,n; j = 1,2,\ldots,S)$.

Set $X_j^* = X_j(0)$ and $f_j^* = f(X_j(0))$, where X_j^* is optimal solution, f_j^* is optimal function value; $k = 0$, then

$$\gamma_j^i(k+1) = 4\gamma_j^i(k)(1-\gamma_j^i(k)), \tag{8}$$

$$x_j^i(k+1) = a_i + \gamma_j^i(k+1)(b_i - a_i) , \tag{9}$$

where $i = 1,2,\ldots,n; \; j = 1,2,\ldots,S$.

Step 4: When $k \le M$, if $f(X_j(k+1)) < f_j^*$, then, $X_j^* = X_j(k+1)$, $f_j^* = f(X_j(k+1))$; if $f(X_j(k+1) \ge f_j^*$, X_j^* and f_j^* hold the line, increase $k = k+1$, and go to Step 3. When $k > M$, output X_j^* and f_j^* , and obtain $f_{min}^* = \min(f_j^*)$, X_{min}^* is the solution of f_{min}^* . If optimal function values meet optimization convergence criterion

$$\left\{ \frac{1}{S} \sum_{j=1}^{S} \left[f_j^* - f_{avg} \right]^2 \right\}^{1/2} < \varepsilon , \tag{10}$$

where $f_{avg} = \dfrac{1}{S} \sum_{j=1}^{S} f_j^*$, then output the global optimization solution X_{min}^* and the optimal function values f_{min}^* ; otherwise $a_i = \min(x_j^i)$, $b_j = \max(x_j^i)$, where $i = 1,2,\ldots,n$; $j = 1,2,\ldots,S$, and go to Step 2.
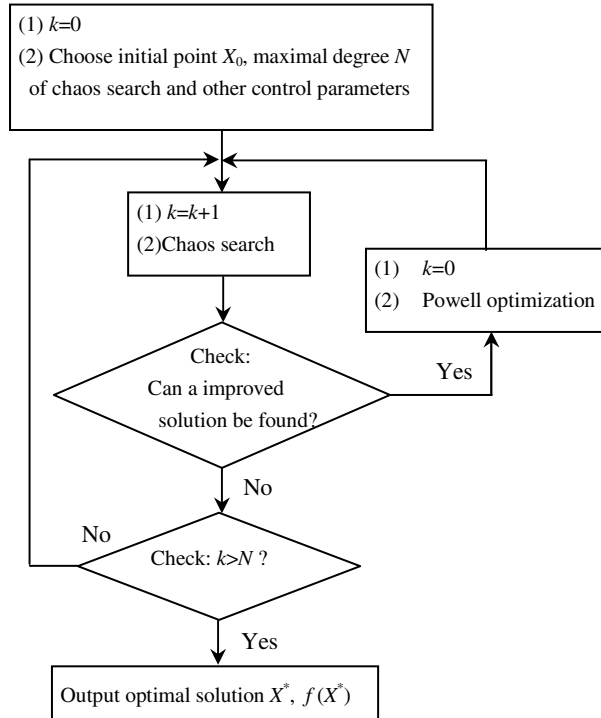
Fig. 2. Procedure of hybrid optimization algorithm

3.2 Powell Search

Powell method is an extension of the basic pattern search method. It is the most widely used direct search method and can be proved to be a method of conjugate directions. A conjugate directions method will minimize a quadratic function in a finite number of steps. Since a general nonlinear function can be approximated reasonably well by a quadratic function near its minimum, a conjugate directions method is expected to speed up the convergence of even general nonlinear objective function. The procedure of Powell search may be obtained from literature [10].

3.3 Hybrid Optimization Algorithm

In order to improve the performance of chaos optimization, we present a hybrid chaos optimization algorithm based on chaos optimization and Powell search. The procedure of hybrid optimization algorithm is shown in Fig. 2. As shown in Fig. 2, first, the search by using COA is prone to arrive near global optimization solution. Then, the local search using Powell search can be quickened, and satisfying result can be obtained.

4   Examples

One numerical example is given here to check the validity and efficiency of the proposed evaluation method. The measurement data from the plane surface are given in Table 1 [4].

Table 1. Data measured

x(cm)	y(cm)						
	0	10	20	30	40	50	60
0	10.71	8.81	6.31	4.01	1.91	0.80	0.00
10	4.30	1.90	-0.82	-3.17	-5.24	-6.65	-7.38
20	-2.69	-5.00	-7.77	-10.22	-12.29	-13.93	-14.66
30	-0.67	-11.01	-14.03	-16.65	-18.53	-20.56	-21.10
40	-13.63	-16.19	-19.10	-21.86	-24.09	-26.03	-26.53
50	-18.26	-20.70	-23.86	-26.55	-28.91	-31.11	-31.42
60	-22.04	-24.50	-27.90	-30.64	-33.22	-35.53	-35.87
70	-25.43	-28.11	-31.66	-34.52	-37.05	-39.16	-39.24
80	-28.58	-30.99	-34.67	-37.59	-39.99	-41.96	-42.08
90	-31.71	-33.68	-37.13	-39.62	-41.96	-43.41	-43.93

Table 2. Flatness results

Calculation method	A (μm/cm)	B (μm/cm)	Flatness error (μm)
Least-square	-0.483	-0.223	9.214
Simplex search	-0.486	-0.216	9.134
Powell search	-0.487	-0.212	9.132
Literature [4]	-0.482	-0.196	8.76
GA	-0.461	-0.494	8.317
Hybrid algorithm	-0.5142	-0.467	8.231
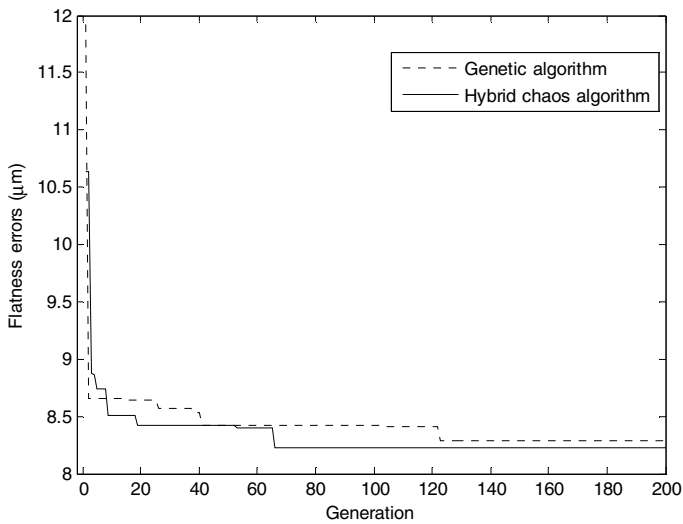
Fig. 3. The evaluation of flatness error

The procedures were programmed in the MATLAB programming language. Fig. 3 shows the searching processes of flatness errors. As seen in the figure, it takes about 140 iterations to find the minimum zone solution of the flatness error. The calculation result of flatness using the hybrid algorithm is f=8.231 μm. The results of flatness evaluation from different methods are provided in Table 2. As shown in the table, the comparison shows that the global optimum solution of flatness evaluation problem using the proposed procedure can be obtained and accord with the fact of measured profile. Figs. 3 shows the optimizing processes of the method for data sets (for clearer comparison, the optimizing processes for GA is showed in the figure). As seen in the figure, it takes about 70 iterations to find the optimal solution in using hybrid optimization algorithm to evaluate flatness error. This further illustrates effectiveness of the proposed method.

In order to verify the flexibility of the proposed method, many measured data of flatness error evaluation available in the literature are selected and all the results are found to be more accurate than those obtained by the LSM and conventional optimization methods.

5 Conclusions

The flatness evaluation problem was formulated as unconstraint optimization problems. In this paper, a hybrid optimization approach based on chaos optimization and Powell search to evaluate flatness error was presented. The optimization solution procedure was developed to solve the optimization problem. The techniques were compared to some existing techniques. It is shown through example that the procedure of this paper provides exact values of flatness error. The result also shows that the proposed procedure converges to the global optimum more rapidly than conventional methods. The evaluation method is the same with the others form and position error evaluation.

Acknowledgements

References

1. Wang, M., Cheraghi, H., Masud, A.S.: Circularity Error Evaluation Theory and Algorithm. Precision Engineering 23(2), 164–176 (1999)
2. Cheraghi, S.H., Jiang, G.H., Ahmad, J.S.: Evaluating the Geometric Characteristics of Cylindrical Features. Precision Engineering 27(2), 195–204 (2003)
3. Kanad, T., Suzuki, S.: Evaluation of Minimum Zone Flatness by Means of Nonlinear Optimization Techniques and Its Verification. Precision Engineering 15(2), 93–99 (1993)
4. Huang, S.T., Fan, K.C., Wu, J.H.: A New Minimum Zone Method for Evaluating Flatness Error. Precision Engineering 15(1), 25–32 (1993)
5. Cheraghi, S.H., Lim, H.S., Motavalli, S.: Straightness and Flatness Tolerance Evaluation: an Optimization Approach. Precision Engineering 18(1), 30–37 (1996)
6. Lee, M.K.: A New Convex-hull Based Approach to Evaluating Flatness Tolerance. Computer-Aided Design 29(12), 861–868 (1997)
7. Samuel, G.L., Shunmugan, M.S.: Evaluation of Straightness and Flatness Error Using Computational Geometric Techniques. Computer-Aided Design 31(13), 829–843 (1999)
8. Zhu, X.Y., Ding, H.: Flatness Tolerance Evaluation: an Approximate Minimum Zone Solution. Computer-Aided Design 34(9), 655–664 (2002)
9. Huang, J.P.: An Efficient Approach for Solving the Straightness and the Flatness Problems at Large Number of Data Points. Computer-Aided Design 35(1), 15–25 (2003)
10. Singiresu, S.R.: Engineering Optimization. John Wiley & Sons, New York (1996)
11. Arun, V.H.: Chaos. Manchester University Press, Manchester (1986)
12. Otawara, K., Jordan, M.I.: Controlling Chaos with an Artificial Neural Network. In: IEEE International Joint Conference on Fuzzy System, pp. 1943–1948. IEEE Press, Los Alamitos (1995)

Type-2 Fuzzy Sets Applied to Pattern Matching for the Classification of Cries of Infants under Neurological Risk

Karen Santiago-Sánchez, Carlos A. Reyes-García, and Pilar Gómez-Gil

National Institute of Astrophysics, Optics and Electronics, Computer Science Department,
Luis Enrique Erro No. 1, Tonantzintla, México 72840
ksantiago@ccc.inaoep.mx, kargaxxi@inaoep.mx, pgomez@acm.org

Abstract. Crying is an acoustic event that contains information about the functioning of the central nervous system, and the analysis of the infant´s crying can be a support in the distinguishing diagnosis in cases like asphyxia and hyperbilirrubinemia. The classification of baby cry has been intended by the use of different types of neural networks and other recognition approaches. In this work we present a pattern classification algorithm based on fuzzy logic Type 2 with which the classification of infant cry is realized. Experiments as well as results are also shown.

Keywords: Fuzzy Sets, Pattern Matching, Neurological Risk.

1   Introduction

The interest to analyze the infants' crying with the intention of helping in the diagnosis of pathologies has increased in the last years. At present computer models are used to automate the pathologies identification through infant cry. In [Cano 2007] it is established that the crying is an acoustic event that contains information about the functioning of the central nervous system, and that the analysis of the infant's crying can be a support in the distinguishing diagnosis in cases like asphyxia and hyperbilirrubinemia. Also there, Cano reports the results obtained in the classification of these two pathologies, for whose processing the authors used a method based on the Kohonen´s Self-organizing maps (SOM) obtain a precision of 85 %. Orozco-García [Orozco 2002] classified the crying of normal and deaf infants, reporting results of 98.5 % of precision for the classification of pathologic and normal cry, using a feedforward neural network trained with the method of conjugated scaled gradient. Orozco used four types of features, namely; *Linear Prediction Coefficients* (LPC), *Mel Frequency Cepstral Coefficients* (MFCC), Fundamental Frequency and Intensity. Amaro (2008) proposes an automatic recognition system of the infants crying based on the combination of characteristics. Classifying two classes, normal and deaf, with a feed forward network neural, it achieved results of 98.66 % of precision with the combination of four characteristics that are: LPC, MFCC, Intensity and Cochleograms.

In [Verduzco 2004] it is explained how it is that pathologies that affect the nervous system and that can have neurological type sequels, produce changes in the crying of infants. In the case of the hyperbilirrubinemia crying is affected in the fundamental

frequency, it increases the percentage of phonation and increases the first formant variability (the highest intensity in a certain frequency). In crying of infants with asphyxia an increase in the fundamental frequency over the cases of crying of normal infants is detected. On the other hand, Jara (2006) did not find significant differences between the fundamental frequency of crying of infants asphyxiating and normal. Nevertheless, it documented high intensities in the signals of crying of infants with asphyxia; he also applied spectrophonographical analysis on the crying of infants presenting neurological risk pathologies. And he found that the fundamental frequency of cries affected by hyperbilirrubinemia increased up to 2120 Hz. with a minimum of 960 Hz. compared with a maximum of 670 Hz. and minimum of 390 Hz, in healthy infants' crying.

The classification of baby cry has been intended by the use of different types of neural networks and other recognition approaches. In this work we present a pattern classification algorithm based on fuzzy logic Type 2 with which the classification of infant cry is realized. In [Saïd, Sayed and Billaudel 2007] a fuzzy pattern matching method is described (FPM) and experiments with overlapped classes are made, reporting a precision average of 92.58 %. In [Chen, Wu and Yachida 1995] use the FPM method for the detection of the face of a person inside an image. A characteristic of this type of information is that they are also overlapped. After applying the FPM method, they conclude that the method is flexible and robust, since they managed to detect faces in different images based on the extraction of colored maps.

2   Type 2 Fuzzy Sets

In 1965 Lotfi Zadeh published by first time the theory were he presents fuzzy sets like those to which an element belongs to a certain degree. The membership degree of an element x to a fuzzy set A is determined by a membership function, this membership degrees are in the [0,1] interval [Zadeh 1965].

Formally a fuzzy set A is defined as a membership function to map the elements in a domain or universe of discourse X to elements in the [0,1] interval:

$$A: X \rightarrow [0,1] \tag{1}$$

The closer is $A(x)$ to 1, the larger will be the membership of object x to the set A. Then the possible membership values vary between 0 (absolutely no membership) and 1 (total membership). This kind of functions is known as of type 1. An example is shown in Figure 1 *(a)*.

Type 2 membership functions describe to type 2 fuzzy sets. To describe this kind of sets we have as a base the image the type 1 membership function image in Figure 1 *(a)*. Now, lets imagine that the membership function is distorted, if we move the vertex of the triangle to the right and to the left in an irregular way we will have what is shown in Figure 1 *(b)*. Then, for a specific value x' there is not a specific membership value u', instead there are several values u' (called primary membership). Nonetheless these values do not have the same importance, that is why weights are assigned to each of them (secondary membership), so that for each value of x, a tridimensional membership functions should be generated. Figure 2 shows an example of the membership functions for each value of x [Mendel y Bob 2002].
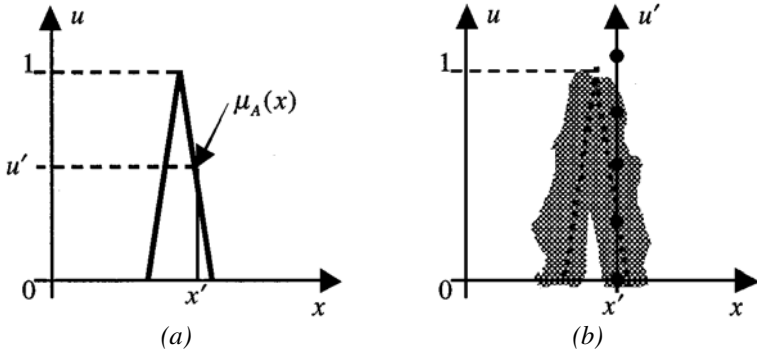
Fig. 1. (a) Type 1 membership function. (b) Shaded Type-2 membership function [Mendel & Bob 2002]

A type 2 fuzzy set, denoted by $\tilde{A}$, is characterized by a type 2 membership function $\mu_{\tilde{A}}(x,u)$ where $x \in X$ and $u' \in J_x$ which is in the [0,1] interval. A formal definition of a type 2 fuzzy set is given by:

$$\tilde{A} = \{((x,u), \mu_{\tilde{A}}(x,u)) | \forall x \in X, \forall u \in J_x \subseteq [0,1]\} \tag{2}$$

where J_x is the membership set of x and $0 \leq \mu_{\tilde{A}}(x,u) \leq 1$.

$\tilde{A}$ can be also expressed as:

$$\tilde{A} = \int_{x \in X} \int_{u \in J_X} \mu_{\tilde{A}}(x,u)/(x,u) \qquad J_x \subseteq [0,1] \tag{3}$$

where $\iint$ denotes the union on all the x and their corresponding calculated u'. Let us recall that the union of two sets A and B is by definition another set that contains the elements of A *or* B. If we look to each element of a type 2 fuzzy set as a subset, then the unions in equation (3) fulfills the definition of union, For each specific value of x and u thee is one union [Mendel y Bob 2002].

The example in Figure 2 describes $\mu_{\tilde{A}}(x,u)$. There, the values for x and u are discrete, in particular, X = {1,2,3,4,5} and U = {0,0.2,0.4,0.6,0.8}. For each value of x, like $x=x'$, the second plane axe represents the values of u and $\mu_{\tilde{A}}(x',u)$ is represented by the vertical part. A secondary membership function is the vertical part of $\mu_{\tilde{A}}(x,u)$. This is $\mu_{\tilde{A}}(x = x', u)$ for $x \in X$ and $\forall u \in J_{x'} \subseteq [0,1]$ it means:

$$\mu_{\tilde{A}}(x = x', u) \equiv \mu_{\tilde{A}}(x') = \int_{u \in J_{x'}} f_{x'}(u)/u \qquad J_{x'} \subseteq [0,1] \tag{4}$$

where $0 \leq f_{x'}(u) \leq 1$ and $\int$ represents the union operation [Mendel y Bob 2002].

Continuing with the example, the type 2 fuzzy membership function described in Figure 2, for $x=2$, has the following values:

$$\mu_{\tilde{A}}(2) = 0.5/0 + 0.35/0.2 + 0.35/0.4 + 0.2/0.6 + 0.5/0.8$$
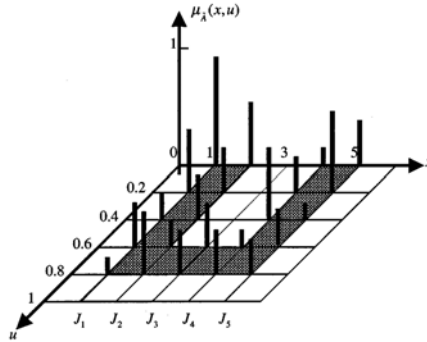
Fig. 2. Example of a type 2 membership function [Mendel & Bob 2002]

where symbol / is not a division, instead it is used to separate the primary membership from the secondary one: [primary membership]/[secondary membership] and the + is not an arithmetic sum but the union of memberships, each primary membership has a secondary membership assigned. Note that the calculation of each value of $\mu_{\tilde{A}}(x')$ is through primary and secondary membership functions.

3   Infant Cry Classification

The infant cry classification process begins with the acoustic processing of the crying signal to obtain the acoustic features in order to form the corresponding data matrixes of each class. Then, a fuzzy pattern matching training process on the classifier is executed. Once the classifier is trained it can be used to recognize unknown infant cry signals.

The samples of infant cry used were directly recorded and collected at the National Mexican Rehabilitation Institute *(Instituto Nacional de Rehabilitación-INR),* and at the Social Security Mexican Institute *(Instituto Mexicano del Seguro Social-IMSS)* in Puebla, Mexico. For this work we used samples from healthy infants (normal) and from babies with the pathologies of asphyxia and hyperbilirrubinemia. The collected files were processed by; cleaning the signals from silence segments, normalizing them to the frequency of 8,000 Hz. Then each file was divided in 1 second segments, after which we obtained 418 hyperbilirrubinemia samples, 340 of asphyxia, and 1136 normal samples.

The feature extraction was done with PRAAT [Boersma P. y Weenink D. 2007]. We used four different extraction methods for the features; Intensity, Cochleogram, LPC, and MFCC. For the extraction of the two last features we applied 50 ms windows (Hamming), in each of which we obtained 16 coefficients. In this way we obtained feature vector for the corresponding characteristic with 19 values for Intensity, 304 for LPC and MFCC and of 510 cochleogram values for each one second segment sample.

4   Fuzzy Pattern Matching with Type 2 Fuzzy Sets

Once the feature vectors of each class are obtained, we proceed to the infant cry recognition and classification phase. For this task we applied the Fuzzy Pattern Matching approach modified to the use of type 2 fuzzy sets (T2-FPM).

The algorithm is divided in two parts, the learning one, where primary and secondary membership information on the classes is collected, the membership of each element to each class is calculated, and a decision to which class each element belongs to is taken.

In the training phase, from the matrixes containing the feature data of each class (*MatrizEntrenamiento^c*), where $c \in (c_1, c_2, \ldots c_m)$ and m is the number of classes, a matrix called *HisT2$_c$* is obtained. This matrix is a histogram which concentrates the information of the training matrix for class c and which is obtained in the following way.

- The largest and the lowest values in the training matrixes, for the class the classifier is going to be trained, are determined.
- A vector b is calculated, whose first value is the lowest value and the last value is the largest one, and the intermediate values are thus obtained by calculating equal intervals between the largest and the lowest values.
- The number of intervals is experimentally established and represents the bars in the histogram.
- Further on the value the value d, which is the number of intervals in which each interval described in b will be divided, is calculated, and a matrix called *MatrizGuía*, which is a guide of the values in the training matrixes, is built.
- Then, for each attribute $x \in X=(x_1, x_2, \ldots x_n)$ found in the class c training matrix:
- Indexes ij are obtained, such that *MatrizGuía$_{ij}$* is the closer value to x and *HisT2$^c_{ij}$* is updated as follows:

$$HisT2^c_{ij} = HisT2^c_{ij} + 1 \tag{5}$$

- Later on the primary and secondary membership values corresponding to the attributes counted in each *HisT2$^c_{ij}$* is calculated in the following way:

$$MsP^c_{ij} = \frac{HisT2^c_{ij}}{HisT2^c_{max}} \tag{6}$$

where *MsP$^c_{ij}$* is the primary membership value of the element *HisT2$^c_{ij}$*, c is the class to which it belongs to and *HisT2$^c_{max}$* is the maximum value contained in the *HisT2* of class c. This primary membership value indicates the membership degree of an element to class c.

The secondary membership is calculated with:

$$MsS^c_{ij} = \frac{HisT2^c_{ij}}{\sum_{c=1}^{C} HisT2^c_{ij}} \tag{7}$$

where MsS^c_{ij} is the secondary membership value of element $HisT2^c_{ij}$, this value is divided by the value obtained from the sum of the values in the $HistT2_{ij}$ of each class. The secondary membership value indicates the importance of that element to class c, because it takes in count its membership to the rest of the classes.

In the classification phase an element $X=(x_1,x_2,...x_n)$ is received, from which the membership to each class is obtained; for doing that the indexes ij or obtained in a way that $MatrizGuía_{ij}$ is the closer value to x_k, where $x_k \in X$, and if $MsS^c_{ij}>P_M$ the membership degree is calculated with equation (8) else it is calculated with equation (9). This is so because depending on secondary membership value we may know if that element is important for the class, then if the secondary membership value is greater than P_M its final membership value is maximized, on the contrary it is maximized. P_M is a parameter which can be experimentally modified, and can take values in the [0,1] interval.

$$\pi^c(x_k) = MsP^c_{ij}+MsS^c_{ij} \tag{8}$$

$$\pi^c(x_k) = (MsP^c_{ij}+MsS^c_{ij}) / 2 \tag{9}$$

where $\pi(x^c_k)$ represents the membership of point x_i to class c, if the secondary membership obtained for point ij in class c is greater than 0.5 it means that that point is important for that class, so that primary and secondary membership values are added up, else the point is not so important for class c in which case the sum of the primary and secondary membership values are divided by two.

Further on the total membership of X to class c is obtained

$$\pi^c(X)=H[\pi(x^c_1), \pi(x^c_2) .... \pi(x^c_n)] \tag{10}$$

where $\pi^c(X)$ is the membership degree of simple X to class c and the function H is the average of the memberships obtained for each point x_k.

The element X is assigned to which it has obtained the greater membership. It is worth to mention that for the experiments performed in this work the following parameters were used: b=20 and d=10, which means that the $HisT2$ generated in each test were of size 20x10 as well as the matrixes of the primary and secondary memberships. Notice that regardless the samples size and of the attributes, each class is represented by matrixes of the same size, which makes the classification process to be a simple and fast algorithm.

5   Results

The classifier was tested using the method of 10-fold cross validation, which consists of dividing in 10 parts the testing set, and testing the classifier with each one. Nine subsets are used for training and one for testing. This process is repeated 10 times using a different test set each time.

5.1   Evaluation Metrics

Recall and precision metrics were used to evaluate the performance of the classifier. Precision metric gives the percentage of elements correctly classified, and it is defined as:

$$precision_c = \left(\frac{E_c}{E_t}\right) * 100 \tag{11}$$

where E_c is the total number of elements classified as class c, and E_t is the total number of elements belonging to class c.

Recall with respect to class c is defined as the probability that an instance x, which belongs to class c, could be correctly classified [Montes and Villaseñor 2007]. Recall is calculated as:

$$recall_c = \left(\frac{E_c}{E_{tc}}\right) * 100 \tag{12}$$

E_c is the total of elements classified as class c and E_{tc} is the total of elements actually classified as c, including those belonging to other classes.

5.2 Experiments

The dataset used to test the classifier contains: 400 patterns for class "normal," 340 patterns for class "asphyxia" and 418 for class "hyperbilirubinemia". Each pattern contains four feature vectors: LPC (304 elements), MFCC (304 elements), intensity (19 elements) and cochleogram (510 elements). Different combinations of these vectors were used for testing the classifier, in order to find out the best features to discriminate among asphyxia and hyperbilirubinemia. For example, for the test case using the four feature vectors, the classifier receives as input 1,137 attributes coming from 340 patterns.

The most interesting results are shown at table 1. Notice that the combination of LPC and Cochleogram vectors gets the best classification results, even better that the experiment using the four vectors. Other interesting point is that, when only one vector was used as feature vector, the best result was obtained by LPC, and the worst by intensity. A reason for this behavior may be that intensity is the smallest feature vector, holding only 19 attributes for each segment of the signal. In the other hand, cochleogram vector has the biggest number of attributes. However, LPC vector performed better than cochleogram when discriminating among asphyxia and hyperbilirubinemia.

Table 1. Experiments using three classes: asphyxia, normal and hyperbilirubinemia

Feature(s)	Mean	
	Precision	Recall
Cochleograms	85,00%	85,73%
Intensity	61,31%	57,63%
LPC	**89.15**	**92.70**
MFCC	78,73%	85,60%
LPC-Cochleograms	**91.74%**	**92.53%**
LPC-Cochleograms-MFCC	84,67%	82,83%
LPC-MFCC-Cochleograms- Intensity	**89,85%**	**93,92%**

Table 2. Experiments using two classes: normal and hyperbilirubinemia

Feature(s)	Mean	
	Precision	**Recall**
Intensity	75,32%	83,32%
LPC	**86,92%**	**86,12%**
MFCC	**90,88%**	**91,57%**
Cochleograms	73,86%	81,30%
LPC-MFCC	**95,56%**	**95,77%**
LPC-Intensity	72,51%	80,57%
LPC-Cochleograms	86,79%	89,67%
MFCC-Cochleograms	80,93%	84,37%

Some experiments discriminating among two classes were performed in order to figure out which features best describe hyperbilirubinemia. Each feature vector was used by itself to classify, and also combinations of two feature vectors were tested. Results are shown at table 2. As in the case of three classes, best results were obtained using LPC. Second best was obtained with MFCC. When two feature vectors were combined, LPC-MFCC got the best results, getting 95.56% precision. This may indicate that LPC is the best characteristic vector for this case. For the same case, a 95.77% of recall was obtained.

For three-class classification, the best results were obtained using LPC-Cochleograms. Acceptable results were also obtained when all four feature vectors were used (LPC, MFCC, Cochleograms and intensity). The fact that two feature vectors perform better than four may be explained because intensity proved to be bad discriminator when used by itself (see table 1).

6   Conclusions

We developed a cry-signal classifier based on pattern matching and type-II fuzzy logic, able to discriminate normal cries from cries when asphyxia or hyperbilirubinemia is present. The classifier was tested using different combinations of feature vectors as well as different combinations of the three available types of cries. Best results were obtained when LPC and MFCC feature vectors were used to discriminate two classes. For the case were the classifier worked to discriminate three classes, best results were obtained using feature vectors LPC and cochleograms. For the case were the classifier discriminated among normal and hyperbilirubinemia, the best results were obtained using LPC and MFCC, giving the insight that these are the best feature vectors to characterize hyperbilirubinemia.

Pattern matching combined with type-II fuzzy logic prove to be a versatile and easy-to-implement method. The characteristics of the method allow a clean representation of each class, and it makes easy to analyze the importance of each feature for each class. Currently, our research lab is working analyzing other

features and other kinds of classifiers, in order to get a better classification for this type of signals. Also, some work in the use of dynamics invariant metrics is in progress.

References

1. Amaro Camargo, E.: Caracterización Automática del Llanto de Bebé para su Estudio con Modelos de Clasificación (Automatic Characterization of cry signals in the study of classification models). Master thesis in computer science. National Institute of Astrophysics, Optics and Electronics (2008)
2. Barkin, R.: Urgencias Pediátricas. Pediatric emergencies, ch. 11, pp. 107–115. Elsevier, Spain (2000)
3. Boersma, P., Weenink, D.: PRAAT: Doing phonetics by computer (Ver.5.0.1). Software (January 2007), http://fonsg3.hum.uva.nl/praat/
4. Cano, S.D., et al.: 15 Años del Cry Analysis en Cuba: Resultados y Perspectivas (Fifth teen years of cry analysis at Cuba: Results and perspectives). In: VI Congreso Internacional de Informática en Salud (Fourth international congress on informatics in Health). Santiago de Cuba (2007)
5. Chen, Q., Wu, H.Y., Yachida, M.: Face Detection by Fuzzy Pattern Matching Department of Systems Engineering, Osaka University, IEEE 0-8186-7042-8/95, pp. 591–596 (1995)
6. Chung, F.: Uncertain Fuzzy Clustering: Insights and Recommendations, Hanyang University, Korea. IEEE Computational Intelligence Magazine (February 2007)
7. Dworkin, P.H.: Pediatria- Serie médica nacional para el studio independiente (National Medical Series in Pediatrics for independent study), 4th edn., pp. 151-157 y 170-173. McGraw Hill Interamericana, New York (1995)
8. Golub, H., Corwin, M.: Infant Cry: a Clue to Diagnostic. Pediatrics 69, 197–201 (1982)
9. Hernández, J.N.: Análisis Espectrofonográfico del llanto en recién nacidos de término con riesgo neurológico (Spectrum and phonetic analysis of cry signals in newborns with neurological risk. Thesis in the specialty of Medical Pediatrics. Department of Neonatology. National Institute of Perinatology, México (2006)
10. Mendel, J.M.: Type-2 Fuzzy Sets and Systems: an overview, University of Southern California, USA
11. Mendel, J.M., Bob John, R.I.: Type-2 Fuzzy Sets Made Simple. IEEE Transactions on Fuzzy Systems 10(2) (2002)
12. Montes, M., Villaseñor, L.: Notas del curso de clasificación de texto (Text classify-cation. Course notes), Speech Tech-nologies Lab. National Institute of Astrophysics, Optics and Electronics (2007)
13. Orozco García, J.: Clasificación de Llanto de Bebé utilizando una Red Neural de Gradiente Conjugado Escalado (Cry signals classification using a conjugate-gradient Neural Net) Master thesis in Computer Science. National Institute of Astrophysics, Optics and Electronics. Puebla, Mex (2002)
14. Saïd Bouguelid, M., Sayed Mouchaweh, M., Billaudel, P.: New Results on Diagnosis by Fuzzy Pattern Recognition, CReSTIC, Université de Reims-Champagne-Ardennes Moulin de la Housse, France (2007)

15. Epidemiología, Sistema único de información, Mortalidad Infantil, Perfil Epidemiológico de las defunciones por Asfixia del Recién nacido (Epidemiology, a unique information system, infant mortality and epidemic profile of death by asphyxia in the neonate) National System of Epidemic watching Num. 45, 23, week 45, November 5-11 (2006)
16. Verduzco, A., et al.: Características del llanto de infantes de alto riesgo (Characteristics of cry signals for neonates with high risk). Ciencias Clínicas (Clinic Sciences) 5(1), 21–29 (2004)
17. Zadeh, L.A.: Fuzzy Sets, Department of Electrical Engineering and Electrics Research Laboratory, University of California, Berkeley, California (1965)
18. Tychtl, Z.k., Psutka, J.: Speech Production Based on the Mel-Frequency Cepstral Coeffi-cients, Department of Cybernetics, University of West Bohemia, Czech Republic (1999)

Real-Time Sound Synthesis of Plucked String Instruments Using a Data Parallel Architecture

Huynh Van Luong[1], Sangjin Cho[2], Jong Myon Kim[1,*], and Uipil Chong[1]

[1] School of Computer Engineering and Information Technology
huynhlvd@yahoo.com, {jmkim07,upchong}@ulsan.ac.kr
[2] School of Electrical Engineering
sjcho75@ulsan.ac.kr

Abstract. Recent advances in physics-based sound synthesis have offered huge potential possibilities for the creation of new musical instruments. Despite that research on physics-based sound synthesis is going on for almost three decades, its higher computational complexity has limited its use in real-time applications. Conventional serial computation is inadequate for handling the physics-based sound synthesis of most instruments. To yield computation time compatible with real-time performance, we introduce a parallel approach to the physics-based sound synthesis. In this paper, with a parallel processing engine we implemented the physical modeling for one of traditional Korean plucked string instruments, called Gayageum, which has 12 silk strings. Analysis and simulation results suggest that our parallel approach has the potential to support the real-time sound synthesis of the Gayageum instrument. Moreover, our parallel approach outperforms today's DSPs in terms of performance and energy efficiency.

Keywords: Music synthesis, Physical modeling, Plucked-string instrument, Parallel processing, Data parallel architectures.

1 Introduction

Physics-based methods of musical sound synthesis have received increasing attention in the last three decade [1],[2]. Currently, the musical synthesis algorithms based on physical models are becoming more and more efficient with high quality to imitate natural instruments. Physical modeling can provide useful information for the acousticians about which are the most important phenomena during sound production and how would the sound of the instrument change by varying its physical properties. Several approaches to physical modeling can be found in [3-5]. In [3], a very simple system which contains three components in a feedback loop (e.g., a lowpass filter, a delay, and a gain) was introduced to synthesize the sound of plucked-string instruments. In [4],[5], physical modeling using digital waveguides was used to implement the sound synthesis of plucked string instruments such as acoustic guitar, banjo and mandolin. In this method, the same wave equation can be applied to any perfectly elastic medium which is displaced along one-dimensional waveguides and extensions to two and three

---

* Corresponding author.

dimensions are also possible. While digital waveguides are physically relevant abstractions, its computational has limited its use in real-time applications.

Application-specific integrated circuits (ASICs) can meet the needed performance for such sound synthesis algorithms, but they provide limited, if any, programmability or flexibility needed for varied application requirements.

General-purpose microprocessors (GPPs) offer the necessary flexibility and inexpensive processing elements. However, they will not be able to meet the much higher levels of performance required by high computational synthesis workloads. This is because they lack the ability to exploit the full data parallelism available in these applications.

Among many computational models available for multimedia applications, single instruction multiple data (SIMD) processor arrays are promising candidates for such applications since they replicate the data, data memory, and I/O to provide high processing performance with low node cost. Whereas instruction-level or thread-level processors use silicon area for large multi-ported register files, large caches, and deeply pipelined functional units, SIMD processor arrays contain many simple processing elements (PEs) for the same silicon area. As a result, SIMD processor arrays often employ thousands of PEs while possibly distributing and co-locating PEs with the data I/O to minimize storage and data communication requirements.

This paper presents a parallel approach to the physical modeling of one of Korean traditional plucked string instruments, called Gayageum, to meet the computational time required by such algorithms. This paper also evaluates our parallel approach using a representative SIMD array architecture including 12 processing elements (PEs) in comparison to today's high-performance digital signal processor (DSP) using architectural and workload simulation. Analysis and simulation results suggest that our parallel approach has the potential to deliver three orders of magnitude greater performance and two orders of magnitude of energy efficiency than available DSPs which have the same 32-bit fixed-point datapath implemented in the 130nm CMOS technology.

The rest of the paper is organized as follows. Section 2 presents background information including Gayageum, a Korean traditional plucked string instrument, the physical modeling of plucked string instruments using digital waveguides, the modeled data parallel architecture, and methodology infrastructure for the performance and energy efficiency evaluation of our parallel approach. Section 3 describes a parallel implementation of the physical modeling of the Gayageum instrument. Section 4 analyzes the performance of our parallel approach and compares our approach to today's DSPs on both performance and energy efficiency, and Section 5 concludes this paper.

2   Background Information

This section describes the selected plucked-string instrument, Gayageum, the physical modeling of plucked string instruments using digital waveguides, the modeled data parallel architecture, and methodology infrastructure for the performance and energy efficiency evaluation of our parallel approach to the physical modeling of Gayageum.

2.1   Gayageum: A Korean Traditional Plucked String Instrument

Gayageum is a Korean traditional plucked-string instrument. It has twelve silk strings supported by twelve anjoks shown in Figure 1. The strings are plucked with the

fingers to produce a clear and delicate tune. There are two kinds of gayageum: pungyu gayageum (also called beopgeum) and sanjo gayageum. They are used for different musical genres. Pungyu gayageum is used for court music and sanjo gayageum is used for folk music. Sanjo gayageum differs from pungyu gayageum in size and structure. Pungyu gayageum is a little bigger than sanjo gayageum. The body of sanjo gayageum is made of part chestnut tree and part paulownia. Pungyu gayageum is entirely made of paulownia dug out the back side.
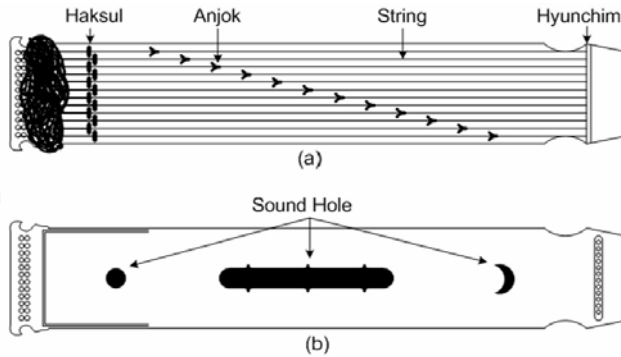
Fig. 1. Gayageum: a Korean traditional plucked-string instrument

2.2 Physical Modeling of Plucked String Instruments Using Digital Waveguides

Many individuals and research groups have studied the physical modeling of plucked string instruments [9],[10]. When plucked string instruments are plucked, energy propagates to the vibrating string and to the body of the instrument. The body and the string start to resonate on particular frequencies until the vibration dampened due to losses in the body and string [9]. The string acts as the sound source while the plucking event and body response act as an excitation signal to the string model [10].

More recently developed digital waveguide methods [4],[5] follow a different path to the physical modeling: the wave equation is first solved in a general way to obtain traveling waves in the medium interior. The traveling waves are explicitly simulated in the waveguide model. The traveling waves must be summed together to produce a physical output. In the lossless case, a traveling wave between two points in the medium can be simulated using nothing but a digital delay line. In the general linear case, the commutativity of linear time-invariant systems allows the losses and dispersion to be lumped at discrete points such that most of the simulation still consists multiply-free delay lines. Figure 2 shows a block diagram of the string model which consists of excitation (plucking), string vibration, and body radiation.

The excitation signal $x(n)$ is stored in a wave-table and used to excite the string model. L denotes delay line length, $F(z)$ is the fractional delay, and $H_l(z)$ represents the loop filter. Output signal $y(n)$ is obtained from the following equation (1).

$$Y(z) = \frac{1}{1 - z^{-L} F(z) H_l(z)} X(z).$$
(1)
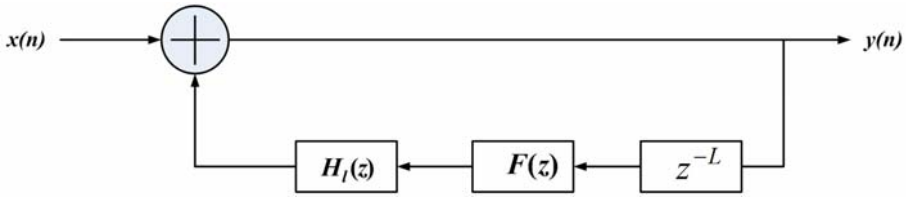
Fig. 2. The block diagram of the string model

Plucked string instruments usually consist of several strings. To model the plucked string instrument, tremendous computational and memory requirements are required. We prefer to overcome the computational burden inherent in the physical modeling of plucked string instruments with a parallel approach using a representative SIMD array system. The next section briefly introduces an overview of the baseline SIMD array architecture.

2.3 Baseline SIMD Array Architecture

Figure 3 shows a block diagram of the baseline SIMD array architecture [11] for the physical modeling of plucked string instruments. This SIMD processor architecture is symmetric, having an array control unit (ACU) and an array consisting of processing elements (PEs). When data are distributed, the PEs execute a set of instructions in a lockstep fashion. Each PE is associated with a specific portion of an input data, allowing input data to be retrieved and processed locally. Each PE has a reduced instruction set computer (RISC) data-path with the following minimum characteristics:

- ALU - computes basic arithmetic and logic operations,
- MACC - multiplies 32-bit values and accumulates into a 64-bit accumulator,
- Sleep - activates or deactivates a PE based on local information,
- Pixel unit - samples pixel data from the local image sensor array,
- Three-ported general-purpose registers (16 32-bit words),
- Small amount of local storage (256 32-bit words),
- Nearest neighbor communications through a NEWS (north-east-west-south) network and serial I/O unit.
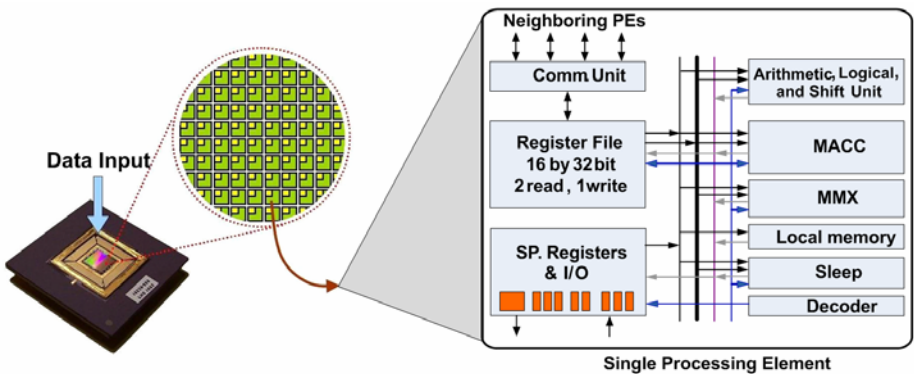
Fig. 3. A block diagram of the SIMD array system

Table 1 summarizes parameters of our modeled architecture. An overall methodology infrastructure is presented next.

Table 1. Modeled architecture parameters

Parameter	Value
System Size (Number of PEs)	12
VLSI Technology	130 nm
Clock Frequency	720 MHz
intALU/intMUL/Barrel Shifter/intMACC/Comm	1 / 1 / 1 / 1 / 1
Local Memory Size [word]	256 [32-bit word]

2.4 Methodology Infrastructure

Figure 4 shows our methodology infrastructure which is divided into three levels: application, architecture, and technology. At the application level, we used an instruction-level simulator to profile execution statistics, such as issued instruction frequency, PE utilization, and PE memory usage. At the architecture level, we used the heterogeneous architectural modeling (HAM) of functional units for processor arrays [12]. The design parameters were then passed to the technology level. At the technology level, we used the Generic System Simulator (GENESYS) [13] to calculate technology parameters (e.g., latency, area, power, and clock frequency) for each configuration. Finally, we combined the database (e.g., cycle times, instruction latencies, instruction counts, area, and power of the functional units), obtained from the application, architecture, and technology levels, to determine execution times and energy efficiency for each case.
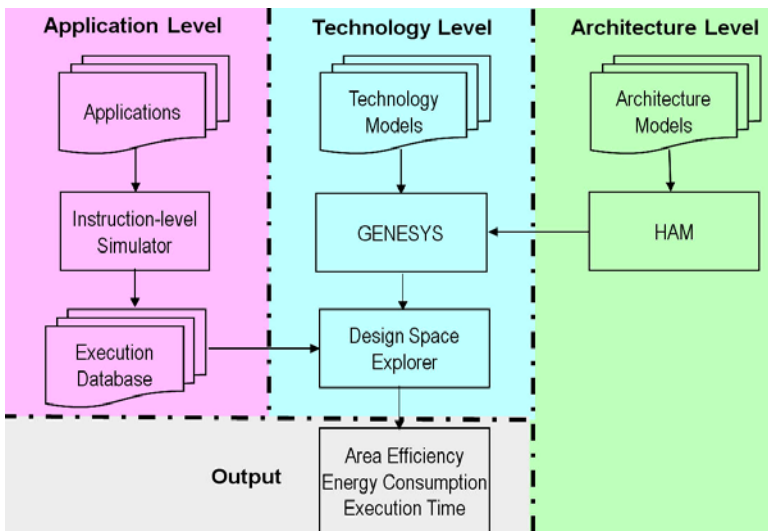
Fig. 4. A methodology infrastructure

3   Parallel Approach to Physical Modeling of Gayageum

In this study, we used a 12-processors array to handle 12 string models of Gayageum, which consists of 12 silk thread strings, where each PE operates for modeling of each string, as shown in Figure 5. Each PE has own local memory which contains excitation data as well as immediate data for synthesizing. In addition, each PE has the parameters of a_1 (loop filter coeff), g (loop filter gain), and h_0, h_1 (Lagrange interpolator coefficients) in its local memory to process acoustical characteristics of each string including delay line length L, the chosen fractional delay $F(z)$ in (2), and the loop filter $H_l(z)$ in (3) in parallel. As a result, we can efficiency obtain the synthesized output sound of Gayageum.

$$F(z) = h_0 + h_1 z^{-1}. \tag{2}$$

$$H_l(z) = \frac{(1+a_1)g}{1+a_1 z^{-1}}. \tag{3}$$

The next section describes more details about the performance evaluation of our parallel approach and compares the parallel approach to today's DSPs in terms of performance and energy efficiency.
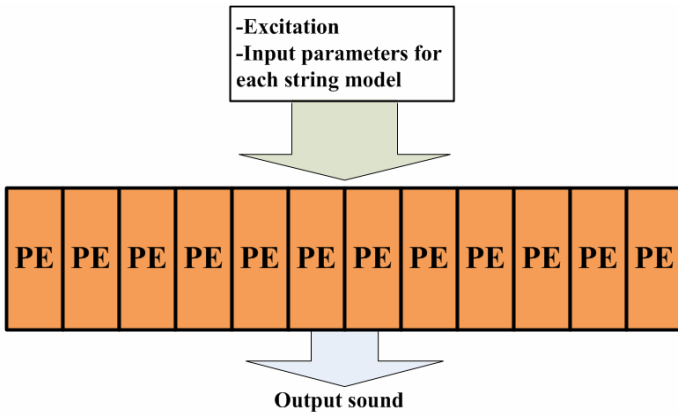
Fig. 5. A 12-processors array model for physical modeling of Gayageum

4   Performance Evaluation

To evaluate the performance and efficiency of our parallel approach to the physical modeling of Gayageum, we used cycle accurate simulation and technology modeling. We developed the physical modeling algorithm in its respective assembly language for the SIMD array system. We then combined the execution statistics (e.g., instruction count, execution cycle time, and PE utilization) with GENESYS predictions to evaluate the benchmark's energy consumption and energy efficiency. The metrics of execution time, sustained throughput, and energy efficiency of each case form the basis of the study comparison, defined in Table 2.

Table 2. Summary of evaluation metrics

Execution time	Sustained throughput		Energy efficiency	
$t_{exec} = \dfrac{C}{f_{ck}}$	$\eta_E = \dfrac{O_{exec} \cdot U \cdot N_{PE}}{t_{exec}} [\dfrac{\text{Gops}}{\text{sec}}]$		$\eta_E = \dfrac{O_{exec} \cdot U \cdot N_{PE}}{Energy} [\dfrac{\text{Gops}}{\text{Joule}}]$	

where C is the cycle count, f_{ck} is the clock frequency, O_{exec} is the number of executed operations, U is the PE utilization, and N_{PE} is the number of processing elements.

4.1   Performance Comparison

This section compares the performance of our parallel approach and commercial processors (TMS320C6416 DSP [14], ARM7TDMI [15], and ARM926EJ-S [16]). A comparison between our parallel approach and commercial products carries unavoidable errors. However, the objective of this study is to show the potential of the proposed parallel approach for improved performance and energy efficiency, rather than a precise performance and efficiency comparison. In the experiments, we assume that our baseline SIMD array and TI TMS320C6416 operate at the same 720 MHz clock rate, while ARM926EJ-S and ARM7TDMI operate at 250 MHz and 133 MHz, respectively. In addition, all the architectures implemented in the 130nm CMOS technology have the same 32-bit fixed-point datapath. In addition, all for versions for the program have the same data sets and calling sequences. Figure 6 shows the performance comparison of the physical modeling of Gayageum among four different architectures (SIMD130, TMS320C6416 DSP, ARM7TDMI, and ARM926EJ-S). Our parallel approach provides three orders of magnitude speedup over other processors for all cases. The magnitude of the speedup suggests that our parallel approach has the potential to provide improved performance for the physical modeling.
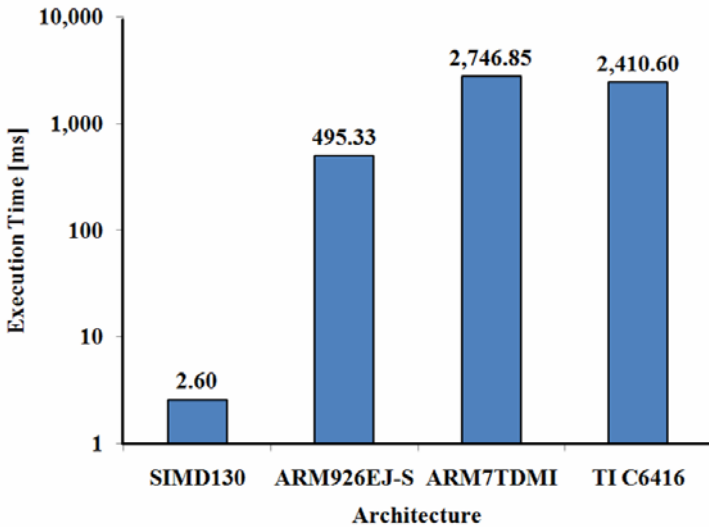
Fig. 6. Execution time comparisons

Figure 7 shows additional data presenting the distribution of issued instructions for the four different versions (SIMD130, ARM926EJ-S, ARM7TDMI, and TMS320C6416) of the programs. Each bar divides the instructions into the arithmetic-logic-unit (ALU), memory (MEM), control unit (CONT), and communication (COMM). The use of our parallel approach reduces a significant number of instruction counts for all the cases in which the number of COMM instructions for both TI DSP and ARM processors are zero since they are single processors. Overall, SIMD130 outperforms both TI DSP and ARM processors in terms of consistently reducing the number of instructions and execution cycles required for the physical modeling benchmark. These performance results are combined with power parameters for each function unit to calculate energy efficiency, which is presented next.
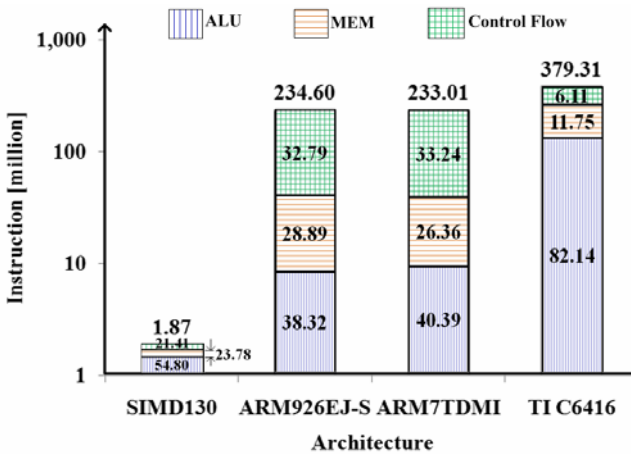
Fig. 7. Instruction count comparison

4.2 Energy Efficiency Comparison

A second evaluation parameter for the four versions of each program is energy efficiency which is defined as the task throughput achieved per unit of Joule. The ARM processors evaluated in this study are low-power designs, while the TMS320C6416 DSP is high-performance design. All the versions considered in this study were implemented in the same 130 nm CMOS process. The focus of this evaluation is energy-related performance since energy is a critical design parameter for handheld products. Results are listed in Table 3.

This analysis suggests that SIMD130 obtains substantial execution throughput over other processors. In addition, SIMD130 utilizes the greater energy efficiency over other processors, providing longer encoding times. Figure 8 shows each version in the throughput-encoding time plane. Even accounting for unavoidable errors in computing research and products, this figure suggests that further study is merited.

Table 3. Performance Comparison among SIMD130, ARM Processors, and TI C6416

Parameter	Units	SIMD130	ARM926EJ-S	ARM7TDMI	TI C6416
Technology	[nm]	130	130	130	130
Clock Frequency	[MHz]	720	250	133	720
Average Power	[mW]	9.58	120.00	23.94	950.40
Average Throughput	[MIPS]	6,821	275	120	5,760
Execution time [12-notes]	[ms]	2.60	495.33	2,746.85	2,410.60
Consumed Time (10 W h battery)	[hr]	401,903.97	168.24	152.07	16.53
Energy	[uJoule]	25	59,439	65,760	604,783
Energy Efficiency	[Gops/Joule]	711.96	2.29	5.00	6.06
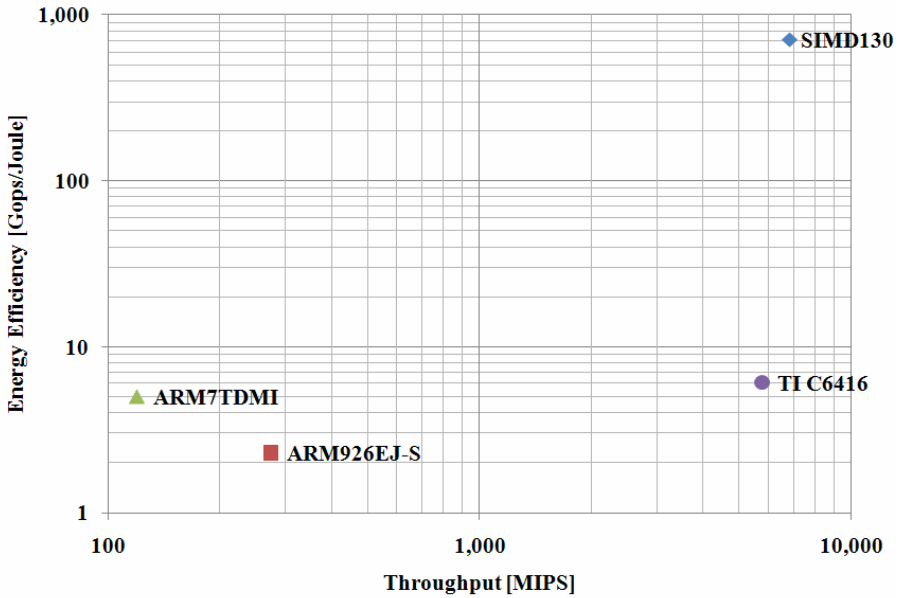
Fig. 8. Energy efficiency and throughput comparison

5   Conclusion

As emerging sound synthesis methods demand more and more tremendous computational throughput with limited power, the need for high efficiency and high performance processing is becoming an important challenge in the field of sound synthesis of

musical instruments. In this paper, we have presented a parallel approach to the physical modeling of Gayageum, a Korean traditional plucked string instrument, using a representative SIMD array system. In addition, we have compared our parallel approach to today's DSPs in terms of processing performance and energy efficiency. Analysis and simulation results for the physical modeling of Gayageum indicate that our parallel approach has the potential to deliver three orders of magnitude greater performance and two orders of magnitude greater energy efficiency than available ARM and DSP processors. These results demonstrate that our parallel approach is a suitable candidate for emerging sound synthesis of musical instruments.

Acknowledgement

References

1. Bilbao, S.: Robust Physical Modeling Sound Synthesis for Nonlinear Systems. IEEE Signal Processing Magazine 24(2), 32–41 (2007)
2. Serra, X.: State of the Art and Future Directions in Musical Sound Synthesis. In: Proceedings of the IEEE 9th Workshop on Multimedia Signal Processing, pp. 9–12 (2007)
3. Karplus, K., Strong, A.: Digital Synthesis of Plucked String and Drum Timbres. Computer Music J. 7(2), 43–55 (1983)
4. Smith, J.O.: Physical Modeling Using Digital Waveguides. Computer Music J. 16(4), 74–87 (1992)
5. Karjalainen, M., Valimaki, V., Tolonen, T.: Pluck-String Models: From the Karplus-Strong Algorithm to Digital Waveguides and Beyond. Computer Music J. 22(3), 17–32 (1998)
6. Alles, H.G.: Music Synthesis Using Real Time Digital Techniques. Proceedings of the IEEE 68(4), 436–449 (1980)
7. Valimaki, V., Huopaniemi, J., Karjalainen, M., Janosy, Z.: Physical Modeling of Plucked String Instruments with Application to Real-Time Sound Synthesis. J. Audio Eng. Soc. 44(5), 331–353 (1996)
8. Karjalainen, M., Valimaki, V., Jinosy, Z.: Towards High-Quality Sound Synthesis of the Guitar and String Instruments. In: Proceedings of the 1993 International Computer Music Conference, Tokyo, Japan, pp. 56–63 (1993)
9. Karjalainen, M., Smith, J.O.: Body Modeling Techniques for String Instrument Synthesis. In: Proceedings of 1996 International Computer Music Conference, Hong Kong, pp. 232–239 (1996)
10. Karjalainen, M., Valimaki, V.: Model-Based Analysis/Synthesis of the Acoustic Guitar. In: Proceedings Stockholm Music Acoustics Conference, Stockholm, Sweden, pp. 443–447 (1993)
11. Gentile, A., Wills, D.S.: Portable Video Supercomputing. IEEE Trans. on Computers 53, 960–973 (2004)

12. Chai, S.M., Taha, T.M., Wills, D.S., Meindl, J.D.: Heterogeneous Architecture Models for Interconnect-Motivated System Design. IEEE Trans. VLSI Systems, Special Issue on System Level Interconnect Prediction 8(6), 660–670 (2000)
13. Eble, J.C., De, V.K., Wills, D.S., Meindl, J.D.: A Generic System Simulator (GENESYS) for ASIC Technology and Architecture Beyond 2001. In: Proceedings of the Ninth Ann. IEEE Intl. ASIC Conf., pp. 193–196 (1996)
14. TMS320C6416 DSP datasheet,
    `http://focus.ti.com/lit/ds/symlink/tms320c6416.pdf`
15. ARM7TDMI datasheet,
    `http://www.atmel.com/dyn/resources/prod_documents/doc0673.pdf`
16. ARM926EJ-S datasheet,
    `http://www.datasheetarchive.com/ARM9EJ-S-datasheet.html`

Robust Acoustic Source Localization with TDOA Based RANSAC Algorithm

Peihua Li and Xianzhe Ma

School of Computer Science and Technology, Heilongjiang University
Hei Long Jiang Province, 15001, China

Abstract. Acoustic source localization has been an hot research topic with widespread applications in many fields. In the noisy environment or when the reverberation is considerable, the source localization problem becomes challenging and many existing algorithms deteriorate. The paper proposes a robust algorithm which combines the RANdom SAmple Consensus (RANSAC) algorithm, and the Generalized Cross-Correlation (GCC) based Time Difference of Arrival (TDOA). Experiments in real world data show that the proposed algorithm has significantly better performance than the traditional algorithm.

Keywords: Acoustic source localization, Random Sample Consensus, Generalized Cross-Correlation, Time Difference of Arrival.

1 Introduction

Acoustic source localization based on microphone array has been an hot research topic with widespread applications in many fields, such as in video-conference assisting in the camera pick up the speaking person, in radar, sonar localizing radiating sources. Despite many years of research, in the noisy environment or when the reverberation is considerable, the problem of acoustic source localization remains challenging [1].

Among varying algorithms for acoustic source localization, those based on Time delay of Arrival (TDOA) are quite popular [2]. These methods first compute TDOAs between microphone pairs based on Generalized Cross-Correlation (GCC) [3,4], and then compute the source position by optimizing some criterions [5]. In clean environments, these methods work well however the performance may deteriorate in challenging conditions. Researchers propose varying algorithms for dealing with noise or reverberation, including fusion algorithm based on multiple sensor pairs, multichannel cross-correlation algorithm [6] or adaptive eigenvalue decomposition algorithm [2].

In this paper,we present a robust algorithm aiming at estimating acoustic source location in reverberant rooms. The approach combines Random Sample Consensus (RANSAC) algorithm [7] and the Generalized Cross-Correlation (GCC) based Time Difference of Arrival (TDOA).

The structure of the paper is as follows. Section 2 presents the proposed algorithm. Section 3 gives the experiments with real world data. The concluding remarks are given at last in section 4.

2   Proposed Algorithm for Sound Source Localization

2.1   The Generalized Cross-Correlation

The generalized cross-correlation method, proposed by Knapp and Carter in 1976, is the most popular technique for estimating TDOA. The signals received at two spatially separated microphones can be mathematically modeled as follows:

$$x_1(t) = s_1(t) + n_1(t),$$
$$x_2(t) = s_1(t + \tau) + n_2(t), \tag{1}$$

where $s_1(t)$ is the acoustic signal without contamination, $n_1(t)$ and $n_2(t)$ are background noise received at the two microphones, respectively. Here we assume that $s_1(t)$ is uncorrelated with noise $n_1(t)$ or $n_2(t)$.

The generalized cross-correlation of $x_1(t)$ and $x_2(t)$ is defined as

$$R_{12}(\tau) = \int_{-\infty}^{\infty} W_{12}(\omega) X_1(\omega) X_2^*(\omega) e^{j\omega\tau} d\omega, \tag{2}$$

where $X_1(\omega)$ is the Fourier transform of $x_1(t)$, $X_2^*(\omega)$ is the complex conjugate of the Fourier transform of $x_2(t)$, and $W_{12}(\omega)$ is a weighing function. The delay estimate is obtained at the peak position of the generalaized cross-correlation $R_{12}(\tau)$.

Proper selection of the weighing function can, to some degree, alleviate the effect of noise. In this paper we select the following weighing function

$$W_{12}(\omega) = \frac{1}{|X_1(\omega) X_2^*(\omega)|^\lambda}, \tag{3}$$

where $0.5 \leq \lambda \leq 1$ [8], and we select $\lambda = 1$.

2.2   Sound Source Localization with TDOAs

Given a pair of microphones m_{i1} and m_{i2}, let $\overrightarrow{m_{i1}}$ and $\overrightarrow{m_{i2}}$ denote their position, respectively. The TDOA τ_i of this pair satisfies the following equation:

$$\| r_s - \overrightarrow{m_{i1}} \| - \| r_s - \overrightarrow{m_{i2}} \| = \tau_i \cdot c, \tag{4}$$

where c is the sound speed in air, r_s is the source position. Obviously, from Equation 4, the source S should be locate on a hyperboloid as shown in Figure 1. For the source that is far from the sensors, the hyperboloid asymptotically converges to a cone [5].

Given n pairs of TDOAs $\tau_i, i = 1, \cdots, n$ estimated from microphone pairs, the source can be determined by the intersection of the hyperboloids of all pairs.
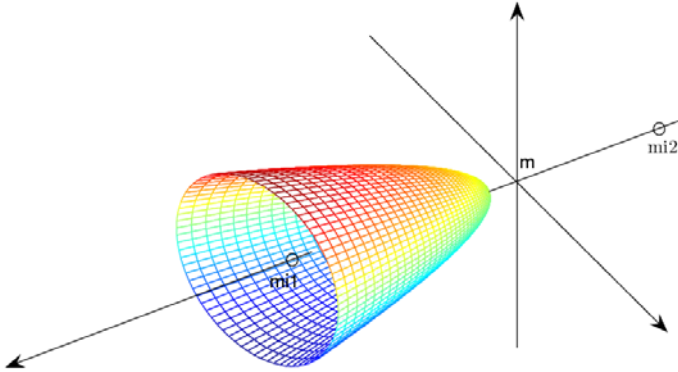
Fig. 1. The source is located on a hyperboloid determined by Equation 4

However in practice due to noise in estimation of TDOAs, the intersection set may always be empty. So usually the following criterion is used:

$$\hat{r}_s = \arg\min_{r_s} J_{TDOA}(r_s) = \sum_{i=1}^{n} \frac{(\tau_i - T((\overrightarrow{m_{i1}}, \overrightarrow{m_{i2}}), r_s))^2}{var[\tau_i]} \ , \tag{5}$$

where $T((\overrightarrow{m_{i1}}, \overrightarrow{m_{i2}})$ denotes the true TDOA, and $var[\tau_i]$ denotes the variation of estimated TDOAs. Other criterions may also be used to estimate the source location given pairs of TDOAs, and details may be referred to Ref. [5].

2.3   RANSAC-Based Localization Algorithm

The RANdom SAmple Consensus (RANSAC) algorithm, first proposed by Fischler and Bolles, is very robust to estimate parameters of a mathematical model from a set of observed data which contain significant outliers.

Our source localization algorithm based on the RANSAC Algorithm is described as follows.

(1) Input N-pair of TDOAs τ_i, $i = 1, \cdots, n$ estimated from N-pair microphones, a tolerable threshold ϵ, and the maximum iteration number K. Set the initial iteration number $k = 1$.
(2) Select randomly a set of n-pair $((n \ll N))$ TDOAs, and estimate the source position $\hat{r}_s$ according to Equation 5.
(3) Among the remaining $(N - n)$-pair of TDOAs, compute the theoretical TDOA $\hat{\tau}_{l_i}$, $i = 1, \cdots, N - n$ with respect to the estimated position $\hat{r}_s$. Make a record of the *persistent set* S_k of τ_{l_j}, $j = 1, \cdots, M_k$ for which $|\hat{\tau}_{l_j} - \tau_{l_j}| < \varepsilon$, $j = 1, \cdots, M_k$.
(4) $k = k + 1$, if $k < K$, then goto step (1).

(5) Among the *persistent sets* S_k, $k = 1, \cdots, K$, find out the one S^* with the maximal number of elements. Estimate the source position with the TDOAs in the set S^*.

(6) Output S^*, stop.

3   Experiment

The data were recorded by two microphone array that are placed 0.8m apart, each of which has a radius of 10cm and eight microphones in an instrumented meeting room [9]. The data are freely available at the web site http://www.idiap.ch/av16_3corpus/. In this paper we used the data set seq01-1p-0000 (single human speaker, static, at several locations).

We can estimate 56-pair of TDOAs from the two microphone arrays. In each TDOA estimation procedure, the signals were segmented in individual frames using a 512-point (25.6ms) half-overlapping Hammming window and a 1024-point FFT was used. The method introduced in section 2.1 is used to compute the peak point of the generalized cross-correlation function.

Table 1 gives comparison of source position $r_s = (x, y, z)$ between the proposed algorithm and the traditional one. It can be seen that, for all cases, the proposed algorithm has significantly better performance in terms of localization error.

Figure 2 shows the source position error in every frame for the position g. It can be seen that in highly reverberant environment, the tradition algorithm has awfully large errors. In contrast, the proposed algorithm demonstrates persistent, nice performance.

Table 1. Comparison of source position (x,y,z)(unit: m) between the proposed and tradition algorithms

Positions	Mean Error of the Proposed	Mean Error of the Traditional
a	(0.69,0.57,0.87)	(1.98,1.47,6.86)
b	(0.32,1.09,0.91)	(1.28,3.19,7.37)
c	(0.40,0.47,0.76)	(0.73,0.94,3.18)
d	(0.52,0.54,0.96)	(0.99,1.22,5.10)
e	(0.23,0.28,0.58)	(0.37,0.62,2.03)
f	(0.36,0.50,0.83)	(0.67,1.02,3.46)
g	(0.22,0.54,0.65)	(1.16,1.97,6.09)
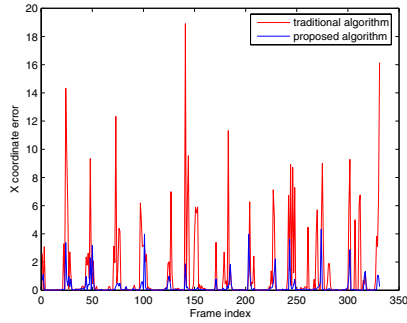
(a) X coordinate error
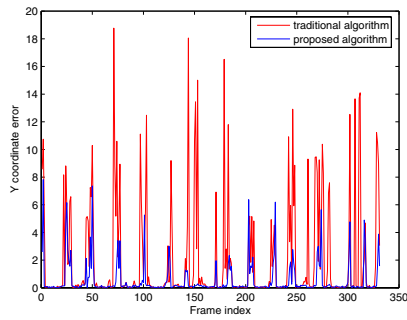
(b) Y coordinate error
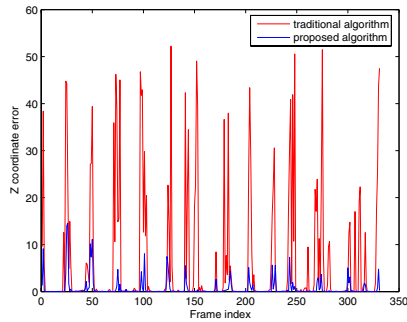
(c) Z coordinate error

Fig. 2. The source position error in every frame for the position g

4   Conclusion

This paper presents a sound source localization algorithm. The approach combines TDOA based technique with a very robust RANSAC algorithm. Experiments are made in real-world data in which the noise and and reverberation are considerable, and comparison with traditional algorithm shows that the

proposed algorithm has significant performance improvement. The current work is computationally intensive and future work is to address this by optimizing the program or developing parallel algorithm.

Acknowledgements

References

1. Benesty, J., Chen, J., Huang, Y.: Microphone Array Signal Processing. Springer, Berlin (2008)
2. Chen, J., Benesty, J., Huang, Y.: Time delay estimation in room acoustic environments: an overview. EURASIP Journal Applied Signal Processing, 1–19 (2006)
3. Knapp, C.H., Carter, G.C.: The generalized correlation method for estimation of time delay. IEEE Trans. Acoust. Speech Signal Process. 24(4), 320–327 (1976)
4. Ianniello, J.P.: Time delay estimation via cross-correlation in the presence of large estimation errors. IEEE Trans. Acoust. Speech Signal Process 30, 998–1003 (1982)
5. Brandstein, M.S.: A framework for speech source localization using sensor arrays. Brown University, USA (1995)
6. Chen, J., Benesty, J., Huang, Y.: Robust time delay estimation exploiting redundancy among multiple microphones. IEEE Trans. Speech Audio Processing 11, 549–557 (2003)
7. Fischler, M.A., Bolles, R.C.: Random Sample Consensus: A Paradigm for Model Fitting with Applications to Image Analysis and Automated Cartography. Comm. of the ACM 24, 381–395 (1981)
8. Rabinkin, D., Renomeron, R.J., Dahl, A., et al.: A DSP implementation of source loeation using microphone arrays. In: Proceedings of the SPIE, pp. 88–99 (1996)
9. Lathoud, G., et al.: An Audio-Visual Corpus for Speaker Localization and Tracking. In: Proceedings of the MLMI 2004 Workshop (2004)

Facial Expression Recognition with Local Binary Pattern and Laplacian Eigenmaps

Zilu Ying[1, 2], Linbo Cai[1], Junying Gan[1], and Sibin He[1]

[1] School of Information, Wuyi University, Jiangmen, Guangdong, 529020
`ziluy@163.com`
[2] School of Electronics and Information Engineering, Beihang University, Beijing 100083

Abstract. A new approach to facial expression recognition is constructed by combining the Local Binary Pattern and Laplacian Eigenmaps. Firstly, each image is transformed by an LBP operator and then divided into 3×5 non-overlapping blocks. The features of facial expression images are formed by concatenating the LBP histogram of each block. Secondly, linear graph embedding framework is used as a platform, and then Laplacian Eigenmaps is developed under this framework and applied for feature dimensionality reduction. Finally, Support Vector Machine is used to classify the seven expressions (anger, disgust, fear, happiness, neutral, sadness and surprise) on JAFFE database. The maximum facial expression recognition rate of the proposed algorithm reaches to 70.48% for person-independent recognition, which is much better than that of LBP+PCA and LBP+LDA algorithms. The experiment results prove that the facial expression recognition with local binary pattern and Laplacian Eigenmaps is an effective and feasible algorithm.

Keywords: Facial expression recognition, Local binary pattern, Graph embedding, Laplacian Eigenmaps.

1 Introduction

With advances in computer technology, artificial intelligence and pattern recogni- tion technology have obtained rapid development. As one of biometrics technology, facial expression recognition (FER) has attracted many researchers' attention [1-2]. Two key steps in FER are expression feature extraction and dimensionality reduction. Local Binary Pattern (LBP) proposed by Ojala, as a powerful method of texture description, is widely used in feature extraction due to its high discriminative power, tolerance against illumination changes and computational simplicity [3]. In feature dimensionality reduction step, many methods have come into being. Among the linear algorithm, i.e. the subspace learning algorithm, principal component analysis (PCA) and linear discriminant analysis (LDA) are the two most popular ones. On the other hand, ISOMAP, LLE, Laplacian Eigenmaps (LE) are three recently developed nonlinear algorithms to conduct nonlinear dimensionality reduction for the data set lying on or nearly on a lower dimensional manifold. More recently, a framework algorithm for dimensionality reduction named Graph Embedding was introduced by Yan [4]. In this framework, the above mentioned algorithms, such as PCA, LDA, ISOMAP, LLE and

Laplacian Eigenmaps, can all be reformulated in this unified framework; and their differences lie in the strategy to design the graph and the embedding type.

In this paper, we present a novel facial expression recognition method with LBP and LE. The remainder of the paper is organized as follows. The LBP operator and its application for facial expression feature extraction are introduced in Section 2. In Section 3, by using graph embedding as a general platform, the Laplacian Eigenmaps algorithm is reformulated. The experiments and the result analysis of facial expression recognition method on JAFFE database are described in Section 4. Finally, we give the conclusion remarks in Section 5.

2   Feature Extraction with LBP

The original LBP operator, introduced by Ojala et al. [3], is a powerful method of texture description. Fig. 1 shows an example of computation of LBP operator in a 3×3 neighborhood which contains 9 gray values. Pixels encircling the central pixel are labeled 1 if their values are greater than or equal to the value of the central pixel; 0, otherwise. The LBP binary code of the center pixel in the neighborhood is composed of those labels anticlockwise. Finally, the local binary pattern for center pixel is obtained by converting the binary code into a decimal one.
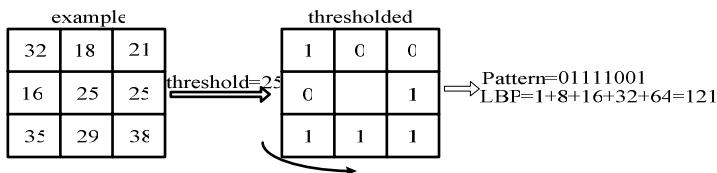
Fig. 1. The computation example of LBP in a 3×3 neighborhood

However, the original LBP operator is not very robust against local changes in the texture, caused, for example, by varying viewpoints or illumination directions. So, features calculated in a local 3×3 neighborhood can not capture large scale structures that may be the dominant texture features of images. In order to overcome the disadvantage of the original LBP operator, it is necessary to extend it to a multi- resolution LBP operator [5]. Let R denote the radius of a circle, the multi-resolution LBP operator is denoted as Eq. (1).

$$LBP_{P,R} = \sum_{j=0}^{j=P-1} S(g_j - g_c)2^j .$$ (1)

Where, $S(x)$ is defined as $S(x) = \begin{cases} 1, x \geq 0 \\ 0, x < 0 \end{cases}$, and $g_j (j = 1, \cdots, P-1)$ corresponds to the gray values of P uniform pixels on a circle radius R ($R>0$) that form a circularly symmetric set of neighbors. Fig. 2 shows several different radiuses of LBP operators.

P=8, R=1   P=8, R=2   P=16, R=4
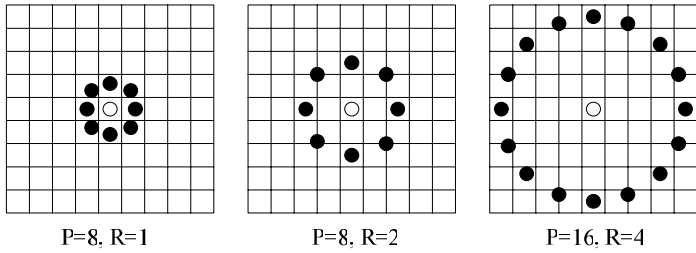
Fig. 2. Three circularly symmetric LBP operators with different radiuses

The parameters P and R influence the performance of LBP operator. Using an operator that produces a large amount of different labels makes the histogram long and thus more time consumed for calculating the distance matrix. On the other hand, choosing a small number of labels makes the feature vector shorter but means losing more information. Moreover, a small radius of the operator makes the information encoded in the histogram more local [6]. According to statistics, the probability of each pattern is not the same in images. The researchers introduced a uniform pattern operator, restricting the LBP binary code, which is at most one 0-1 and 1-0 transition, denoted as $LBP_{P,R}^{u2}$. In an $LBP_{8,R}^{u2}$ operator, for example, there are 256 labels for standard LBP and 58 uniform patterns. Accordingly, there are 65536 labels and 242 uniform patterns in an $LBP_{16,R}^{u2}$ operator. All of non-uniform pattern is replaced by a fixed pattern. So, the corresponding number of patterns is 59 for $LBP_{8,R}^{u2}$ and 243 for $LBP_{16,R}^{u2}$ operators respectively.

$LBP_{8,2}^{u2}$

(a) The original image   (b) LBP image   (c) LBP histogram
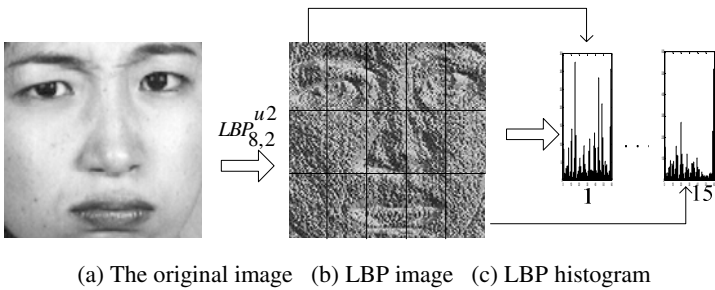
Fig. 3. Image description with LBP

Fig. 3 shows an illustration of the process of feature extraction by LBP operator. First, the facial expression image is transformed by a uniform pattern operator. Then, the LBP image is equably divided into several non-overlapping blocks. In our experiments, we choose the 3×5 partition for the feature extraction with LBP. As shown in Fig. 3 (b), the features of two eyes are approximately included in the first row. The second row includes the features of the most face and nose. And the third row mostly shows the features of mouse. The LBP histograms are calculated for each block.

Finally, the LBP histograms of all blacks are concatenated into a long series of histogram as a single vector showed in Fig. 3 (c). The final long histogram is used as the facial expression image features [7].

3 Feature Dimensionality Reduction with LE

Laplacian Eigenmaps (LE), proposed by Mikhail Belkin and Partha Niyogi [8], considers that points which are near in high dimensional space should also be near when projected into lower dimensional space. The core algorithm is very simple. It consists of a few local computations and one sparse eigenvalue problem. The solution reflects the intrinsic geometric structure of the manifold which is modeled by an adjacency graph and can find the optimal linear approximations to the Eigen functions of the Laplace Beltrami Operator on the face manifold. What sets Laplacian Eigen- maps apart is that the choice of weights used in constructing the graph. The subsequent spectral analysis is formally justified as a process which "simplifies" the manifold structure [9].

Denote the sample set as $X = [x_1, x_2, \cdots, x_N]$, $x_i \in R^m$. We construct a weighted graph with N vertices, one for each point, and the set of edges connecting neighboring points to each other. The LE algorithm is presented as follows [10]:

Step 1. Constructing the Graph.

We put an edge between vertices i and j if x_i and x_j are "close". There are two variations:

(a) ε -neighborhoods, $\varepsilon \in R$. Vertices i and j are connected by an edge if $\|x_i - x_j\|^2 < \varepsilon$.

(b) k nearest neighbors, $k \in N$. Vertices i and j are connected by an edge if $i \in N_k(j)$ or $j \in N_k(i)$ where $N_k(i)$ is the indices of the k nearest neighbors for the sample x_i.

Step 2. Choosing the weights.
There are also two variations for weighting the edges:

(a) Heat kernel [11]. If vertices i and j are connected, put $W_{ij} = \exp\{-\|x_i - x_j\|^2 / t\}$, where parameter $t \in R$.

(b) Simple-minded. $W_{ij} = 1$, if and only if vertices i and j are connected by an edge.

Step 3. Laplacian eigenvalue Computation.

Let $G = \{X, W\}$ be an undirected weighted graph with vertex set X and the adjacency matrix $W \in R^{N \times N}$. The diagonal degree matrix D and the Laplacian matrix L of the graph G are defined as

$$L = D - W, D_{ii} = \sum_{j \neq i} W_{ij}, \forall i \qquad (2)$$

We solve the generalized eigenvector problem as defined by

$$Lv = \lambda Dv \, , \qquad (3)$$

Let $v_0, v_1, \cdots, v_{n-1}$ be the eigenvectors of equation (3) corresponding to the sorted eigenvalues $\lambda_0 \le \lambda_1 \le \lambda_2 \le \cdots \le \lambda_{n-1}$. We leave out v_0 and use the first m eigenvectors $(v_1, v_1, \cdots, v_m)$ for embedding in m-dimensional Euclidian space.

The LE algorithm as well can be obtained by graph embedding framework which can be used as a general platform to develop new algorithm for dimensionality reduction by designing graphs with special motivation [8]. In this paper, specially, we choose graph embedding framework as a platform to generate the Laplacian Eigenmaps whose adjacency matrix is calculated from the Gaussian function $W_{ij} = \exp\{-\|x_i - x_j\|^2 / t\}$ if $i \in N_k(j)$ or $j \in N_k(i)$; 0, otherwise.

4   Experiments and Analysis

To prove the validity and efficiency of our method, experiments are carried out on Japanese Female Facial Expression (JAFFE) database which consists of 213 gray scale images of 10 Japanese females. There are 21 images per person with 7 facial expressions (anger, disgust, fear, happiness, neutral, sadness and surprise). Each expression has 3 to 4 copies of images. In our experiments, we select 3 copies of each expression of each individual (210 images in all) as experiment samples.

All experiments are carried out by "Leave one out" approach, i.e. expression images of 1 out of 10 subjects are used as the testing set, and the other 9 subjects are used as the training sets. Repeat the process for each individual. The final recognition rate is the average recognition rate of ten subjects.

At the beginning of our experiments, all these images are resized to a size of 256×256 pixels and preprocessed by gray normalization and histogram equalization. Gamma correction is used for illumination compensation and $LBP_{8,2}^{u2}$ operator is used.

After LBP operation, all images are divided into 3×5 blocks. We concatenate the histograms of 15 blocks to describe the feature of the image. Linear graph embedding (LGE) framework is used as a platform. The Laplacian Eigenmaps algorithm is constructed on LGE to achieve feature dimensionality reduction. Here, we set the parameter of LE as 2 for better performance. Finally, support vector machines (SVMs) are applied to classify the features into expression categories. There are two kinds of SVM classifiers. One is the one-against-one (OAO) and another is one-against-all (OAA). For a C-classification problem, OAOSVM strategy needs $C \times (C-1)/2$ classifiers. However, OAASVM strategy needs only C classifiers. To simplify the experiments, here, we use the OAASVM strategy for expression classification and the kernel used is a Gaussian radial basis function.

In order to demonstrate the performance of our method, we compare it with that of LBP+PCA and LBP+LDA. The expression recognition rates varies with the dimensions of the embedding space are shown in Fig. 4. As shown in Fig. 4, we see that when the reduced feature dimensions reach 10, the expression recognition accuracy are close
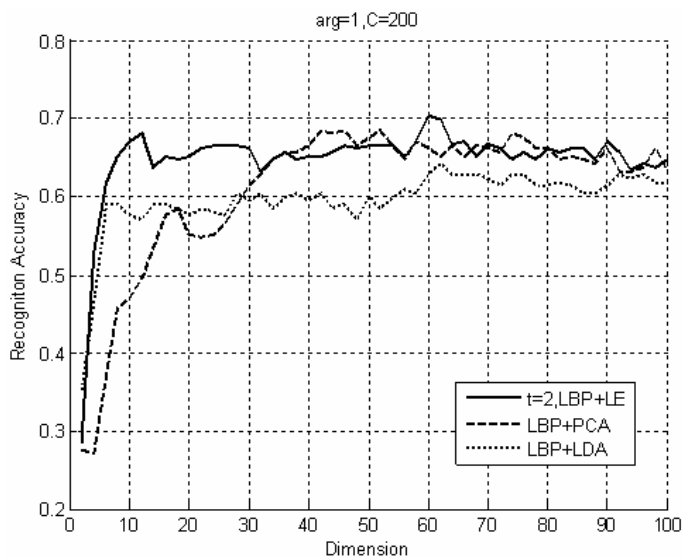
Fig. 4. FER accuracy versus dimensions of the embedding space on JAFFE database

Table 1. FER rates of each expression on JAFFE database

Expression	LBP＋PCA (%)	LBP＋LDA (%)	LBP+LE (%)
Anger	90	73.33	90
Disgust	56.67	70	60
Fear	70	56.67	66.67
Happiness	80	73.33	83.33
Neutral	50	66.67	70
Sadness	56.67	46.67	50
Surprise	76.67	63.33	73.33
Total	68.57	64.29	70.48

to a limit value and will not change much with the dimensions continuing to increase for both the LBP+LE and LBP+LDA algorithms, but the recognition rate continue to increase until the reduced dimension reaches 40 for LBP+PCA algorithm. For the reduced dimensions changing from 2 to 100, the average recognition rate of LBP+LE algorithm is about 64.82%, 60.18% for LBP+PCA algorithm and 59.66% for LBP+LDA algorithm. So, LBP+LE method can obtain better recognition accuracy than that of others. When the dimension increases to 60, LBP+LE algorithm gets the highest recognition rates of 70.48%.

The corresponding recognition rates of above-mentioned algorithms for 7 expressions on JAFFE database is shown in Table 1. As shown in table 1, compared with other expressions, the accuracy of anger is more satisfactory. It gets the highest recognition rate of 90% for LBP+PCA and LBP+ LE. Happiness gets the second highest recognition rate of 80% for LBP+PCA, 73.33% for LBP+LDA and 83.33% for LBP+LE.
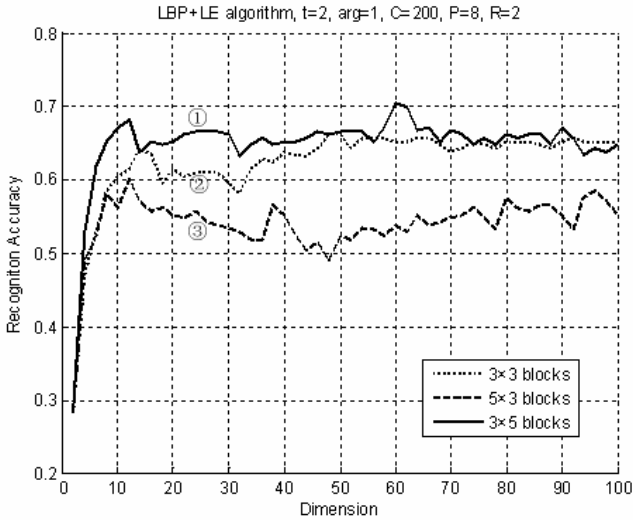
Fig. 5. The influence of different partition methods on FER rates

However, disgust and sadness are not easy to be recognized. The accuracies of these two expressions are almost lower than others. The total accuracy of each method is 68.57%, 64.29% and 70.48% respectively. It can make a conclusion that LBP+LE combination gets better performance than that of LBP+PCA and LBP+LDA.

It is worth noting that several parameters have a great impact on the final recognition rate, especially those parameters of LBP operator, such as the ways of partition, the parameters of P and R. Expression recognition rates changing with the dimensions of the embedding space of different partition methods of LBP operator are shown in Fig.5. As shown in Fig.5, the partition method has a significant influence on the recognition accuracy. Line 1 denotes the recognition result of 3×5 blocking approach. When the dimension reaches to 60, it gets the highest rate of 70.48%. Line 2 denotes the recognition result of 3×3 blocking approach. When the dimension is close to 50, it gets the maximum rate of 66.19%. Line 3 denotes the recognition result of 5×3 blocking approach, which can obtain the highest rate of 60% as the dimension is near to 12. It is clear that the blocking approach of LBP have a significant influence on FER rate. In our experiments, the LBP operator with division of a image into 3×5 blocks can obtain better result than that of 3×3 blocking approach and 5×3 blocking approach.

5   Conclusion

Two key steps of facial expression recognition are feature extraction and dimensionality reduction. LBP is a powerful algorithm for describing the texture feature of images due to its high discriminative power, tolerance against illumination changes and computational simplicity. And LE is a nonlinear dimensionality reduction method which reflects the intrinsic geometric structure of the manifold and has few local computations and can be formulated to a sparse eigenvalue problem. Combined the

advantages of these two approaches, this paper constructs a new approach to FER by combining LBP and LE. Extensive experiments are carried out on JAFFE database and the influence of block division strategies is also studied. The maximum recognition rate in our approach reaches to 70.48%. According to the comparison with other methods, it proves the efficiency of our approach.

However, some potential problems still exist, such as the parameters and the blocking approach of LBP. On the one hand, how to optimize parameters adaptively deserves research profoundly. On the other hand, what blocking approach is the best for facial expression recognition still needs more research in the future work.

Acknowledgements

References

1. Yang, G.L., Wang, Z.L., Wang, G.J.: A Survey of Facial Expression Recognition. J. Techniques of automation and applications 25(4), 1–6 (2006)
2. Wang, H., Leng, Y., Wang, Z., Wu, X.-j.: Generalized PCA Face Recognition by Image Correction and Bit Feature Fusion. In: King, I., Wang, J., Chan, L.-W., Wang, D. (eds.) ICONIP 2006. LNCS, vol. 4233, pp. 227–235. Springer, Heidelberg (2006)
3. Ojala, T., Pietikainen, M., Harwood, D.: A Comparative Study of texture Measures with Classification based on Featured Distributions. J. Pattern Recognition 29(1), 51–59 (1996)
4. Yan, S.C., Xu, D., Zhang, B.Y., et al.: Graph Embedding: A General Framework for Dimensionality Reduction. In: IEEE Computer Society Conference 2005 on Computer Vision and Pattern Recognition, vol. 2, pp. 830–837 (2005)
5. Ojala, T., Pietikainen, M., Maenpaa, T.: Multiresolution Gray Scale and Rotation Invariant Texture Classification with Local Binary Pattern. J. Pattern Analysis and Machine Intelligence, IEEE Transaction 24(7), 971–987 (2002)
6. Ahonen, T., Hadid, A., Pietikäinen, M.: Face Recognition with Local Binary Patterns. In: Pajdla, T., Matas, J(G.) (eds.) ECCV 2004. LNCS, vol. 3021, pp. 469–481. Springer, Heidelberg (2004)
7. Feng, X., Pietikainen, M., Hadid, A.: Facial Expression Recognition with Local Binary Patterns and Linear Programming. J. Pattern Recognition and Image Analysis 15(2), 46–48 (2005)
8. Belkin, M., Niyogi, P.: Laplacian Eigenmaps for Dimensionality Reduction and Data Representation. J. Neural Computation 15(6), 1373–1396 (2003)
9. Chen, J.F., Yuan, B.Z., Pei, B.N.: Face Recognition Using Two Dimensional Laplacian Eigenmap. J. Journal of electronics (China) 25(5), 616–621 (2008)
10. Belkin, M., Niyogi, P.: Laplacian Eigenmaps and Spectral Techniques for Embedding and Clustering. In: Advances in Neural Information Processing Systems 14 (NIPS 2001), pp. 585–591 (2002)
11. Xiao, B.: Heat Kernel Analysis on Graphs. Submitted for degree of Doctor of Philosophy, Department of Computer Science, The University of York (2007)

A Novel Face Detection Method Based on Contourlet Features

Huan Yang, Yi Liu, Tao Sun, and Yongmi Yang

School of Computer Science and Technology
Shandong University, Jinan, 250101, China
yangyang6899@gmail.com

Abstract. This paper primarily investigates a novel face detection method based on contourlet features. In this method, a face-pyramid is developed through contourlet transform, which includes both low and high frequency information to represent face features on multiresolutions and multidirections. The most discriminative features are then selected from the face-pyramid and are trained to construct the classifier by using the cascade boosting algorithm (Adaboost). Speed and capability are important issues for current face detection systems. This method extensively reduces feature demensions and the negative sample numbers step by step, so that the speed is increased radically. Mean-face template matching is adopted finally in the system to ensure a detection of one face in a scanned image. Extensive experiments are conducted and the results show that the proposed method is efficient in detecting frontal faces from cluttered images.

Keywords: Contourlet transform; Face-pyramid; Adaboost; Template matching.

1 Introduction

In the recent years, many effective face detection algorithms have been reported due to its wide applications [1], [2]. The milestone of the development is the robust real-time face detection scheme based on a boosted cascade of Haar features developed by Paul Viola and Michael J. Jones [3], [4]. After this survey, Lienhart R., Kuranov A., Viola P. et al. improved the boosting method, such as Discrete Adaboost, Real Adaboost, Float Adaboost, Asymmetric Adaboost [5], [6], [7]. And someone tried to extract new face features [8]. Rainer Lienhart described some new rotated Haar-like features to resolve the rotated face detection [9]. Yunyang Yan described Walsh Features based on Adaboost to realize the multi-view face detection [10]. Gabor filter and some second generation wavelet transform have also been adopted to extract face features in some face detection systems [11], [12, [13].

This paper attempts to introduce the contourlet transform into face detection system. Contourlet transform is a new second generation wavelet and is an efficient directional multiresolution image representation [14], [15]. Its excellence is able to express smooth contours efficiently using sparse coefficients, which is the natural feature of image. Its distinguishing merits have been demonstrated in several image processing tasks, such as image denoising, image fusion, image contour detection etc. [16], [17].

In this paper, the preprocessed image is decomposed to Laplacian Pyramid and Gauss Pyramid through contourlet transform. Then we construct a face-pyramid, including both low and high frequency information. The face-pyramid is selected and boosted by Adaboost algorithm. Experiment comparative results prove the strong discrimination and good efficiency of those contourlet features. The effectiveness of the proposed method is demonstrated by the experiment results in view of facial variations in position, scale and expression.

2 Contourlet Features Extraction

2.1 Contourlet Transform

The Contourlet transform, introduced by Minh D. and Martin Vetterli [14], is a new image representation scheme, which not only possess the main features of wavelets (namely, multiscale and time frequency localization), but also show a high degree of directionality and anisotropy. It is constructed by combining two distinct and successive decomposition stages: a multiscale decomposition followed by a multidirectional decomposition. The first stage uses Laplacian pyramid (LP) scheme to decompose an image into one coarse image plus a set of LP bandpass images. The second stage applies directional filter bank (DFB) to divide each LP bandpass image into a number of wedge shaped subbands, and thus capturing directional information. Finally, the image can be represented as a set of subbands at multiscales.

We can see these two states from Fig.1. H and G are analysis and synthesis filters respectively. M is the sampling matrix. In frequency partition, there are 8 real wedge-shaped frequency bands. Subbands 0–3 correspond to the mostly horizontal directions, while subbands 4–7 correspond to the mostly vertical directions.

The LP in the contourlet filter bank uses orthogonal filters and downsampling by 2 in each dimension [14]. The lowpass filter defines a unique scaling function $\phi(t) \in L_2(\Re^2)$. Let

$$\phi_{j,n} = 2^{-j} \phi(\frac{t - 2^i n}{2^j}), j \in Z, n \in Z^2.$$

(1)

Then the family $\{\phi_{j,n}\}_{n \in Z^2}$ is an orthogonal basis for an approximation subspace V_j at the scale 2^j. $\{V_j\}_{j \in Z}$ provide a sequence of multiresolution nested subspaces.
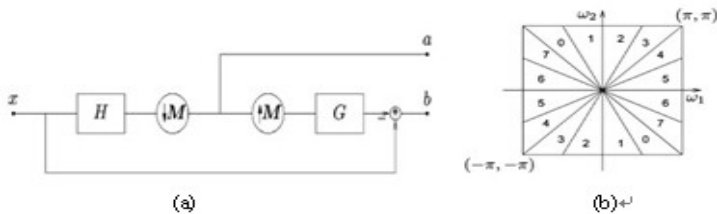
Fig. 1. Contourlet Transform. (a) LP decomposition. (b) Frequency partition.

The difference images live in a subspace W_j is the orthogonal complement of V_j and V_{j-1}. The synthesis filters are highpass filters, defines a continuous function $\varphi^{(i)}(t)$. Let

$$\varphi_{j,n}^{(i)}(t) = 2^{-j}\,\varphi^{(i)}\left(\frac{t-2^j n}{2^j}\right), j \in \mathbf{Z}, n \in \mathbf{Z}^2. \tag{2}$$

For scale 2^j, $\{\varphi_{j,n}^{(i)}\}_{0 \le i \le 3, n \in \mathbf{Z}^2}$ is a tight frame for W_j. The DFB also use orthogonal filters. The DFB is applied to the difference images or the detail subspaces W_j. With orthogonal filters, the discrete contourlet transform has a redundancy ratio less than $4/3$.

The contourlet transform has other several distinguishing features. The approximation error of the Contourlet transform is $O((\log M)^3 M^{-2})$, while the Fourier transform is $O(M^{-1/2})$, and the Wavelet transform is $O(M^{-1})$. The contourlet is better in capturing the geometrical smoothness of the contours than wavelet and Fourier transform. The contourlet filter bank leads to compactly supported contourlet frame and provides a space-domain multiresolutional scheme that offers flexible refinements for the spatial resolution and the angular resolution. At the same time, the contourlet transform has fast filter band algorithms and convenient tree structures.

2.2 Face-Pyramid

Through contourlet transform, an image was decomposed into directional multiscale images. All of the low frequency images compose the Gauss Pyramid (GP), and all of the new high frequency images form the Laplacian pyramid (LP). A sample of contourlet decomposition of a face image is shown in Fig.2. Then the GP subbands are partitioned by DFB into several directional subbands. According to the structure of face, like that eyes and mouth are all mostly horizontal distributed, extracting directional contour information is necessary, especially the horizontal details.
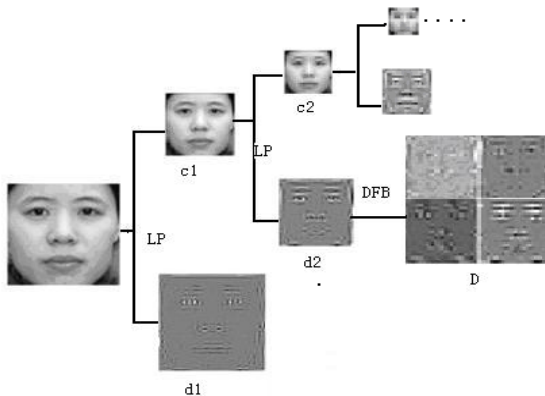
Fig. 2. Contourlet decomposition of a face image. (d2 is decomposed to D in four directions. D is displayed at doubled length.)

Through decomposition, the most energy focuses on the low frequency bands. And the high frequency images represent the edge contour information of the original face image. The general representation of an image using contourlet transform is to choose the series of high frequency images and the last low frequency image. This may loss some critical features in the energy focused images. Through experiment results, it can be seen that lots of the final features are chosen from the low frequency bands. So we choose all of the directional subbands and low frequency images except the following subbands. Because the first one of detail images is influenced by lighting and other noise easily, we only choose the horizontal department from its decomposed subbands. And the original image is abandoned due to its high dimension. Then we array the subbands from coarse scale to fine scale and the new constructed pyramid is called as face-pyramid. The feature vector is obtained by transforming every image (M by N) in the pyramid to images (1 by M*N).

3   The Adaboost Algorithm

In the numerous detection methods, the Adaboost face detection method resolved the speed problem radically and achieved excellent classify performance [3], [4]. In this method the Adaboost algorithm is adopted to train the strong classifier. The Adaboost can seem as a feature selection mechanism. Every weak classifier selects several features and finally a small number of features are combined to form the strong classifier. At the same time, the optimal thresholds are recorded. According to every weak classifier, those images judged as nonface are rejected and the rest are sent to the next state. Almost all of the face images are remained and lots of nonface images are rejected. After several states of processing the number of samples has been reduced and the speed is increased radically. The next classifier is more complicated than the previous one, because the distinguishing task is more and more difficult. The cascade structure is shown in Fig. 3.
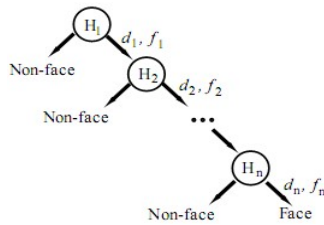
Fig. 3. The Demonstration of the Cascade Structure

For each feature, a weak classifier determines the optimal threshold classification function $h_j(x)$ [4]. It consists of a feature (f_j), a threshold (θ_j) and a parity (p_j).

$$h_j(x) = \begin{cases} 1 & if(p_j f_j(x) < p_j \theta_j) \\ 0 & otherwise \end{cases}. \tag{3}$$

For every state the classifier learning algorithm is shown in Table 1.

Table 1. The learning algorithm

✧ Give example images $(x_i, y_i)_{i=1,...,n}$ where $y_i = 0,1$ for nonface and face examples.

✧ Initialize weights w_1, $i = 1/2m, 1/2l$ for $y_i = 0,1$, where m and l are the number of nonfaces and faces.

✧ For t=1,...,T:

  1. Normalize the weights $w_{t,i} = \dfrac{w_{t,i}}{\sum_{j=1}^{n} w_{t,j}}$.

  2. For each feature j, train a classifier h_j and evaluate the error respect to w_t.

$$\varepsilon_j = \sum_i w_i \, | h_j(x_i) - y_i |.$$

  3. Choose the classifier with the lowest error ε_t.

  4. Update the weights $w_{t+1,i} = w_t \beta_t^{1-e_i}$, if x_i is classified correctly $e_i = 0$.

    Otherwise $e_i = 1$, and $\beta_t = \dfrac{\varepsilon_t}{1-\varepsilon_t}$.

✧ The final strong classifier is

$$C(x) = \begin{cases} 1 & \sum_{t=1}^{T} \alpha_t h_t(x) \geq \dfrac{1}{2}\sum_{t=1}^{T} \alpha_t \\ 0 & otherwise \end{cases}$$

  where $\alpha_t = \log \dfrac{1}{\beta_t}$.

The final false positive rate (FPR) of the cascade is $F = \prod_{i=1}^{K} f_i$, where K is the number of states, and f_i is the FPR of the i th classifier on the samples that get through to it. The correct detection rate (CDR) is $D = \prod_{i=1}^{K} d_i$, where d_i is the detection rate of the i th classifier. As usual, a lower threshold yields higher detection rates and higher false positive rates.

The learning process of every weak classifier in the algorithm is almost according to the algorithm (Tab.1). But we choose the state numbers and the feature numbers for every state, to reduce the selecting time of the original Adaboost algorithm.

We choose 20 states to run the Adaboost. In every state, through adjusting the threshold, we control the CDR almost 100% and FPR less than 50%. For example, in the first state, 20 features can make CDR 98.34% and FPR 30.25%, while the threshold is set to 0.43. If the threshold is set to 0.38, the CDR is 100% and the FRP is 48.75%. Finally the algorithm gets 360 features to construct the strong classifier. The Receiver Operating Characteristic (ROC) curves of experiments are shown in section 4.2.

4   Experiments and Analysis

4.1   The Classifier Training

The CAS-PEAL face database [12] is adopted to train the classifier. This database contains 99,594 images of 1,040 individuals (595 males and 445 females) with varying pose, expression, accessory, and lighting. In experiment, we select 2400 face images from the database and cut the face departments. The cut images are all normalized to 56×56 pixels. The style of training samples influences the performance greatly. Through several times adjustment, the appropriate training samples are obtained. 4200 non-face samples are gathered from Web and also are normalized to 56×56 pixels. All of the training samples are transformed to face-pyramids. After three levels of decomposition, the last low frequency image has 49 features. Therefore the face-pyramid constructed has3577 dimensions. Then the final classifier with 360 features is constructed by using the Adaboost algorithm.

The test set consists of 4000 images (1200 faces and 2800 non-faces) and are normalized to 56×56 pixels. In order to demonstrate the efficiency of the extracted features, two comparative experiments are carried. In the first experiment, three different classifiers based on different features are trained. The GP has 4165 dimensions and The LP has 4116 dimensions. The detection performances on this test set are compared by ROC curves in Fig. 4. From the compare we can find that the face-pyramid based classifier has better performance than the low-frequency based and high-frequency based classifiers.

In the second experiment, the classifiers are constructed based on Haar features, Gabor features, and Contourlet features respectively. The comparable performance is shown in Fig 5. We can see that the classifier based face-pyramid has comparable performance with Haar classifier, and much better than Gabor classifier. At the same time, according to the training process of Adaboost algorithm, more features will result in more computing time. The standard five kinds of Haar features are 4,941,385 on an image of 56×56 pixels. The Gabor features obtained by decomposing at three scales and four directions are 37632. The total contourlet features we constructed are only 8281. So the training and testing time is greatly reduced in this method.
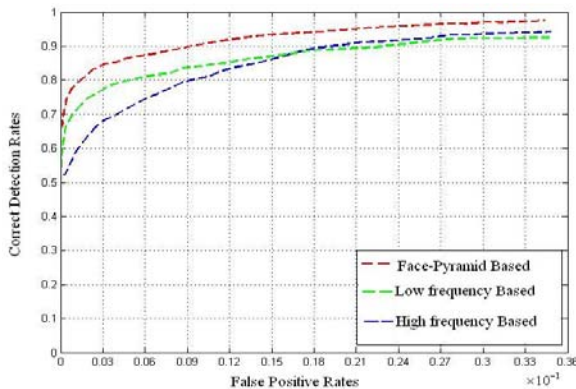
Fig. 4. The ROC curves for 360 features
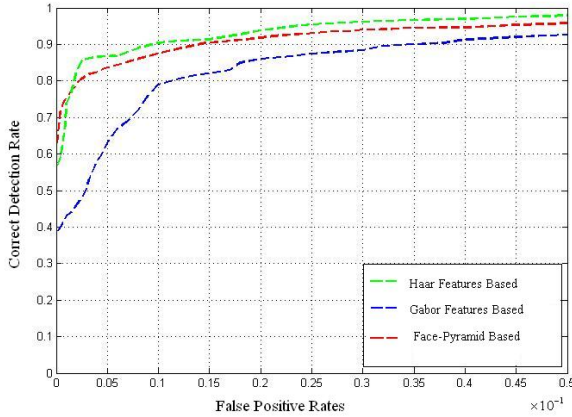
Fig. 5. The ROC curves

Using the constructed classifier with 360 features, we can quickly detect faces from giving images. Operating on 360 by 480 pixel images, faces are detected within 200 milliseconds in this method.

4.2   Detection Results

The detector is tested on the MIT+CMU frontal face test set and CAS-PEAL face database. We slide a window detector across the image horizontally and vertically to detect face. The window size can be adjusted according to the tested image size and every time at 20 scales. For example, according to a MIT+CMU face image, the window size is set as 32*32, 44*44,…,168*168. All of the candidate windows are classified to face or non-face by the trained classifier.

At last, multiple detections will usually occur around each face in a scanned image. We use mean-face template matching to select a single detection from overlapping detections. The mean-face template is the average image of all the training face samples. The face candidate window whose similarity is highest is the finally face detection.

The CAS-PEAL test set includes large-scale face images with varying pose, expression, accessory, and lighting. There is only one frontal face in each image. The pose images we chose have one face in multi-view but within $30°$. The new MIT+CMU test set consists of 180 images showing more than 500 upright faces. We select some images from this two face databases and the detection results are shown in Table 2.

Table 2. Detection results on some frontal face test sets

Face Database	Faces	Correct Detection Rate	False Positives
CAS-PEAL(Frontal)	5595	95.80%	58
CAS-PEAL(Pose)	6144	94.40%	86
MIT+CMU	463	92.87%	29
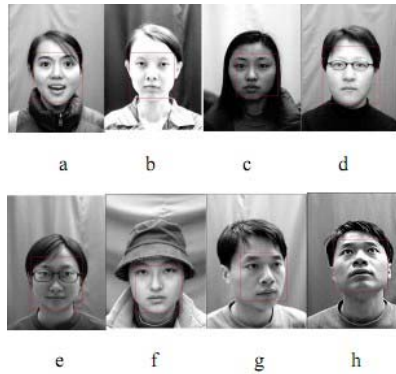
Fig. 6. Face detection results on some images from CAS-PEAL test set. a: expression. b, c: lighting. d, e, f: accessories. g, h: pose.

Fig. 7. Face detection results on some images from CMU-MIT test set

From Table 2 we can see the good performance of the method in many complex conditions. The detector successfully detects 11,160 images from 11,739 images with only 144 false positives. Some examples are shown in Fig. 6 and Fig. 7, and a square indicates a face region successfully detected.

5   Conclusions

In this paper, a novel face detection method based on contourlet features is introduced. The method proposes a new style of feature construction: face-pyramid. Due to the successfully implement of Adaboost, it is adopted to capture the most discriminative features from the formed face-pyramid. The constructed classifier can achieve high correct detection rates at low false positive rates. The CAS-PEAL face database is firstly applied to train the classifier. The experiment results show that the proposed

method is efficient at detecting frontal faces from clutter images. But the performance is not so good in some special situations, such as too low resolution, many ornaments and bad lighting conditions. In the future work, we are trying to study the detection problems on the variety of light conditions and face poses.

Acknowledgments

References

 1. Yang, M.H., Kriegman, D., Ahuja, N.: Detecting Faces in Images: A Survey. IEEE Tr. Pattern Analysis and Machine Intelligence 24, 34–58 (2002)
 2. Erik, H., Boon, K.L.: Face Detection: A Survey. Computer Vision and Image Understanding, 236–274 (2001)
 3. Viola, P., Jones, M.: Rapid Object Detection Using a Boosted Cascade of Simple Features. In: Conf. Computer Vision and Pattern Recognition, vol. 1, pp. 511–518 (2001)
 4. Viola, P., Jones, M.: Robust Real-Time Face Detection. In: Conf. International Journal of Computer Vision, pp. 137–154 (2004)
 5. Li, S.Z., Zhu, L., Zhang, Z.Q., Zhang, H.J.: Learning to detect multi-view faces in real-time. In: Proc. Development and Learning, Proceedings, pp. 172–177 (2002)
 6. Lienhart, R., Kuranov, A., Pisarevsky, V.: Empirical analysis of detection cascades of boosted classifiers for rapid object detection. In: Michaelis, B., Krell, G. (eds.) DAGM 2003. LNCS, vol. 2781, pp. 297–304. Springer, Heidelberg (2003)
 7. Viola, P.: Fast and Robust Classification using Asymmetric AdaBoost and a Detector Cascade. In: Advances in Neural Information Processing Systems, pp. 1311–1318 (2001)
 8. Randazzo, V., Usai, L.: An Improvement of AdaBoost for Face-Detection with Motion and Color Information. In: Conf. Image Analysis and Processing, ICIAP 2007 (2007)
 9. Lienhart, R., Maydt, J.: An extended set of Haar-like features for rapid object detection. In: International Conference on Proc. Image Processing, vol. 901, pp. 900–903 (2002)
10. Yan, Y.Y., Guo, Z.B.: Multi-view Face Detection Based on the Enhanced AdaBoost Using Walsh Features. In: Proc. Software Engineering, Artificial Intelligence, Networking, and Parallel/Distributed Computing, vol. 1, pp. 200–205 (2007)
11. Lin-Lin, H., Shimizu, A.: Classification-Based Face Detection Using Gabor Filter Features. In: Conf. Automatic Face and Gesture Recognition, pp. 397–402 (2004)
12. Chen, J., Shan, S.G.: Novel face detection method based on gabor features. In: Li, S.Z., Lai, J.-H., Tan, T., Feng, G.-C., Wang, Y. (eds.) SINOBIOMETRICS 2004. LNCS, vol. 3338, pp. 90–99. Springer, Heidelberg (2004)
13. Zhang, J.L., Zhang, Z.Y.: Face Recognition Based on Curvefaces. In: Third Intelnational Conference on Natural Computation (ICNC 2007), vol. 2, pp. 627–631 (2007)
14. Minh, N., Martin, V.: The Contourlet Transform: An Efficient Directional Multiresolution Image Representation. IEEE Transactions on Image Processing 14 (2005)
15. Duncan, D., Po, Y., Minh, N.D.: Directional Multiscale Modeling of Images Using the Contourlet Transform. IEEE Transactions on Image Processing 15, 1610–1620 (2006)
16. Li, B.B., Li, X.: A Multiscale and Multidirectional Image Denoising Algorithm Based on Contourlet Transform. In: 2006 IEEE Conf. on Intelligent Information Hiding and Multimedia Signal Processing (2006)
17. Dai, S.W.: Image Denoising Based on Complex Contourlet Transform. In: Conf. on Wavelet Analysis and Pattern Recognition (ICWAPR 2007), vol. 4, pp. 1742–1747 (2007)

Two-Dimensional Heteroscedastic Discriminant Analysis for Facial Gender Classification

Jun-Ying Gan, Si-Bin He, Zi-Lu Ying, and Lin-Bo Cai

School of Information, Wuyi University, Jiangmen, Guangdong, P.R.C. 529020
jygan@wyu.cn, junyinggan@163.com

Abstract. In this paper, a novel discriminant analysis named two-dimensional Heteroscedastic Discriminant Analysis (2DHDA) is presented, and used for gender classification. In 2DHDA, equal within-class covariance constraint is removed. Firstly, the criterion of 2DHDA is defined according to that of 2DLDA. Secondly, the criterion of 2DHDA, log and rearranging terms are taken, and then the optimal projection matrix is solved by gradient descent algorithm. Thirdly, face images are projected onto the optimal projection matrix, thus the 2DHDA features are extracted. Finally, Nearest Neighbor classifier is selected to perform gender classification. Experimental results show that higher recognition rate is obtained by way of 2DHDA compared with 2DLDA and HDA.

Keywords: Gender Classification, Two-dimensional Heteroscedastic Discriminant Analysis, Linear Discriminant Analysis.

1 Introduction

Gender classification using face images is a challenging work due to the similarity between male and female face images[1-2]. Thus, discriminant feature extraction is a key step to improve recognition rate. Linear Discriminant analysis (LDA)[3] is a well-known approach for feature extraction and dimensional reduction. However, it often encounters the Small Sample Size problem (S3 problem) when the number of samples is less than the dimensionality of samples. Then, two-dimensional Linear Discriminant analysis (2DLDA) [4] is proposed, in which discriminant features are extracted directly from 2-D images without a vectorization procedure, the computation cost is reduced and the S3 problem is overcome. However, in both of LDA and 2DLDA, it is assumed that the covariance matrices are equal for all sample classes[5-6]. Thus, when the within-class covariance of each sample class is significantly unequal, optimal performances can not be gained by LDA and 2DLDA.

Heteroscedastic Discriminant analysis (HDA) [7] is extended from LDA, in which equal within-class covariance constraint is removed. HDA can be viewed as a constrained Maximum likelihood (ML) projection [8], the constraint is given by the maximization of the projected between-class covariance volume and each class a single full covariance Gaussian model is satisfied. HDA is widely used in speech recognition and recognition rate is greatly increased than that of LDA [7-8]. But in 1D-based approaches, the transformation matrix is difficult to calculate due to high

dimensionality and extreme sparseness of the data. In this paper, based on 2DLDA and HDA, two-dimensional Heteroscedastic Discriminant analysis (2DHDA) is presented and used for gender classification. Firstly, the criterion of 2DHDA is defined, and log and rearranging terms are taken, then optimal projection matrix is solved by gradient descent algorithm. Secondly, face images are projected onto the optimal projection matrix, thus the discrimination features of face images are extracted. Finally, Nearest Neighbor classifier is selected to perform gender classification. Experimental results show the validity of 2DHDA method.

2   Presented Approach

Suppose there are C sample classes, represented by $A^1, A^2, A^3, \cdots, A^c$ respectively. The total number of samples is N and each class includes n samples, that is $nc = N$. $A_j^i \in \mathbf{R}^{m \times d}$ denotes the j th $(j = 1, 2, 3, \cdots, n)$ sample which belongs to the i th $(i = 1, 2, 3, \cdots, C)$ class. Thus, the mean of the i th sample class is $\overline{A^i} = \dfrac{1}{n} \sum_{j=1}^{n} A_j^i$, and the global mean of all samples is $\overline{A} = \dfrac{1}{N} \sum_{i=1}^{c} \sum_{j=1}^{n} A_j^i$.

2.1   2DLDA Approach

2DLDA's criterion is defined as

$$J(\theta_{\mathrm{2DLDA}}) = \frac{\left| \theta_{\mathrm{2DLDA}}^{\mathrm{T}} S_b \theta_{\mathrm{2DLDA}} \right|}{\left| \theta_{\mathrm{2DLDA}}^{\mathrm{T}} S_w \theta_{\mathrm{2DLDA}} \right|}, \tag{1}$$

where S_w is called within-class covariance matrix and S_b is called between-class covariance matrix of training samples, expressed respectively as

$$S_w = \frac{1}{N} \sum_{i=1}^{c} \sum_{j=1}^{n} \left(A_j^i - \overline{A^i} \right)^{\mathrm{T}} \left(A_j^i - \overline{A^i} \right). \tag{2}$$

$$S_b = \frac{1}{N} \sum_{i=1}^{c} n \left(\overline{A^i} - \overline{A} \right)^{\mathrm{T}} \left(\overline{A^i} - \overline{A} \right). \tag{3}$$

Transformation matrix θ_{2DLDA} is calculated by the solution of the eigenvalue and eigenvector problem of $S_b S_w^{-1}$.

2.2   2DHDA Approach

2DHDA is the heteroscedastic extension of 2DLDA. In 2DHDA, equal within-class covariance constraint is removed and the criterion is defined which maximizes the class discrimination in the projected subspace. The criterion of 2DHDA is defined as

$$J\left(\boldsymbol{\theta}_{\mathrm{2DHDA}}\right) = \prod_{i=1}^{C}\left(\frac{\left|\boldsymbol{\theta}_{\mathrm{2DHDA}}^{\mathrm{T}}\boldsymbol{S}_{b}\boldsymbol{\theta}_{\mathrm{2DHDA}}\right|}{\left|\boldsymbol{\theta}_{\mathrm{2DHDA}}^{\mathrm{T}}\boldsymbol{W}_{i}\boldsymbol{\theta}_{\mathrm{2DHDA}}\right|}\right)^{n} = \frac{\left|\boldsymbol{\theta}_{\mathrm{2DHDA}}^{\mathrm{T}}\boldsymbol{S}_{b}\boldsymbol{\theta}_{\mathrm{2DHDA}}\right|^{N}}{\prod_{i=1}^{C}\left|\boldsymbol{\theta}_{\mathrm{2DHDA}}^{\mathrm{T}}\boldsymbol{W}_{i}\boldsymbol{\theta}_{\mathrm{2DHDA}}\right|^{n}}, \tag{4}$$

where $\boldsymbol{W}_{i}=\dfrac{1}{n}\sum_{j=1}^{n}\left(\boldsymbol{A}_{j}^{i}-\overline{\boldsymbol{A}^{i}}\right)^{\mathrm{T}}\left(\boldsymbol{A}_{j}^{i}-\overline{\boldsymbol{A}^{i}}\right)$ denotes the covariance matrix of the ith sample

class. Thus, $\boldsymbol{S}_{w}=\dfrac{1}{C}\sum_{i=1}^{c}\boldsymbol{W}_{i}$. According to equation (1) and (4), if covariance matrix $\boldsymbol{W}_{i}$

of all sample classes is assumed equal, then $J\left(\boldsymbol{\theta}_{\mathrm{2DHDA}}\right)=\left(J\left(\boldsymbol{\theta}_{\mathrm{2DLDA}}\right)\right)^{N}$. However, $\boldsymbol{\theta}_{\mathrm{2DLDA}}$ is invariant to scale transformation of $J(\boldsymbol{\theta}_{\mathrm{2DLDA}})$, then $\boldsymbol{\theta}_{\mathrm{2DHDA}}=\boldsymbol{\theta}_{\mathrm{2DLDA}}$ is satisfied and 2DHDA is become 2DLDA. By taking log and rearranging terms, we get

$$H\left(\boldsymbol{\theta}_{\mathrm{2DHDA}}\right) \triangleq N\log\left|\boldsymbol{\theta}_{\mathrm{2DHDA}}^{\mathrm{T}}\boldsymbol{S}_{b}\boldsymbol{\theta}_{\mathrm{2DHDA}}\right| - \sum_{i=1}^{C}n\log\left|\boldsymbol{\theta}_{\mathrm{2DHDA}}^{\mathrm{T}}\boldsymbol{W}_{i}\boldsymbol{\theta}_{\mathrm{2DHDA}}\right|, \tag{5}$$

H has two useful properties of invariance[5]. For every nonsingular matrix $\varphi\in\mathbf{R}^{l\times l}$, $H(\varphi\boldsymbol{\theta}_{\mathrm{2DHDA}})=H(\boldsymbol{\theta}_{\mathrm{2DHDA}})$. This means that subsequent feature space transformations of the range of $\boldsymbol{\theta}_{\mathrm{2DHDA}}$ will not affect the value of the criterion. The second is that the criterion is invariant to row or column scalings of $\boldsymbol{\theta}_{\mathrm{2DHDA}}$ or eigenvalue scalings of $\boldsymbol{\theta}_{\mathrm{2DHDA}}\boldsymbol{\theta}_{\mathrm{2DHDA}}^{\mathrm{T}}$. Using matrix differentiation, the derivative of H is given by

$$\frac{\partial H\left(\boldsymbol{\theta}_{\mathrm{2DHDA}}\right)}{\partial\boldsymbol{\theta}_{\mathrm{2DHDA}}} = 2N\left(\boldsymbol{\theta}_{\mathrm{2DHDA}}^{\mathrm{T}}\boldsymbol{S}_{b}\boldsymbol{\theta}_{\mathrm{2DHDA}}\right)^{-1}\boldsymbol{\theta}_{\mathrm{2DHDA}}\boldsymbol{S}_{b} - \sum_{i=1}^{C}2n\left(\boldsymbol{\theta}_{\mathrm{2DHDA}}^{\mathrm{T}}\boldsymbol{W}_{i}\boldsymbol{\theta}_{\mathrm{2DHDA}}\right)^{-1}\boldsymbol{\theta}_{\mathrm{2DHDA}}\boldsymbol{W}_{i} \tag{6}$$

However, there is no close-form solution for $\dfrac{\partial H\left(\boldsymbol{\theta}_{\mathrm{2DHDA}}\right)}{\partial\boldsymbol{\theta}_{\mathrm{2DHDA}}}=0$. Instead, the gradient descent algorithm is used for the optimization of H and $\boldsymbol{\theta}_{\mathrm{2DHDA}}$ is solved. Usually, face images are projected onto the whole $\boldsymbol{\theta}_{\mathrm{2DHDA}}$, the most discriminant features are could not extracted, thus, former d column vectors of $\boldsymbol{\theta}_{\mathrm{2DHDA}}$ are selected as projection axes, then, the extracted features expressed as

$$\boldsymbol{Y}=\boldsymbol{A}\boldsymbol{\theta}_{\mathrm{2DHDA}}(:,1:d) \tag{7}$$

where $\boldsymbol{\theta}_{\mathrm{2DHDA}}(:,1:d)$ denotes the former d column vectors of $\boldsymbol{\theta}_{\mathrm{2DHDA}}$ and $\boldsymbol{Y}$ represents the extracted feature matrix of sample $\boldsymbol{A}$.

2.3 Nearest Neighbor Classifier

After a transformation of 2DHDA, Nearest Neighbor classifier is selected to perform gender classification. Suppose $\boldsymbol{Y}_{\mathrm{test}}$ denotes the feature matrix of an arbitrary testing sample $\boldsymbol{A}_{\mathrm{test}}$, $\boldsymbol{Y}_{j}^{i}$ denotes the feature matrix of training sample $\boldsymbol{A}_{j}^{i}$. Then, the distance between $\boldsymbol{Y}_{\mathrm{test}}$ and $\boldsymbol{Y}_{j}^{i}$ can be expressed as

$$D\left(Y_j^i, Y_{\text{test}}\right) = \sum_{k=1}^{d} \left\| Y_j^i\left(:,k\right) - Y_{\text{test}}\left(:,k\right) \right\|^2 \tag{8}$$

where $Y_j^i\left(:,k\right)$ and $Y_{\text{test}}\left(:,k\right)$ are the k th column vector of Y_j^i and Y_{test} respectively. $\left\| Y_j^i\left(:,k\right) - Y_{\text{test}}\left(:,k\right) \right\|^2$ is the Euclidean Distance between $Y_j^i\left(:,k\right)$ and $Y_{\text{test}}\left(:,k\right)$. If $D\left(Y_q^p, Y_{\text{test}}\right) = \min_{i,j} D\left(Y_j^i, Y_{\text{test}}\right)$ is satisfied, the testing sample A_{test} is classified to the p th class, where Y_q^p represents the feature matrix of training sample A_q^p , and p , q are constants.

3 Experiments

3.1 Experimental Objects

Experiments are based on Feret color face database[9] and face database from University of Essex, UK[10]. In Feret color face database, the images are varying in position, lighting and expression. We selected 10 male individuals, 10 female individuals with each individual 20 face images. Thus, there are 400 face images for experiments. In experiments, the images are chopped and resized to 100×90, then transformed to gray-scale images, as shown in Fig. 1.

Fig. 1. Face images in Feret color face database

Fig. 2. Face images in face database from university of Essex

In the face database from university of Essex, the images are with a resolution of 200×180, and with each individual 20 face images that vary in position, rotation, expression and lighting. We select 19 male and 19 female individuals, totally 760 face images for gender classification experiments. The original images are color images, we transformed them to gray-scale images and chopped them with a resolution of 80×70, as shown in Fig. 2.

3.2 Experimental Results and Analysis

In gender classification, there are only two classes, that are male and female respectively, thus in equation (4), $C = 2$. When the gradient descent algorithm is used for the optimization of H, θ_{2DLDA} is selected as the initial matrix of θ_{2DHDA} for iterations. Finally, Nearest Neighbor classifier is used for gender classification. The classification model is shown in Fig. 3.
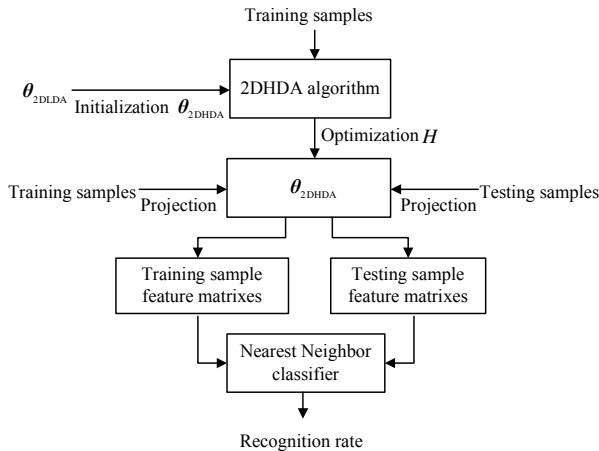
Fig. 3. Gender classification model

Based on Feret color face database, firstly, the former 5 individuals of male and female, totally 200 face images are selected as training samples, the remains as testing samples. Experimental results are shown in Fig. 4. Secondly, for male and female, the former 4, 6, 8 individuals, totally 160, 240, 320 face images are selected as training samples respectively, experimental results are listed in Table 1.

Fig. 4 illustrates that, when totally 200 images are selected as training samples, the highest recognition rate of 2DHDA is 85.00%, which is 4.5% higher than that of 2DLDA. Table 1 shows that when 320 images are selected as training samples, the recognition rate of 2DHDA is 88.75%. However, the recognition rate of 2DLDA is only 83.75% and that of HDA is only 80.00%. When 160 and 240 images are selected as training samples respectively, we can know that the recognition rates of 2DHDA are also higher than that of 2DLDA and HDA. In table 1, when HDA is used for gender classification, PCA is used as a pretreatment step for dimensional reduction.
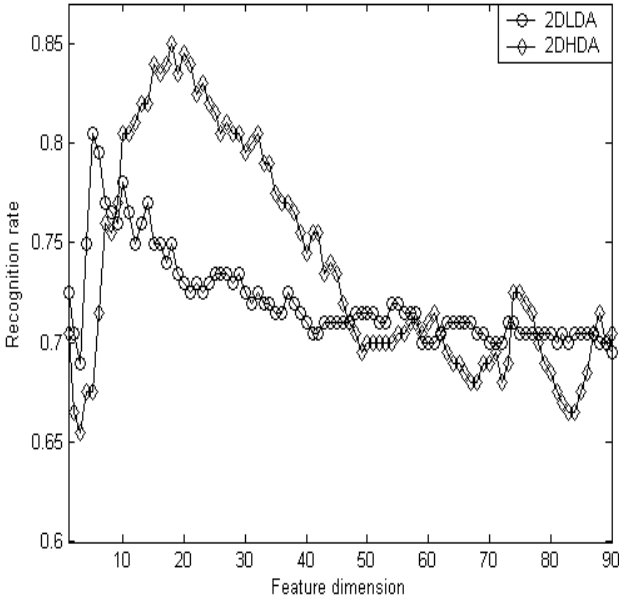
Fig. 4. Experimental results based on Feret face database when 200 images selected as training samples

Table 1. Experimental results based on Feret face database when different numbers of training samples are selected

Algorithms	160	240	320
2DHDA	84.58%	86.25%	88.75%
2DLDA	80.00%	85.62%	83.75%
HDA	76.66%	81.25%	80.00%

Based on face database from university of Essex, firstly, 20 individuals with 10 male and 10 female, totally 400 face images are selected as training samples and the remains as testing samples. When different numbers of feature dimension are selected, the results are shown in Fig. 5. Then, for male and female there are 9, 11 and 13 individuals for each class, totally 360, 440 and 520 face images are selected as training samples respectively, and the remains are selected as testing samples. Experimental results are listed in Table 2.

Fig. 5 demonstrates that, when 400 images are selected as training samples, the highest recognition rate of 2DHDA is 79.44%. The highest recognition rate of 2DLDA 70.89%, which is 8.55% lower than that of 2DHDA. Table 2 expresses that, the recognition rate of 2DHDA is higher than that of 2DLDA and HDA when 360, 440 and 520 images are selected as training samples respectively.
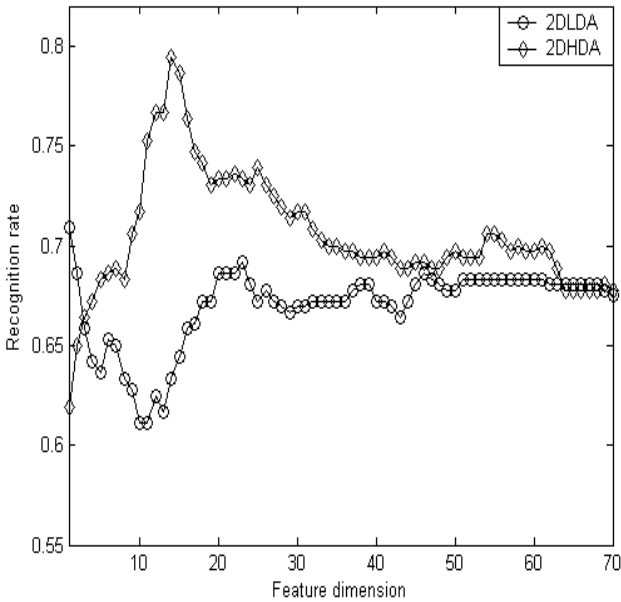
Fig. 5. Experimental results with 400 training samples based on face database from university of Essex

Table 2. Correct recognition rates based on face database from universityof Essex when different numbers of training samples are selected

Algorithms	360	440	520
2DHDA	74.50%	79.06%	77.50%
2DLDA	73.75%	70.31%	67.92%
HDA	65.75%	68.33%	69.16%

4   Conclusions and Future Work

In this paper, we presented the 2DHDA algorithm for gender classification using face images. Experimental results based on Feret color face database and face database from University of Essex show that 2DHDA is more effective than 2DLDA and HDA algorithms. However, when the images in a more complex background condition, how to improve the recognition rate need be further studied; the gradient descent algorithm is easy to run in local optimalization. Thus, in gender classification how to select the iteration factor and iteration number to gain the global optimal results need be further studied.

Acknowledgements

References

1. Castrillón-Santana, M., Vuong, Q.C.: An Analysis of Automatic Gender Classification. In: Rueda, L., Mery, D., Kittler, J. (eds.) CIARP 2007. LNCS, vol. 4756, pp. 271–280. Springer, Heidelberg (2007)
2. Wu, J., Smith, W.A.P., Hancock, E.R.: Facial Gender Classification Using Shape from Shading and Weighted Principal Geodesic Analysis. In: Campilho, A., Kamel, M.S. (eds.) ICIAR 2008. LNCS, vol. 5112, pp. 925–934. Springer, Heidelberg (2008)
3. Kyperountas, M., Tefas, A., Pitas, I.: Weighted Piecewise LDA for Solving the Small Sample Size Problem in Face Verification. IEEE Transactions on Neural Networks 18, 506–519 (2007)
4. Sanguansat, P., Asdornwised, W., Jitapunkul, S., Marukatat, S.: Two-Dimensional Discriminant Analysis of Principal Component Vectors for Face Recognition. IEICE Transactions on Information and Systems 7, 2164–2170 (2006)
5. Campbell, N.: Canonical Variate Analysis – A General Formulation. Australian Journal of Statistics 26, 86–96 (1984)
6. Hastie, T., Tibshirani, R.: Discriminant Annlysis by Gaussian Mixtures tech.rep. AT&T Bell Laboratories (1994)
7. Saon, G., Padmanabhan, M., Gopinath, R., Chen, S.: Maximum Likelihood Discriminant Feature Spaces. In: IEEE International Conference on Acoustics, Speech, and Signal Processing, vol. 2, pp. 1129–1132 (2000)
8. Gsles, M.J.F.: Maximum Likelihood Multiple Subspace Projection for Hidden Markomodels. IEEE Transactions on Speech and Audio Processing 10, 3–47 (2002)
9. The Feret face database, `http://face.nist.gov/colorferet/`
10. The face database, University of Essex, UK, `http://cswww.essex.ac.uk/mv/allfaces/faces94.html`

A Method of Plant Classification Based on Wavelet Transforms and Support Vector Machines

Jiandu Liu[1], Shanwen Zhang[2], and Shengli Deng[3]

[1] Missile Institute, Air-Force Engineering University, Sanyuan, 713800, P.R. China
[2] Hefei Institute of Intelligent Machines, Chinese Academy of Science,
Hefei, Anhui 230031, P.R. China
[3] Air-Force 93861 Army, Sanyuan, 713800, P.R. China
ljdzxj@sohu.com, wjdw716@163.com

Abstract. As one of the most important morphological taxonomy features, plant leaf with many strong points has significant influence on research. In this paper, we propose a novel method of plant classification from leaf image set based on wavelet transforms and support vector machines (SVMS). Firstly, the leaf images are converted into the time-frequency domain image by wavelet transforms without any further preprocessing such as image enhancement and texture thinning, and then feature extraction vector is conducted. Then the effectiveness of the proposed method is evaluated by the classification accuracy of SVM classifier. The experimental results about the data set with 300 leaf images show that the method has higher recognition rate and faster processing speed.

Keywords: Plant leaf image feature extraction; Wavelet transforms; Support vector machines.

1 Introduction

Compared with other methods, such as cell and molecule biology methods, classification based on leaf image is a better choice. Sampling leaves and photoing leaf images are low-cost and convenient. One can easily transfer the leaf image to a computer and a computer can extract features automatically in image processing techniques. Some systems employ descriptions used by botanists [5-8]. But it is not easy to extract and transfer those features to a computer automatically. This paper tries to prevent human interference in feature extraction. It is also a long discussed topic on how to extract or measure leaf features [9-15]. That makes the application of pattern recognition in this field a new challenge [1] [16]. According to Du et al. [1], data acquisition from living plant automatically by the computer has not been implemented. Several other approaches used their pre-defined features. Miao et al. proposed an evidence-theory-based rose classification [3] involving many features of roses. Gu et al. tried leaf recognition using skeleton segmentation by wavelet transform and Gaussian interpolation [17]. Wang et al. used a moving median center (MMC) hypersphere classifier [18]. Similar method was proposed by Du et al. [1]. Their another paper proposed a

modified dynamic programming algorithm for leaf shape matching [19]. Ye et al. compared the similarity between features to classify plants [2].

It is important and necessary to mine this data to select only genes contributing to the problem domain and to filter irrelevant data. Feature selection is the problem of identifying such genes or features with significant information content to improve the generalization performance and inference of classification models by overcoming the 'curse of dimensionality' which causes the risk of "over fitting". In this paper, we propose a method of feature selection based on wavelet power spectrum which is found fit for a wide range of data sets and also works with smaller number of samples. It can be used in conjunction with other classification methods. The algorithm is very simple and requires comparatively less time to be executed. Unlike most of the other methods, it is relatively a very simple algorithm.

The Wavelet Transform [19,20] overcomes some of these limitations; it can provide a constant-Q analysis of a given signal by projection onto a set of basis functions that are scale variant with frequency. Each wavelet is a shifted scaled version of an original or mother wavelet. These families are usually orthogonal to one another, important since this yields computational efficiency and ease of numerical implementation. Other factors influencing the choice of Wavelet Transforms over conventional methods include their ability to capture localised features. Also, developments aimed at generalisation such as the Best-Basis Paradigm of Coifman and Wickerhauser make for more flexible and useful representations.

Support Vector Machine (SVM) with kernels is a widely used technique for classification [21,22]. Recently, distance-based algorithms with variant data-dependent distance metrics are proposed. However, many classification methods fail for classification problems when the number of data points is much less than the data-dimension. It is required for those data sets to reduce the data dimension so that the classification methods can be well applied on the reduced points.

The rest of the paper is organized as follows. Section 2 and Section 3 introduce wavelet packet transforms and support Vector Machines (SVMs), respectively. Section 4 proposes a method of plant classification. Experimental results and analysis are illustrated in Section 5. A conclusion is given in Section 6.

2   Wavelet Packet Transforms

Wavelet transforms (WT), enabling the time-frequency representation of the instrumental signals, gain more and more attention in different fields of microarray technology. The main advantages offered by WT and explored until now are associated with signal compression and denoising. In the case of the wavelet packet transform (WPT), multiresolution analysis, is now regard as standard tools by researchers in many applications. It is necessary to propose opportunities for exchanging between practitioners and researchers about wavelet. The decomposition of a signal, the length of which equals an integer power of two $(n = 2^p)$ on an orthonormal wavelet basis results in the same number of coefficients as the original signal. WPT is an extension of the DWT to the full binary tree. Any set of n coefficients, which together cover the entire time-frequency range, forms a complete orthonormal basis and can be used to reconstruct the original signal. Usually, there are only few decomposition coefficients

useful. Over-complete representation of the signal in the time-frequency domain as obtained by WPT gives us the possibility of adaptation i.e. of choosing among many representations the one best suited for our purpose. There are many possible criteria of the 'best basis' selection. According to one of them, the best basis is the one with the minimum entropy of the distribution of the wavelet transform coefficient [19,20].

In the discrete wavelet packet transform (DWPT), both the scaling and wavelet co-efficients are subject to the high-pass and low-pass filtering when computing the next level scaling and wavelet coefficients. With the standard transforms, scaling coefficients identify the frequency band $[0,1/2^{J+1}]$, with J the coarsest level, while wavelet coefficients at level j describe the frequency band $[1/2^{j+1},1/2^{j}]$. The DWPT partition the whole frequency band, $[0,1/2]$, into equal length frequency bands. DWPT induces a finer partition of the frequency space. It is a collection of functions $2^{-j/2}u_n(2^{-j}t-k), j,k \in Z, n \in Z_+$ generated from the following sequence of functions:

$$u_{2n}(t)=\sqrt{2}\sum_{k\in z}h_k u_n(2t-k)$$
$$u_{2n+1}(t)=\sqrt{2}\sum_{k\in z}g_k u_n(2t-k)$$

(1)

where h and g are the quadrature mirror filters, $\sum_{n\in Z}h_{n-2k}h_{n-2l}=\delta_{kl}, \sum_{n\in Z}h_n=\sqrt{2}$, $g_k=(-1)^k h_{1-k}, k \in Z$, $u_0(t)$ and $u_1(t)$ are the scaling function and basic wavelet, respectively. The wavelet packet $\{2^{-j/2}u_n(2^{-j}t-k), j,k \in Z, n \in Z_+\}$ is a localized function of unit energy with scale 2^j, translation $2^j k$, and an oscillation parameter of n.

For a discrete signal, the decomposition coefficients of wavelet packets can be computed iteratively by

$$x_{2n,j+1}^k=\sum_l h_{l-2k}x_{n,j}^l \quad , \quad x_{2n+1,j+1}^k=\sum_l g_{l-2k}x_{n,j}^l$$

(2)

where $X_{n,j}=\{x_{n,j}^k\}$ is the decomposition coefficient sequence of the nth node at level j of the wavelet packet tree. Without loss of generality, $j=1,2,3,4$.

The original signal can be reconstructed iteratively by

$$x_{n,j}^l=\sum_k h_{l-2k}x_{2n,j+1}^k+\sum_k g_{l-2k}x_{2n+1,j+1}^k$$

(3)

The wavelet packet functions as well as the corresponding decomposition coefficients can be organized as a binary tree as shown in Fig.1. Each node corresponds to a frequency band. The leaf nodes of any connected subtree that has the same root node as the full tree form an orthonormal basis and can represent a signal of finite energy completely.
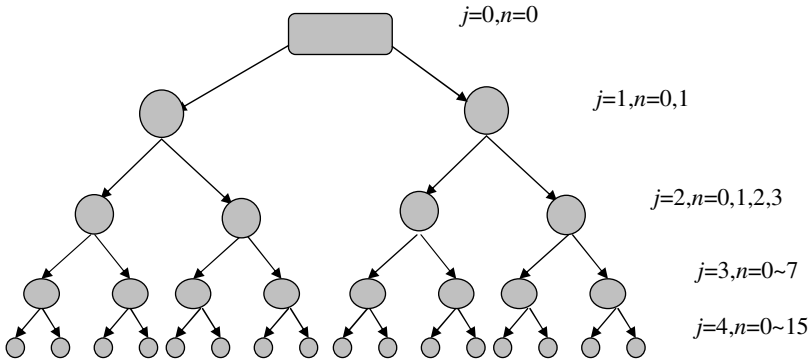
Fig. 1. Wavelet packet tree

As an example, the wavelet packet functions of the filled nodes form an orthonormal basis. The decomposition coefficients are organized as the same tree, in which the nth node at level j includes the wavelet packet function $\{2^{-j/2}u_n(2^{-j}t-k), 0 \le k < 2^{-j}N\}$.

3 Support Vector Machines (SVM)

SVM is a relatively new type of machine learning model, originally introduced by Vapnik and co-workers, and successively extended by a number of other researchers. This model, which is of remarkably robust performance with respect to sparse and noisy data, is becoming into the system of choice in a number of applications from text categorization to protein function prediction.

When used for classification, SVM can separate a given set of binary labeled training data with a hyper-plane that is maximally distant from then (the maximal margin hype-plane). For the cases in which no linear separation is possible, they can work in combination with the technique of 'kernels', which automatically realizes a nonlinear mapping to a feature space, Generally, the hyper-plane found by the SVM in feature space corresponds to a nonlinear decision boundary in the original space.

Without loss of generality, let the i_{th} input sample $\beta^i = (\beta^i_1, ... \beta^i_n)$ be the realization of the random vector β. And this input sample is labeled by the random variable $\gamma \in \{-1,1\}$. Assume the $\varnothing : U \Rightarrow V (U \subseteq R^p, V \subseteq R^q)$ is a mapping from the input space U to a feature space V, and that we have a set of samples θ of m samples labeled data points: $\theta = \{(\beta^1, \gamma^1), ..., (\beta^m, \gamma^m)\}$. The SVM learning algorithm is to find a hyper-plane (ω, b) such that the quantity:

$$\chi = \min_i \gamma^i \{< \omega, \phi(\beta^i) > -b\} \tag{4}$$

is maximized, where $< \bullet >$ denote an inner product, the vector ω has the same dimensionality as V, $\| \omega \|_2$ is held as a constant, b is a real number and χ is called the margin. The quantity $(< \omega, \phi(\beta^i) > -b)$ corresponds to the distance between the points β^i and the decision boundary. When multiplied by the label γ^i, it gives a positive value for all correct classifications and a negative value for all the incorrect ones. The minimum of this quantity over all the data is positive if the data is linearly separable, which is called as the margin. Given a new data sample β to be classified, a label is assigned according to its relationship to the decision boundary, and the corresponding decision function is:

$$f(\beta) = sign(< \omega, \phi(\beta) > -b) \tag{5}$$

4   A Method of Plant Classification

One of the most important aspects of any pattern detection technique is extracting inherent features in a dataset. Consequently any texture classification technique would only be as good as the features extracted from the pattern being analyzed. The process of feature extraction is compounded in difficulty by the dimensionality of the feature space. The complexity of features selection is leaded to a great extent by employing an adaptive WPT and using it to extract the most discriminate of subbands from the complete transform. The following statistical features of the wavelet packet decomposition are computed for each node of wavelet packet tree. They are to be useful for our application.

- *Mean:* $M_{n,j} = \dfrac{1}{K_{n,j}} \sum\limits_{k=0}^{K_{n,j}-1} x_{n,j}^k$;

- *Energy*: $E_{n,j} = \dfrac{1}{K_{n,j}} \sum\limits_{k=0}^{K_{n,j}-1} [x_{n,j}^k]^2$;

- *Entropy*: $H_{n,j} = -\sum\limits_{k=0}^{K_{n,j}-1} \left| x_{n,j}^k \right| \cdot \log_2 \left| x_{n,j}^k \right|$;

where, $K_{n,j}$ is the length of decomposition coefficient sequence $x_{n,j}$.

The question about plant classification has complicated nature of incertitude. In many conditions, we always get cursory information in the beginning of investigation. So the key of the problem is how to engender the final rule using useful feature and original data. To deal with the decision information using wavelet transforms gets features. We regard mean, energy and entropy as classification features. Then SVM classifier is applied.

5   Experimental Results and Analysis

The shape feature is one of the most important features for characterizing an object, which is commonly used in object recognition, matching and registration. In addition,

the shape matching and recognition is also an important part of machine intelligence that is useful for both decision-making and data processing. More importantly, the recognition based on shape feature is also a central problem in those fields such as pattern recognition, image technology and computer vision, etc., which have received considerable attention recent years.

During the experiment, we select 15 kinds of plants with 100 leaf images for each plant, as shown in Fig.2, we regard mean, energy and entropy as classification features, and employ radial basis function (RBF) kernel in SVM learning algorithm to validate the selected features. The covering decision systems can be divided into consistent covering decision systems and inconsistent decision systems.
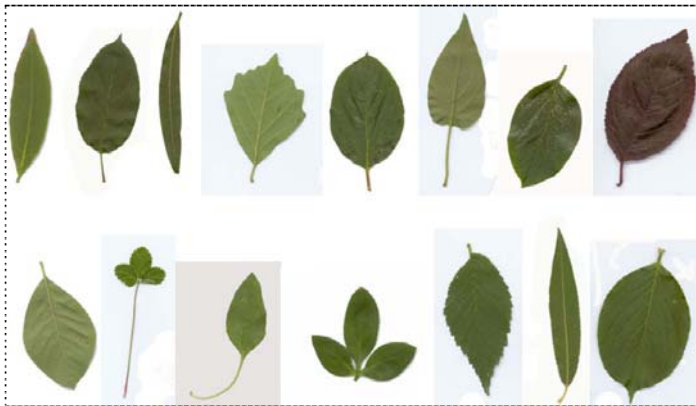
Fig. 2. Original 15 plant leaf images

All the plant leaf images are captured with the same plant leaf image capture device. Plant leaf image samples from the same palm may be with a little rotation and shift. The size of all the plant leaf image images is 240×240. The following experiments are designed for testing the accuracy and efficiency of the proposed method. The data collection process involves four steps:

- Find 15 plants, each has 100 leaf images.
- Compute the wavelet transforms of the leaf images and exact the features with mean, energy and entropy
- Classify leaf image using SVM.

We extract the classifying feature from the coefficients of leaf image. We try the following two kinds of support vector machines: 1) Linear SVM (no kernel); 2) Radial basis function SVM (RBF kernel). Where, we select wavelet function Daubechies 5. The results can be seen from table1. In order to compare, the plant leaf image classification based on other methods are given in table 1.

Table 1. Results of plant leaf image identification rate

Method	PCA+ SVM	ICA + SVM	Wavelet transforms +SVM
Identification rate	92%	94%	95%
Response time (s)	12	14	8

The response time is recorded to evaluate the efficiency of the proposed method. From the table1, we can see the timing response is shorter using the method which is presented in this paper.

6   Conclusion

This paper introduces a feature extraction approach for plant leaf recognition. The classifier SVM is adopted for it has fast speed on training and simple structure. The features of plant leaf are extracted and processed by wavelet transforms to form the input vector of SVM. The experimental results indicate that our algorithm is workable with a recognition rate greater than 95% on 15 kinds of plant leaf images. Compared with other methods, this algorithm is fast in execution, efficient in recognition and easy in implementation. Future work is under consideration to improve it.

Acknowledgment

References

1. Du, J.X., Wang, X.F., Zhang, G.J.: Leaf shape based plant species recognition. Applied Mathematics and Computation 185 (2007)
2. Ye, Y., Chen, C., Li, C.T., Fu, H., Chi, Z.: A computerized plant species recognition system. In: Proceedings of 2004 International Symposium on Intelligent Multimedia, Video and Speech Processing, Hong Kong (October 2004)
3. Miao, Z., Gandelin, M.H., Yuan, B.: An oopr-based rose variety recognition system. Engineering Applications of Artificial Intelligence 19 (2006)
4. de Oliveira, P., Falvo, R.M., Pdua, J.G., Bernacci, L.C.M., Vieira, L.C.G., Oliveira, C.X., Bruno, O.M.: Leaf shape analysis using the multiscale minkowski fractal dimension, a new morphometric method: a study with passifliora (passifloraceae). Canada Journal of Botany 83 (2005)
5. Dallwitz, M.J.: A general system for coding taxonomic descriptions. Taxon 29 (1980)
6. Fu, H., Chi, Z., Feng, D., Song, J.: Machine learning techniques for ontology-based leaf classification. In: IEEE 2004 8th International Conference on Control, Automation, Robotics and Vision, Kunming, China (2004)
7. Warren, D.: Automated leaf shape description for variety testing in chrysanthemums. In: Proceedings of IEE 6th International Conference Image Processing and Its Applications (1997)
8. Brendel, T., Schwanke, J., Jensch, P., Megnet, R.: Knowledgebased object recognition for different morphological classes of plants. In: Proceedings of SPIE, vol. 2345 (1995)
9. Li, Y., Zhu, Q., Cao, Y., Wang, C.: A leaf vein extraction method based on snakes technique. In: Proceedings of IEEE International Conference on Neural Networks and Brain (2005)

10. Fu, H., Chi, Z.: Combined thresholding and neural network approach for vein pattern extraction from leaf images. IEE Proceedings-Vision, Image and Signal Processing 153(6) (2006)
11. Nam, Y., Hwang, E.-J., Byeon, K.: ELIS: An efficient leaf image retrieval system. In: Singh, S., Singh, M., Apte, C., Perner, P. (eds.) ICAPR 2005. LNCS, vol. 3687, pp. 589–597. Springer, Heidelberg (2005)
12. Fu, H., Chi, Z.: A two-stage approach for leaf vein extraction. In: Proceedings of IEEE International Conference on Neural Networks and Signal Processing, Nanjing, China (2003)
13. Wang, Z., Chi, Z., Feng, D.: Shape based leaf image retrieval. IEE Proceedings-Vision, Image and Signal Processing 150(1) (2003)
14. Qi, H., Yang, J.G.: Sawtooth feature extraction of leaf edge based on support vector machine. in Proceedings of the Second International Conference on Machine Learning and Cybernetics.(2003)
15. Hong, S.M., Simpson, B., Baranoski, G.V.G.: Interactive venationbased leaf shape modeling. Computer Animation and Virtual Worlds. 16 (2005)
16. Gouveia, F., Filipe, V., Reis, M., Couto, C., Bulas-Cruz, J.: Biometry: the characterisation of chestnut-tree leaves using computer vision. In: Proceedings of IEEE International Symposium on Industrial Electronics, Guimaraes, Portugal (1997)
17. Gu, X., Du, J.-X., Wang, X.-F.: Leaf recognition based on the combination of wavelet transform and gaussian interpolation. In: Huang, D.-S., Zhang, X.-P., Huang, G.-B. (eds.) ICIC 2005. LNCS, vol. 3644, pp. 253–262. Springer, Heidelberg (2005)
18. Wang, X.-F., Du, J.-X., Zhang, G.-J.: Recognition of leaf images based on shape features using a hypersphere classifier. In: Huang, D.-S., Zhang, X.-P., Huang, G.-B. (eds.) ICIC 2005. LNCS, vol. 3644, pp. 87–96. Springer, Heidelberg (2005)
19. Mallat, S.A.: Theory for multiresolution signal decomposition: The wavelet representation. IEEE Trans. Pattern Anal. Machine Intell. 31, 674–693 (1989)
20. Kadambe, S., Boudreaux-Bartels, G.F.: Application of the Wavelet Transform for Pitch Detection of Speech Signals. IEEE Transactions on Information Theory 32, 712–718 (1992)
21. Mukherjee, S., Rifkin, R.: Support Vector Machine Classification of Microarray Data, CBCL Paper 182, Artificial Intelligence Lab. Memo. 1676. MIT, Cambridge (1999)
22. Brown, M.P., Grundy, W.N., Lin, D., Cristianini, N., Sugnet, C., Ares, M.J., Haussler, D.: Support vector machine classification of microarray gene expression data. Proc. Natl. Acad. Sci. USA 97, 262–267 (2000)

Locality Preserving Fisher Discriminant Analysis for Face Recognition

Xu Zhao and Xiaoyan Tian

The Laboratory of Computer Software and Theory,
Beijing University of Technology
PingLeYuan 100, ChaoYang District, Beijing
P.R. China
zhaoxu166@gmail.com, tianxiaoyan@emails.bjut.edu.cn

Abstract. Dimensionality reduction is a key technology for face recognition. In this paper, we propose a novel method, called Locality Preserving Fisher Discriminant Analysis (LPFDA), which extends the original Fisher Discriminant Analysis by preserving the locality structure of the data. LPFDA can get a subspace projection matrix by solving a generalized eigenvalue problem. Several experiments are conducted to demonstrate the effectiveness and robustness of our method.

Keywords: Fisher Discriminant Criterion; Locality Preserving Projection; Face Recognition.

1 Introduction

Face recognition is a key branch of computer vision; also it can be employed in many fields, such as human-computer interface, security validation, etc. So, over the past few years, face recognition has received lots of attention by researchers, and a great deal of techniques has been developed. Appearance-based method is one main approach among them, which involves two main steps: one is to represent the face image in a feature space, and the other is how to classify the images in the feature space. In order to deal with the first issue, several classic techniques have been developed, which can be categorized into two classes, namely, the linear methods and the nonlinear methods. Principal Component Analysis (PCA) [1] and Fisher Discriminant Analysis (FDA) [2] are two classic linear features extraction methods based on Euclidean distance. PCA attempts to preserve the global structure of the data points' distribution, and FDA attempts to make the samples as separable as possible in the low dimensional space. But in the real world, there is no trusty evidence that the image feature space is Euclidean, so it is more reasonable to postulate that the image feature space is a manifold embedded in the high dimensional space, and the traditional linear methods are difficult to find the true structure of the data points. Recently, some nonlinear algorithms have been developed, such as Locally Linear Embedding (LLE) [3] and [4], ISOMAP [5], Locality Preserving Projections [6] and [7] Neighborhood Preserving Embedding [8], Tangent Distance Preserving Mapping

[9] etc. However, all these methods are designed to preserve the locality of the data in the lower dimensional space rather than discriminant ability. So, the results of the projection will result in the overlap of data form different class. Developing an effective algorithm which holds both the linear character and the nonlinear character, and also possesses the discriminant ability is an important issue.

In order to deal with this problem, we propose a novel algorithm named Locality Preserving Fisher Discriminant Analysis (LPFDA), which extends the original Fisher Discriminant Analysis by preserving the locality structure of the data. LPFDA inherits the discriminant ability of FDA, and the locality preserving ability makes LPFDA more reliable than FDA. Also we can consider LPFDA as a combination of FDA with Locality Preserving Projections (LPP) at a certain degree.

The rest of the paper is organized as follows: in section 2, we briefly review Fisher Discriminant Analysis and Locality Preserving Projections, and then in section 3, we elaborate our proposed method Locality Preserving Fisher Discriminant Analysis (LPFDA). Section 4 presents the experiment results. Finally in section 5, the conclusion is given.

2 Paper Preparation

2.1 Review of FDA

Let $X = [x_1, x_2 \ldots x_n]$ be a set of vectors, where x_i (i =1, 2,…, n) $\in$ R^h, and each x_i belongs to one of the C classes. Let n_i (i=1,2,…,c) be the number of the ith class data points, and we have n=n_1+n_2+…+n_c. Also we define $m_i = \dfrac{1}{n_i} \sum\limits_{x \in class\ i} x$ is the mean of the data points belonging to the ith class, and $m = \dfrac{1}{n} \sum\limits_{x} x$ be the mean of all the data points.

FDA wants to project X into a low dimensional subspace in which the ratio of between-class distribution and within-class distribution can be maximized. So the objective function of FDA is defined as:

$$\max \ \frac{p^T S_b p}{p^T S_w p} \tag{1}$$

S_w is the within-class scatter matrix, $S_w = \sum\limits_{i=1}^{c} S_i$, where

$$S_i = \sum_{x \in class\ i} (x - m_i)(x - m_i)^T \ .$$

S_b is the between-class scatter matrix, $S_b = \sum\limits_{i=1}^{c} n_i (m_i - m)(m_i - m)^T$

(1) is called the generalized rayleigh quotient and can be solved by solving a generalized eigenvalue problem:

$$S_b p = \lambda S_w p \qquad (2)$$

By projecting the original data set to the subspace constructed by the solutions of (2), FDA can make the ratio of between-class distribution and within-class distribution can be maximized.

2.2 Review of LPP

LPP is a linear approximation of the nonlinear Laplacian Eigenmap [10] for learning a locality preserving subspace which preserves the intrinsic geometry of the data and local structure. A neighborhood relationship graph G is built to depict the local structure of the data points.

Let $X = [x_1, x_2 \ldots x_n]$ be a set of vectors, where x_i (i =1, 2,…, n) $\in R^h$. LPP seeks a matrix P to project X into a low dimensional space in which the local structure of X can be preserved.

The objective function of LPP is defined as:

$$\min \sum_i \sum_j \left\| y_i - y_j \right\|^2 w_{ij} \qquad (3)$$

Where w_{ij} is some certain measures weight which depicts the nearness relationship between two data points.

There are two common ways to compute w_{ij}:
Heat Kernel:

$$w_{ij} = \begin{cases} \exp(-\left\| x_i - x_j \right\|^2 /t) & \text{if } x_i(x_j) \text{ is among k nearest neighbors of } x_j(x_i) \\ 0 & \text{otherwise} \end{cases}$$

Constant Weight:

$$w_{ij} = \begin{cases} 1 & \text{if } x_i(x_j) \text{ is among k nearest neighbors of } x_j(x_i) \\ 0 & \text{otherwise} \end{cases}$$

The weight w_{ij} define a matrix W, whose entry w_{ij} is nonzero only when x_i and x_j are adjacent. Also notice that the entries of W are nonnegative and that W is sparse and symmetric.

There is also an imposed constraint to (1), namely, $y^T D y=1$, where D is a diagonal matrix, with $D_{ii} = \sum_j w_{ij}$.

Finally, the minimization problem reduces to the following form:

$$\min P^T X L X^T P$$
$$P^T X D X^T P = I \qquad (4)$$

Where L=D-W, is the Laplacian [11] of the graph G constructed by LPP.

In order to get the optimal projection matrix P_{LPP}, we just need to get the eigenvectors corresponding to the minimum eigenvalues of the generalized eigenvalue problem:

$$XLX^T P = \lambda XDX^T P \qquad (5)$$

When we get P_{LPP}, we can obtain the projection results of X easily by $Y = P_{LPP}^T X$.

3   Locality Preserving Fisher Discriminant Analysis (LPFDA)

3.1   Justification

Since LPFDA is designed to hold both the linear character and the nonlinear character, and also possess the discriminant ability, we can innovate a heuristic object function by combining the FDA and the LPP:

$$\max \frac{p^T S_b p}{p^T S_w p} - p^T XLX^T p \qquad (6)$$

Where S_b is the between-class scatter matrix, S_w is the within-class scatter matrix. X is the image set, and p is the projection direction. The definition of matrix L is the same as in section 2.2. Since LPP wants to find a projection direction w to make $p^T XLX^T p$ as small as possible, we can instead here choose $- p^T XLX^T p$ and make it as large as possible in the low dimension space.

Alternatively, we can reformulate (6) as:

$$\max p^T S_b p - p^T XLX^T p$$

$$s.t. \quad p^T S_w p = 1 \qquad (7)$$

The constrained maximization problem (7) can be solved by Lagrange multipliers:

$$L(w) = p^T S_b p - p^T XLX^T p + \mu(p^T S_w p - 1)$$
$$= p^T (S_b - XLX^T) p + \mu(p^T S_w p - 1)$$

Let the partial derivative with respect to p to zero we can get:

$$(S_b - XLX^T) p + \mu S_w p = 0 \qquad (8)$$
$$\xrightarrow{\lambda = -\mu} (S_b - XLX^T) p = \lambda S_w p$$

Let $A = S_b - XLX^T$, $B = S_w$, (8) can be rewritten as:

$$Ap = \lambda Bp \qquad (9)$$

If matrix B is invertible, we can reduce (9) to the common eigenvalue problem:

$$B^{-1}Ap = \lambda p \qquad (10)$$

In real operation, we can add a diagonal matrix with small entry value, such as 0.01, to the matrix B to ensure it has full rank.

The optimal solution vector p of (7) is the eigenvector corresponding to the largest eigenvalue of $B^{-1}A$. Usually one projection direction is not enough, so we use the eigenvectors corresponding to the largest d eigenvalues of $B^{-1}A$ (d is the dimension of the subspace) to be the projection directions, and they form a projection matrix P_{h*d}. Here h is the dimension of the original space.

When P_{h*d} is gotten, the data points in the original space can be mapped into a subspace by

$$Y=P_{h*d}{}^{T}X \qquad (11)$$

And for a new test sample x, it can be mapped to the subspace by

$$y=P_{h*d}{}^{T}x \qquad (12)$$

3.2 The Algorithm

The concrete algorithmic procedure is formally stated as follows:

Input: $X = [x_1, x_2 \dots x_n]$, x_i (i =1, 2,…, n) $\in R^h$
Output: $Y = [y_1, y_2 \dots y_n]$, x_i (i =1, 2,…, n) $\in R^d$ (d<<h)
Step1: Compute the between-class scatter matrix S_b and the within-class scatter matrix S_w;
Step2: Compute the matrix L
    a.   Construct the neighborhood relationship graph G, and compute the weigh w_{ij} of each edge of G, which equals to compute the weigh matrix W;
    b.   Let L=D-W, where $D_{ii} = \sum_j w_{ij}$

Step3: Solve the generalized eigenvalue problem $(S_b - XLX^T)p = \lambda S_w p$, and get projection matrix P_{h*d}, which is constituted of the eigenvectors corresponding to the d largest eigenvalues of matrix $S_w^{-1}(S_b-XLX^T)$.
Step4: Project X to Y by $Y=P_{h*d}{}^{T}X$.

4   Experiments and Discussions

In this section, we conduct several experiments on different datasets to demonstrate the effective and the robustness of our proposed method Locality Preserving Fisher Discriminant Analysis (LPFDA). Remember that in all the following experiments the parameter k used to find the nearest neighborhoods in LPP is set to 10. Also when constructing the graph G in LPFDA, we put an edge between two data points just when they belong to the same class.

4.1 ORL

In this experiment, we use a famous face database ORL [12]. ORL contains 40 different subjects, and each subject has ten different images. The images include variation in face expression (smile or not, open / closed eyes) and pose. Fig.1 illustrates two sample subjects of the ORL database along with variations in facial expression and pose.

Fig. 1. Sample faces images from the ORL database along with variations in facial expression and pose

We test LPFDA against PCA, FDA and LPP to demonstrate the predominance of LPFDA. We form the training set by randomly selecting 5 images per individual with labels, and the rest of the database was considered to be the testing set. Nearest neighborhood classifier is employed in the experiments. We conduct the experiments for 10 times, and report the average results here. The results are given in Fig.2. The horizontal axis represents the dimension of the subspace and the vertical axis stands for the recognition rate.
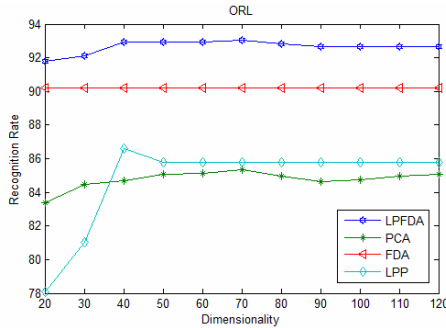
Fig. 2. The recognition rates versus the subspace dimension on ORL database

It can be seen that LPFDA outperforms all the 3 methods a lot. Although FDA performs well, however, LPFDA exceeds it nearly 3 percent. Here, LPP has the same performance when the dimension is larger than 50, which is because we run PCA to preserve 90 percent principal component prior to conduct LPP, and the subspace dimension obtained by PCA is smaller than 50. However, LPP' performance is far worse than LPFDA and FDA, which is because LPP do not consider explicitly the class labels of the data points, so when comparing with LPFDA and LDA, it eats their dust. Another point we should pay attention to is that LPFDA gets the second highest recognition rate (92.95%, which is less than the first highest recognition rate for 0.1 percent) at the dimension 40 which equals to the number of the classes. This indicates that LPFDA inherits the character of FDA.

Like Fisherface, we display the first 10 "eigenfaces" obtained by LPFDA in Fig.3, for simplicity, we call them LPFisherfaces. In order to compare, we also show Eigenfaces, Fisherfaces and Laplacianfaces along with LPFisherfaces.
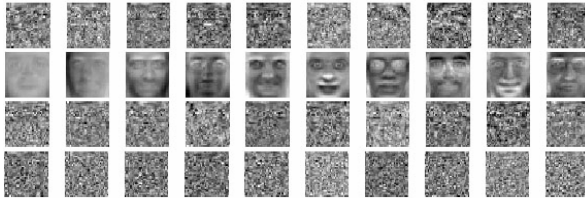
Fig. 3. From top to bottom: the first 10 LPFisherfaces, Eigenfaces, Fisherfaces and Laplacianfaces obtained from ORL data base

All the images in the original space are projected into the subspace spanned by the LPFisherfaces.

We also list the best recognition rate achieved by each method along with the corresponding subspace dimension d in Table 1.

Table 1. The best recognition rate achieved by each method along with the corresponding subspace dimension d

	Subspace Dimension d	Recognition Rate (%)
LPFDA	70	93.05
PCA	70	85.35
FDA	39	90.2
LPP	40	86.6

4.2 YALE

YALE face database [13] contains 165 grayscale images of 15 individuals. There are 11 images per subject, one per different facial expression or configuration: center-light, w/glasses, happy, surprised, and wink, etc. Fig.4 illustrates two sample subjects of the YALE database along with variations in facial expression and configuration.

We also test LPFDA against PCA, FDA and LPP. We form the training set by randomly selecting 6 images per individual with labels, and the rest of the database was considered to be the testing set. Nearest neighborhood classifier is employed in the experiments. The experiments are conducted for 10 times, and we report the average results here. The results are given in Fig.5. The horizontal axis represents the dimension of the subspace and the vertical axis stands for the recognition rate.

Fig. 4. Sample faces images from the YALE database along with variations in facial expression and configuration
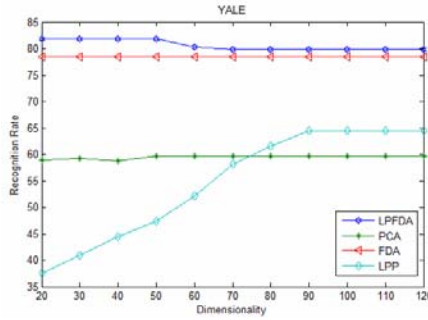
Fig. 5. The recognition rates versus the subspace dimension on YALE database

This time, LPFDA outperforms all the other 3 methods too. It can be seen that LPFDA averagely exceeds PCA and FDA 20 percent and 2.5 percent respectively. Also the second best recognition (81.8%, which is less than the first highest recognition rate for 0.07 percent) achieved by LPFDA appears at the dimension 20 which is closed to the number of the classes. Meanwhile, we notice that LPP gets a similar performance like in ORL data base, which shows the importance of using the labels information to do categorization work again.

Also we display the first 10 LPFisherfaces along with Eigenfaces, Fisherfaces and Laplacianfaces in Fig.6.
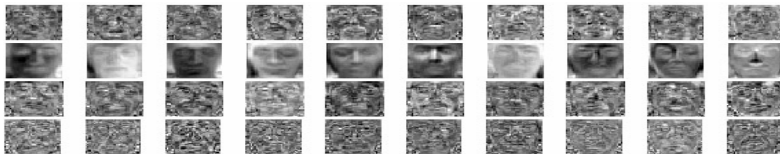
Fig. 6. From top to bottom: the first 10 LPFisherfaces, Eigenfaces, Fisherfaces and Laplacianfaces obtained from YALE data base

The best recognition rate achieved by each method along with the corresponding subspace dimension d is shown in Table 2.

Table 2. The best recognition rate achieved by each method along with the corresponding subspace dimension d

	Subspace Dimension d	Recognition Rate (%)
LPFDA	40	81.87
PCA	50	59.73
FDA	14	78.4
LPP	90	64.4

5   Conclusions and Future Works

In the paper, a novel algorithm named Locality Preserving Fisher Discriminant Analysis (LPFDA) is proposed. LPFDA is derived by designing a heuristic object function which combines FDA with LPP. Through solving an eigenvalue problem, we get the projection vectors in the low dimension space. Face recognition experiments are conducted on two famous databases to demonstrate the effectiveness and robustness of LPFDA.

Refer to the future work, we will try to develop a kernel extended version of LPFDA, and study how to harmonize the impact factor of the linear portion and the nonlinear portion in LPFDA.

References

1. Jolliffe, I.T.: Principal Component Analysis. Springer, New York (1986)
2. Belhumeur, P.N., Hespanha, J.P., Kriegman, D.J.: Eigenfaces vs. Fisherfaces: Recognition Using Class Specific Linear Projection. IEEE Trans. Pattern Analysis and Machine Intelligence 19(7), 711–720 (1997)
3. Roweis, S., Saul, L.: Nonlinear Dimensionality Reduction by Locally Linear Embedding. Science 290, 2323–2326 (2000)
4. Saul, L., Roweis, S.: Think Globally, Fit Locally: Unsupervised Learning of Low Dimensional Manifolds. Journal of Machine Learning Research 4, 119–155 (2003)
5. Tenenbaum, J.B., de Silva, V., Langford, J.C.: A Global Geometric Framework for Nonlinear Dimensionality Reduction. Science 290, 2319–2323 (2000)
6. He, X., Niyogi, P.: Locality Preserving Projections. In: Proc. Advances in Neural Information Processing System Conf. (2003)
7. He, X., Yan, S., Hu, Y., Niyogi, P., Zhang, H.J.: Face Recognition Using Laplacianfaces. IEEE. Trans. Pattern Analysis and Machine Intelligence 27(3) (2005)
8. He, X., Cai, D., Yan, S., Zhang, H.J.: Neighborhood Preserving Embedding. In: Proc. of the Tenth IEEE International Conference on Computer Vision, vol. 2, pp. 1208–1213 (2005)
9. Zhao, X., Jiang, Z.L.: A Nonlinear Structure Preserving Dimensionality Reduction Algorithm. In: Proceedings of the 10 IASTED International Conference on Signal Processing and Applications, pp. 200–204 (2008)
10. Belkin, M., Niyogi, P.: Laplacian Eigenmaps and Spectral Techniques for Embedding and Clustering. In: Advances in Neural Information Processing Systems 14, Vancouver, British Columbia, Canada (2002)
11. Chung Fan, R.K.: Spectral Graph Theory. Regional Conferences Series in Mathematics, vol. 92 (1997)
12. Samaria, F., Harter, A.: Parameterisation of a Stochastic Model for Human Face Identification. In: Proc. Second IEEE Workshop Applications of Computer Vision (1994)
13. Yale Univ. Face Database,
   http://cvc.yale.edu/projects/yalefaces/yalefaces.html

Palmprint Recognition Using Band-Limited Phase-Only Correlation and Different Representations

Yi-Hai Zhu[1,2], Wei Jia[1], and Ling-Feng Liu[1,2]

[1] Intelligent Computation Laboratory, Hefei Institute of Intelligent Machines,
Chinese Academy of Science, P.O. Box 1130, Hefei, Anhui 230031, China
[2] Department of Automation, University of Science and Technology of China,
Hefei 230027, China
`yihaizhu@gmail.com`

Abstract. In this paper, we propose a novel approach for palmprint recognition, which combine band-limited phase-only correlation method and directional representation of palmprint. We also exploit *peak-to-sidelobe ratio* as the similarity measure. The results of experiments conducted on Hong Kong Polytechnic University Palmprint Database show that the proposed approach has higher accurate recognition rates and lower equal error rates than that of the approach, which combine band-limited phase-only correlation and original representation.

Keywords: biometric, palmprint recognition, phase-only correlation.

1 Introduction

Nowadays, palmprint recognition, as a new biometric technology, has been receiving wide attentions from researchers [1]. So far, there have been many approaches proposed for palmprint recognition [2]. Kong et al. has made a survey for this technique and divided the approaches into several different categories [2].

Recently, Phase-only correlation (POC) and Band-Limited Phase-Only correlation (BLPOC) have been successfully applied to palmprint recognition [3,4]. In [3], BLPOC was used for palmprint recognition for the first time. Then, in [4], the performance of BLPOC has been improved by two aspects: (a) reducing the registered data size by registering quantized phase information and (b) dealing with nonlinear distortion between palmprint images by local block matching.

In this paper, we also propose a method for palmprint recognition exploiting BLPOC. Compared to the work [3] and [4], we make two improvements. The first one is that we combine the BLPOC and directional representation (DR) of palmprint for recognition, which can obtain higher accurate recognition rates and lower equal error rates (EER) than the approach combining BLPOC and original representation (OR) of palmprint. The second one is that we exploit *peak-to-sidelobe ratio* (PSR) similarity measure to further improve the recognition rates.

The experiments were conducted on Hong Kong Polytechnic University (PolyU) Palmprint Database. And experimental results have demonstrated the effectiveness of the proposed method.

2   Fundamentals of Band-Limited Phase-Only Correlation

In this section, we will introduce the fundamentals of BLPOC [3,4]. Firstly, the definition of POC is described as follows:

Consider two $N_1 \times N_2$ images, $f(n_1,n_2)$, and $g(n_1,n_2)$. Let $F(k_1,k_2)$ and $G(k_1,k_2)$ denote the 2D DFTs of the two images. Here, $F(k_1,k_2)$ is given by:

$$F(k_1,k_2) = \sum_{n_1=0}^{N_1} \sum_{n_2=0}^{N_2} f(n_1,n_2) e^{-j2\pi(\frac{n_1 k_1}{N_1}+\frac{n_2 k_2}{N_2})}$$
$$= A_F(k_1,k_2) e^{j\theta_F(k_1,k_2)} \tag{1}$$

where $A_F(k_1,k_2)$ is amplitude and $\theta_F(k_1,k_2)$ is phase. $G(k_1,k_2)$ can be defined in the same way. The cross-phase spectrum $R_{FG}(k_1,k_2)$ is given by:

$$R_{FG}(k_1,k_2) = \frac{F(k_1,k_2)\overline{G(k_1,k_2)}}{|F(k_1,k_2)\overline{G(k_1,k_2)}|} = e^{j\theta(k_1,k_2)} \tag{2}$$

where $\overline{G(k_1,k_2)}$ is the complex conjugate of $G(k_1,k_2)$ and $\theta(k_1,k_2)$ denotes the phase difference $\theta_F(k_1,k_2) - \theta_G(k_1,k_2)$. The POC function $r_{fg}(n_1,n_2)$ is the 2D Inverse DFT (2D IDFT) of $R_{FG}(k_1,k_2)$ and is given by:

$$r_{fg}(n_1,n_2) = \frac{1}{N_1 N_2} \sum_{k_1 k_2} e^{j\theta(k_1,k_2)} e^{j2\pi(\frac{n_1 k_1}{N_1}+\frac{n_2 k_2}{N_2})} \tag{3}$$

From formulas (2) and (3), we can see that original POC exploit all components of image's 2D DFT to generate the out plane. In [3], ITO et al., found that BLPOC can achieve better recognition performance by removing the high frequency components and only using the inherent frequency band for matching.

Here we denote the center area of $\theta_F(k_1,k_2)$ and $\theta_G(k_1,k_2)$ as $\theta_F(k_1,k_2)_{BL}$ and $\theta_G(k_1,k_2)_{BL}$, whose size is $J_1 \times J_2$. Thus, the BLPOC function is given by:

$$r_{fg}(n_1,n_2)_{BL} = \frac{1}{J_1 J_2} \sum_{k_1 k_2} e^{j(\theta_F(k_1,k_2)_{BL} - \theta_G(k_1,k_2)_{BL})} e^{j2\pi(\frac{n_1 k_1}{J_1}+\frac{n_2 k_2}{J_2})} \tag{4}$$

3   Directional Representation of Palmprint

In our prior work [5,6,7], we exploited DR of palmprint for appearance based approaches, and better recognition results have been reported. In this paper, we try to combine BLPOC and DR to see whether the better recognition rates can be obtained or not. Here, we also use the MFRAT to extract directional representation of palmprint [5,6]. The MFRAT is defined as follows:

Denoting $Z_p=\{0,1,\ldots,p-1\}$, where p is a positive integer, the MFRAT of real function $f[x,y]$ on the finite grid Z_p^2 is defined as:

$$r[L_k] = MFRAT_f(k) = \sum_{i,j \in L_k} f[i,j] \tag{5}$$

where L_k denotes the set of points that make up a line on the lattice Z_p^2, which means:

$$L_k = \{(i,j): j = S_k(i-i_0) + j_0, i \in Z_p\} \tag{6}$$

where (i_0,j_0) denotes the center point of the lattice Z_p^2, and k means the index value corresponding to the slope of S_k. That is to say, different k denotes different slopes of L_k. For any given k, the summation of only one line, which passes through the center point (i_0,j_0) of Z_p^2, is calculated. It should be pointed out that all lines at different directionls have an identical number of pixels. The discussion about the differences between finite radon transform (FRAT) and MFRAT can be found in [7].
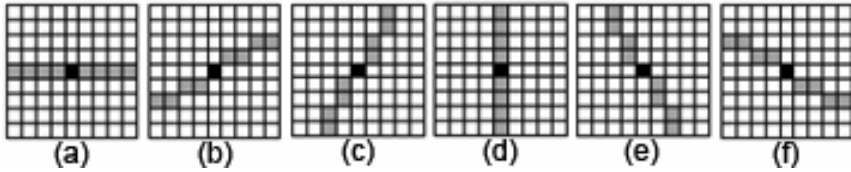
Fig. 1. The 9×9 MFRAT. (a) $k1$ represents 0°; (b) $k2$ represents $\pi/6$; (c) $k3$ represents $2\pi/6$; (d) $k4$ represents $3\pi/6$; (e) $k5$ represents $4\pi/6$; (f) $k6$ represents $5\pi/6$.

In the MFRAT, the index value $k(i_0,j_0)$ of center point $f(i_0,j_0)$ of the lattice Z_p^2 is calculated by following formula:

$$k(i_0, j_0) = \arg(\min_k(r[L_k]))   \quad k = 1,2,\cdots N \tag{7}$$

In this way, the directions of all pixels can be calculated if the center of lattice Z_p^2 moves over an image pixel by pixel. For an image $I(x,y)$ of size $m\times n$, if the values of all pixels are replaced by their index values of direction, the DR of this palmprint can be constructed as follows:

$$DR = \begin{vmatrix} k(1,1) & k(1,2) & \vdots & k(1,n) \\ k(2,1) & k(2,2) & \vdots & k(2,n) \\ \ldots & \ldots & \ldots & \ldots \\ k(m,1) & k(m,2) & \vdots & k(m,n) \end{vmatrix} \tag{8}$$

Fig. 1 shows an example of MFRAT with size 9×9. Table 1 lists the angles and their corresponding index value k. In order to view the DR clearly, we multiply k by 20 to form the final DR image.

Fig. 2 shows three palmprint images and their DRs extracted by 23×23 MFRAT. In Fig.2, The first row shows original images, and the second row is the corresponding directional images. Among them, Fig 2(b) and (c) come from the same palm, but

Table 1. The index values representing different angles

k	7	8	9	10	11	12
Angle	0º	$\pi/12$	$2\pi/12$	$3\pi/12$	$4\pi/12$	$5\pi/12$
k	1	2	3	4	5	6
Angle	$6\pi/12$	$7\pi/12$	$8\pi/12$	$9\pi/12$	$10\pi/12$	$11\pi/12$

(a)	(b)	(c)
(d)	(e)	(f)
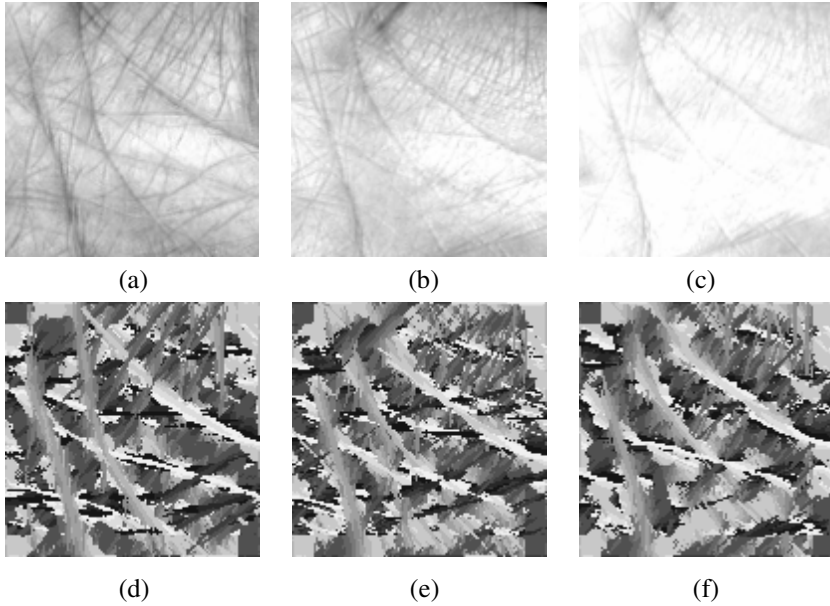
Fig. 2. Plmprint images and their directional representation

were captured in different illumination conditions. Although the illumination conditions changed drastically, however, their DRs are still very similar (see Fig.2d and f).

In fact, the size p of MFRAT is an important parameter, which will influence the recognition rates. We will discuss how to determine a proper p by experiments in Section 4.

4   Experiments

The proposed approach in this paper was tested on the Hong Kong Polytechnic University (PolyU) Palmprint Database. This database contains 7752 grayscale palmprint images from 386 palms corresponding to 193 individuals, in which about 20 samples from each of these palms were collected in two sessions. And the total numbers of images captured in the first session and the second session are 3889 and 3863, respectively. A detailed introduction of PolyU Database can be found in [1]. In our paper, by using the similar preprocessing approach described in literature [1], palmprint is orientated and the ROI, whose size is 128×128, is cropped. In our experiments, we

use the first three palmprints from the first session for training and leave the palm-prints from the second session for testing. Therefore, the numbers of images for training and test are 1158 and 3863, respectively. When a test image A matches with three training images belonging to a same class i, there will generate three matching scores. The largest one will be selected as the final matching score between A and class i. In experiments, The nearest neighbor rule is used for classification. The EER is adopted for evaluate the verification performance, and the recognition rate (RR) is exploited to evaluate the identification performance.

The experiments were conducted on a personal computer with an Intel Duo T7500 processor (2.20GHz) and 2.0G RAM configured with Microsoft Vista and Matlab 7.0.

In [3] and [4], the *peak* was adopted for similarity measure. In this paper, we will also investigate the recognition performance of measures *peak-to-correlation energy* (PCE), and *peak-to-sidelobe ratio* (PSR). As the name suggests, *peak* is the maximum peak value in the correlation out plane (COP). PCE and PSR are defined by:

$$PCE = \frac{peak - mean_{COP}}{std_{COP}}, \quad PSR = \frac{peak - mean_{sidelobe}}{std_{sidelobe}} \tag{9}$$

where $mean_{COP}$ is the average of the COP, std_{COP} is the standard deviation of the COP, $mean_{sidelobe}$ is the average of the sidelobe region surrounding the peak (21×21 pixels with a 5×5 excluded zone around the peak), and $std_{sidelobe}$ is the standard deviation of the sidelobe region values.

The first experiment is to evaluate the recognition performance of three different similarity measures i.e., *peak*, PCE and PSR on OR. In this experiment, determining suitable values of M_1 and M_2 is a key problem that should be solved firstly. Since the ROI image of Palmprint is a square, and its Fourier spectrum is also a square, we let M_1 equal to M_2. That is to say, the selected center area of the 2D DFT spectrum is also a square, whose size is $M_1 \times M_1$. Furthermore, in order to choose the best M_1, we
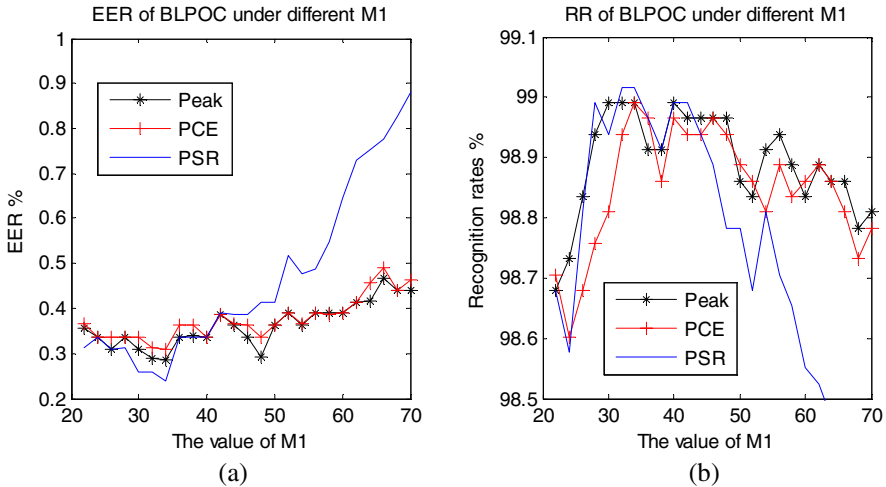
Fig. 3. EERs (a) and RR (b) of BLPOC + OR using different measures under different M_1

Table 2. The lowest EER (LEER) and highest recognition rates (HRR) of BLPOC + OR using different measures, and corresponding band size of $M_1 \times M_1$

Lowest EER of BLPOC + OR				Highest RR of BLPOC + OR			
LEER-Peak	0.285%	$M_1 \times M_1$	34×34	HRR-Peak	98.99%	$M_1 \times M_1$	30×30
LEER -PCE	0.309%	$M_1 \times M_1$	34×34	HRR-PCE	98.99%	$M_1 \times M_1$	34×34
LEER -PSR	0.238%	$M_1 \times M_1$	34×34	HRR-PSR	99.02%	$M_1 \times M_1$	34×34

conduct the tests exploiting different values of M_1. Here, the values of M_1 are set to an even number, and the range of M_1 is {22, 24, ···, 70}. Fig.3a depicts the EERs when the values M_1 changes from 22 to 70. And Fig.3b shows RRs corresponding to different M_1. It can be easily seen that using PSR can generate the lowest EER and highest RR. Table 2 lists the lowest EER and highest RR obtained by *peak*, PCE and PSR, respectively. We can know the results obtained by PSR are best.

In Section 3, we have mentioned that the size of MFRAT will influence the recognition rates. Thus, the second experiment aims to find the size of MFRAT, which can achieve best recognition performance. In order to do this task, we change the size p of MFRAT from 11 to 21 with an interval of 2. Meanwhile, the arrangement of index values of directions is also adopted the one listed in Table 1. It should be noted that in this experiment we only exploit PSR for similarity measure. Fig.4 depicts the EERs and recognition rates under different p. When the size p of MFRAT is set to 13×13, the lowest EER is obtained, which is 0.129%. And when the p is set to 15×15, we can get highest recognition rate, which is 99.94%.

In the first experiment, we concluded that PSR is a better similarity measure than *peak* and PCE. In the third experiment, we will investigate whether this conclusion is true or not when the DR is used. Here, we fix the value of p as 13×13, and see the recognition performance of peak, PCE and PSR by changing the band width of M_1 from 22 to 70 with an interval of 2.

(a)   (b)
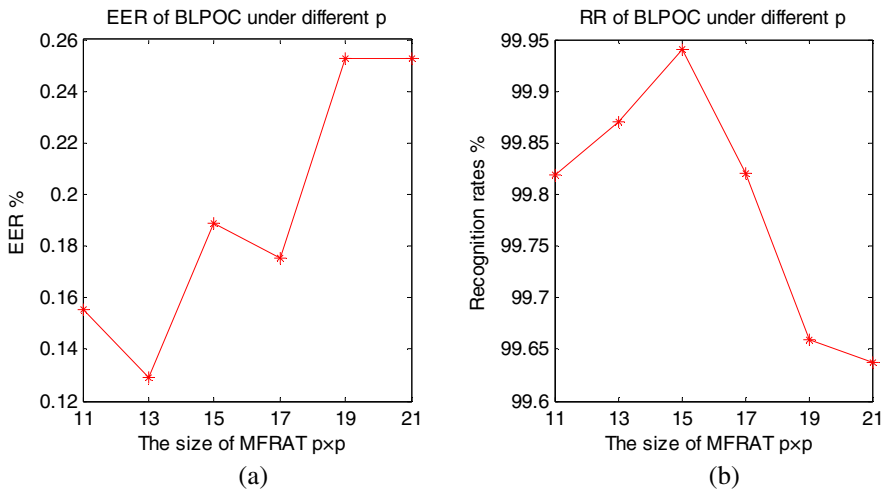
Fig. 4. EERs (a) and RR (b) of BLPOC + DR using PSR as similarity measures while the size of MFRAT p changed from 11 to 21 with an interval of 2
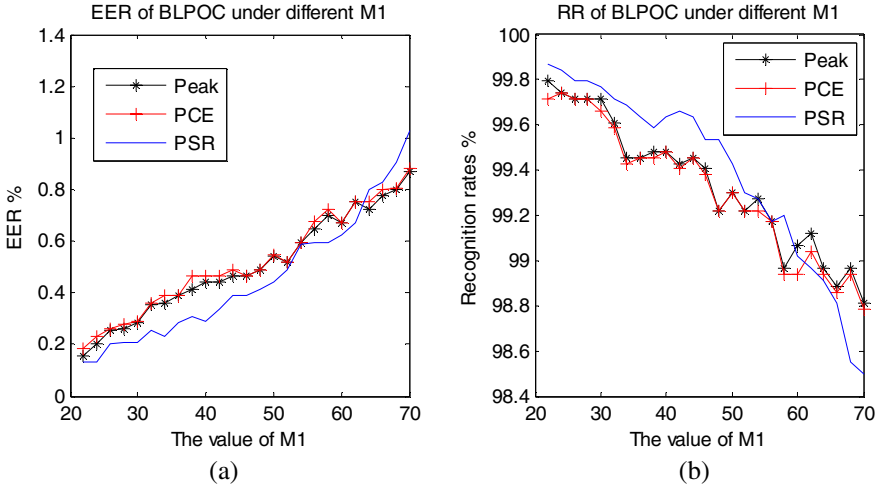
Fig. 5. EERs (a) and RR (b) of BLPOC + DR using different measures under different M_1, while the size p of MFRAT is set to 13×13

Table 3. The lowest EER (LEER) and highest recognition rates (HRR) of BLPOC + DR using different measures, and corresponding band size of $M_1 \times M_1$, while the size p of MFRAT is set to 13×13

Lowest EER of BLPOC + DR				Highest RR of BLPOC + DR			
LEER-Peak	0.155%	$M_1 \times M_1$	22×22	HRR-Peak	99.79%	$M_1 \times M_1$	22×22
LEER -PCE	0.186%	$M_1 \times M_1$	22×22	HRR-PCE	99.74%	$M_1 \times M_1$	24×24
LEER -PSR	0.129%	$M_1 \times M_1$	22×22	HRR-PSR	99.87%	$M_1 \times M_1$	22×22

Table 4. The comparision of EER between proposed method and the method in [4]

	Proposed method	Method in [4]
EER	0.129%	0.224%

Fig.5 depicts the EERs and recognition rates using *peak*, PCE and PSR as similarity measure, respectively. It can be easily seen that using PSR can also get the lowest EER and highest RR. Table 3 lists the lowest EER and highest RR obtained by *peak*, *PCE* and *PSR*, respectively. We can know the results obtained by *PSR* are best when DR is used.

We compare the EERs of the proposed method with the work [4], which are listed in Table 4. We can see the recognition performance is also better than the method in [4].

5   Conclusion

In this paper, we propose a novel approach for palmprint recognition, which combine band-limited phase-only correlation method and DR of palmprint. We also exploit *PSR*

as the similarity measure. The results of experiments conducted on Hong Kong Polytechnic University Palmprint Database show that using DR and PSR, the recognition performance of BLPOC has been significantly improved. For example, when using PSR as similarity measure, the lowest EER of BLPOC+DR is 0.129%, which is about half of the method BLPOC+OR, and the highest RR of BLPOC+DR is 99.94%, which is about 1% higher than the HRR of the method BLPOC+OR. In addition, the recognition performance of proposed method is better than that of the work in [3] and [4].

Acknowledgements

References

1. Zhang, D., Kong, A., et al.: Online Palmprint Identification. IEEE T PAMI 25(9), 1041–1050 (2003)
2. Kong, A., Zhang, D., Kamel, M.: A Survey of Palmprint Recognition. Pattern Recognition 42(7), 1408–1418 (2009)
3. ITO, K., et al.: A Palmprint Recognition Algorithm Using Phase-Only Correlation. IEICE T Fundamentals (4), 1023–1030 (2008)
4. Litsuka, S., Ito, K., Aoki, T.: A Practical Palmprint Recognition Algorithm Using Phase Information. In: 2008 IEEE International Conference on Pattern Recognition (ICPR), USA (2008)
5. Huang, D., Jia, W., Zhang, D.: Palmprint Verification Based on Principal Lines. Pattern Recognition 41(4), 1316–1328 (2008)
6. Jia, W., Huang, D., Zhang, D.: Palmprint Verification Based on Robust Line Orientation Code. Pattern Recognition 41(5), 1504–1513 (2008)
7. Jia, W., Huang, D., et al.: Palmprint Identification Based on Directional Representation. In: 2008 IEEE International Conference on SMC, Singapore, pp. 1562–1567 (2008)

A Design and Research of Eye Gaze Tracking System Based on Stereovision

Pengyi Zhang[1], Zhiliang Wang[1], Siyi Zheng[1,2], and Xuejing Gu[1,3]

[1] Information Engineering Institute, Beijing University of Science and Technology, Beijing, 100083, China
[2] Beijing Command College of Chinese People's Armed Police Force, 100012, Beijing, China
[3] School of Information, Hebei Polytechnic University, 063009, China
`zpy19811103@163.com, wzl@ies.ustb.edu.cn, sisizheng@163.com,`
`guxuejing@sina.com`

Abstract. A new design for an eye gaze tracking system based on stereovision technique is presented. The system consists of two CCD cameras and two novel light sources for stereovision. The way of getting pupil position is to do subtraction of two images, the "bright pupil" and the "dark pupil", which are gained by illuminating user's eyes alternately. The pupil center is located by ellipse fitting when the Purkinje image is gained in the "dark pupil", so the local gaze direction can be obtained. We also use support vector regression to figure out the mapping relationship from eye parameters to gaze point, and the interference from head motion may be eliminated by using 3D eyeball data. The experimental results show that the system can achieve an average accuracy of 1.8 degree and be robust in gaze tracking under large head movements.

Keywords: Eye Gaze Tracking, Stereovision, Pupil, Purkinje Image.

1 Introduction

Non-contact eye gaze tracking system, the primary design of Human-Computer Interaction (HCI), has made great improvement due to the advance in technology and the expanding of application fields in recent years[1,2]. The system of non-contact eye gaze tracking is impelled to more precision, higher efficiency and lower cost by the updating of iris detecting and pupil tracking technology[3-5].

A non-contact eye gaze tracking system based on stereovision is designed to help the handicapped and the elderly to use computer for information interacting. The system used new methods of detecting and tracking as well as new mapping algorithm. The precision and real-time performance is greatly improved.

To solve the problem of head motion in eye gaze tracking system, the system platform is mainly designed to contain two cameras and loop light sources. The "bright pupil" and the "dark pupil", illuminated by each loop light source between two frames, can be captured by each camera. The position of pupil in image can be obtained by subtraction and connectivity analysis, and the space position of eyeball can be achieved by stereovision technology. The center of pupil and Purkinje image can be obtained by

the algorithm of ellipse fitting and gray detection, using the vector difference between them as the input of mapping model. Based on the characteristics of human gazing movement, a tracking prediction algorithm is designed by combining Kalman filtering and neural network. The experimental data show that the tracking has been improved. Finally, the mapping model is obtained by support vector regression analysis between eye gaze direction and other variables such as the positions of eyeball, pupil and Purkinje image vector.

This paper is organized as follows. In Section 2, the configuration of system hardware is described. We propose the method of pupil detection and tracking as well as how to get the gaze mapping model using 3D eye data in Section 3. Experimental results are shown in Section 4. The paper ends in Section 5 with a conclusion of our work.

2 System Hardware Description

To make user environment more comfortable without the loss of tracking precision, the non-contact eye gaze tracking system is designed to contain two CCD cameras and two loop infrared light sources. And the system also contains image grabber, personal computer and processor, as shown in Fig.1 (a) and Fig.1 (b).

The stereo camera system provides 3D coordinates of pupil for tracking by image capture and 3D-space ranging. Video source acquisition and pretreatment are accomplished by image grabber, and then the information is sent to personal computer for image processing. Once the frame synchronization signal captured, the two light sources illuminate alternatively controlled by processor, then the image of the "bright pupil" and the "dark pupil" (see in Fig.2) can be obtained from two consecutive frame images of either CCD camera. Infrared LED (Light Emitting Diode), as the light source has a peak wavelength of 880nm. To produce the "bright pupil" and "dark pupil" as well as to minimize interference from daylight, we use an optical band pass filter, which has a wavelength pass band only 10nm wide at 880nm.
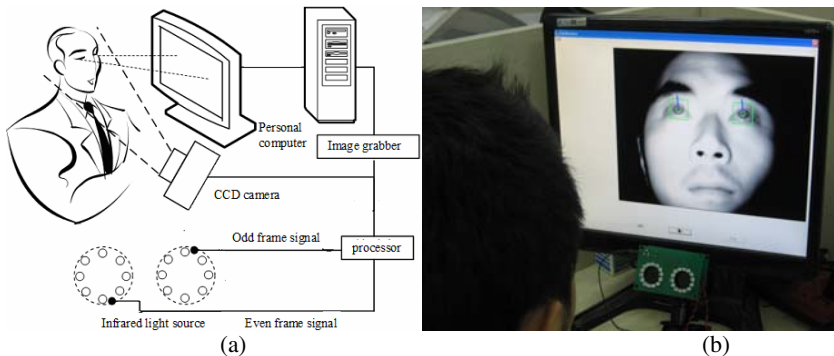
Fig. 1. （a）The eye gaze tracking system configuration. (b)Actual photograph of the eye gaze tracking system.
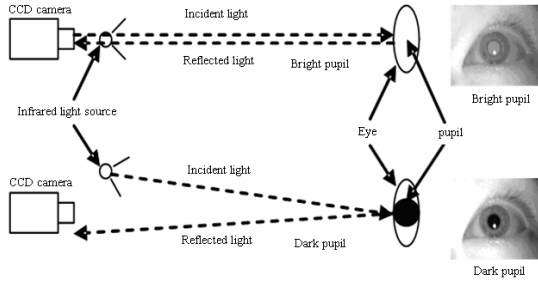
Fig. 2. Principle of bright and dark pupil effects

3 Proposed Methods

3.1 Pupil Detection

3.1.1 Image Subtraction

To get the position of pupil, the first step is to make subtraction between two consecutive frame images which are the "bright pupil" (see in Fig.3(a)) and the "dark pupil" (see in Fig.3(b)). Fig.3(c) is a result from the subtraction between Fig.3(a) and Fig.3(b). Obviously, except for the pupil, facial contour is also displayed in Fig.3(c) because of two images' gray difference. To eliminate interference from facial contour, we use connectivity analysis, and set the criterion and the threshold for the subsequent image processing based on the shape of pupil and its gray value. The processed image is shown in Fig.3(d).

| (a) | (b) | (c) | (d) |
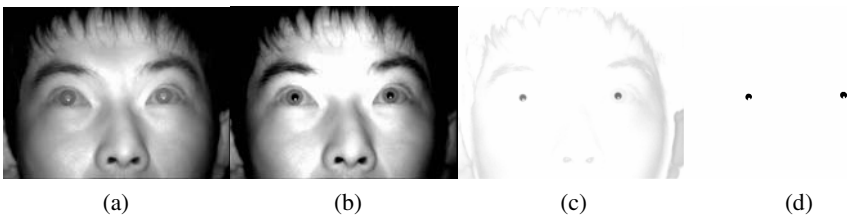
Fig. 3. Image subtraction and processing. (a) The "bright pupil". (b) The "dark pupil". (c) Subtraction between (a) and (b). (d) Image processed result.

3.1.2 Determining the Positions of Pupil and Purkinje Image

After the facial contour interference was removed by the above method, there are only two pupils in the image. The pupils may be detected by searching the entire image to locate two dark regions that satisfy certain size, shape, and distance constraints. Therefore, the Purkinje image may be found in Fig.4(c) because of its higher gray value compared with the "dark pupil".

The center of pupil is obtained by ellipse fitting (see in Fig.4(b)). Among a number of discrete data, an ellipse can be determined by six points. Therefore, pupil center may be located by ellipse fitting that choose six points randomly from thirty boundary ones via least-square method based algebraic distance. The ellipse can be denoted by the equation of conic section:

$$Ax_i^2 + Bx_i y_i + Cy_i^2 + Dx_i + Ey_i + F = 0 \tag{1}$$

Where, (x_i, y_i) is the point coordinate on ellipse. Substitute six points into the Eqn(1) respectively, then the parameters A、B、C、D、E、F are computed. So the center coordinates of ellipse is:

$$X_c = \frac{BE - 2CD}{4AC - B^2} \qquad Y_c = \frac{BD - 2AE}{4AC - B^2}$$

We use above method to do the calculation several times, then the average of those ellipse center coordinates is the ultimate pupil center. The center of pupil and Purkinje image are shown in Fig.4(d).

| (a) | (b) | (c) | (d) |
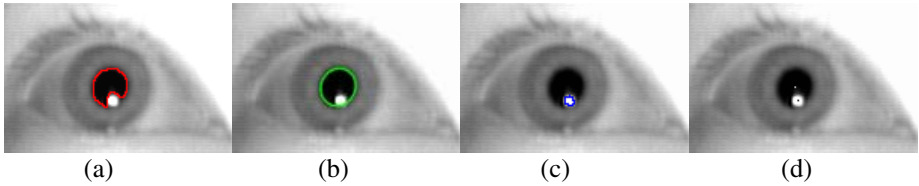
Fig. 4. Detection of gaze parameters. (a) Pupil edge. (b) Ellipse fitting of pupil. (c)The contour of Purkinje image. (d) Center of pupil and Purkinje image.

3.2 Pupil Tracking

The pupil position can be tracked in real-time once the initial condition and prediction model above are determined. In order to find real-time gaze point, the pupil must be tracked constantly in each frame, so the Kalman filtering is selected to track the pupil position.

The prediction model and the measuring model should be set up first based on the principle of Kalman filter. However, it is very difficult to build an accurate model for prediction because gaze moving is greatly random and closely related with personal habit, especially for the distributed information with a complicated non-linear relationship. To solve this problem, we use BP neural network to approximate the state transition model of gaze moving. Suppose $x_t = (c_t, r_t, u_t, v_t)^T$, the state vector, is the input of BP neural network. Where, (c_t, r_t) is the pixel position of pupil and (u_t, v_t) is the vector in c and r directions, respectively. The state vector of pupil at time $t+1$, $x_{t+1} = (c_{t+1}, r_{t+1}, u_{t+1}, v_{t+1})^T$, is the output of BP neural network. There are two layers in BP neural network, the sigmoid function is used for hidden layer and the pureline function is used for output layer. In the case of the same distributed information source,

we choose multi-age samples to train by BP neural network so that the prediction model can be more universal.

After training, the prediction model can be denoted as:

$$x_{t+1} = \boldsymbol{\Phi} x_t + w_t \tag{2}$$

Where, $\boldsymbol{\Phi}$ is the state transition model which is trained by BP neural network, w_t is the system perturbation, fitted normal distribution as $w_t \sim (0, Q)$. We further use a feature detector denoted as $z_t = (\hat{c}_t, \hat{r}_t)$, to estimate pupil position at time t. Therefore, the measuring model is represented as:

$$z_t = \boldsymbol{H} x_t + v_t \tag{3}$$

Where, v_t represents measurement uncertainty, fitted normal distribution as $v_t \sim N(0, R)$. $\boldsymbol{H}$ is the relative state to current measured value. Since the movement of pupil between two consecutive frames is similar and within little distance and z_t only involves pupil position, $\boldsymbol{H}$ can be simply expressed as:

$$\boldsymbol{H} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{bmatrix}$$

We start Kalman filter tracker after we detected the pupil successfully between two consecutive frames. Let t and $t+1$ be the two frames. The initial state vector x_0 can be specified as:

$$r_0 = r_{t+1}, \quad c_0 = c_{t+1}, \quad u_0 = c_{t+1} - c_t, \quad v_0 = r_{t+1} - r_t$$

Via the error of pixels and vectors respectively, the initial estimator error covariance matrix Σ_0 and the initial state covariance matrix Q can be defined as:

$$\Sigma_0 = \begin{bmatrix} 100 & 0 & 0 & 0 \\ 0 & 100 & 0 & 0 \\ 0 & 0 & 25 & 0 \\ 0 & 0 & 0 & 26 \end{bmatrix} \qquad Q = \begin{bmatrix} 16 & 0 & 0 & 0 \\ 0 & 16 & 0 & 0 \\ 0 & 0 & 4 & 0 \\ 0 & 0 & 0 & 4 \end{bmatrix}$$

The error for measuring model as 2 pixels for both c and r directions. Thus,

$$\boldsymbol{R} = \begin{bmatrix} 4 & 0 \\ 0 & 4 \end{bmatrix}$$

Both Q and R are assumed to be stationary constant.

The pupil position can be predicted accurately in each frame based on the initial condition and the above two models, as well as five equations of Kalman filter as Eqn(4) ~ Eqn(8):

$$x_{t+1}^{-1} = \boldsymbol{\Phi} x_t \tag{4}$$

$$\sum\nolimits_{t+1}^{-1} = \Phi \sum\nolimits_{t} \Phi^{t} + Q_{T} \tag{5}$$

$$K_{t+1} = \frac{\sum\nolimits_{t+1}^{-1} H^{T}}{H \sum\nolimits_{t+1}^{-1} H^{T} + R} \tag{6}$$

$$x_{t+1} = x_{t+1}^{-1} + K_{t+1}(z_{t+1} - Hx_{t+1}^{-1}) \tag{7}$$

$$\sum\nolimits_{t+1} = (I - K_{t+1}H)\sum\nolimits_{t+1}^{-1} \tag{8}$$

Where, Eqn(4) is the state prediction, Eqn(5) is the error covariance prediction, Eqn(6) is the update of the Kalman gain, Eqn(8) is the update of the error covariance, and Eqn(7) is the optimal solution. The performance of tracking strategy is described in section 4.1.

3.3 Designing the Gaze Mapping Model

The location of gaze point on computer screen can be worked out by using the collected information previously after pupil detection and tracking. Since the gaze mapping model is a complicated high order nonlinear function, we use the support vetor regression to estimate the mapping model for higher precision.

The gaze mapping model is just a mathematic equation between the gaze point and other variables such as the 3D positions of eyeball, user's pupil and Purkinje image vector. Suppose $V = [d_x, d_y]$ is the vector between pupil center and Purkinje image center, O is center position of eyeball in 3D space, and the gaze point is (s_x, s_y). The center of eyeball O, which can not be gained directly by cameras, can be replaced by the pupil center $P = (p_x, p_y, p_z)$ obtained directly from measurement because of their close distance less than 8mm. In addition, both cameras are calibrated to form a stereo vision system so that the 3D coordinate of the pupil center can be computed. So, $X_g = [d_x, d_y, p_x, p_y, p_z]$ is the input of support vector regression and (s_x, s_y) is the output. The performance of the gaze mapping model is given in section 4.2.

4 Experimental Results

4.1 Performance of Pupil Tracking

We choose 347 samples to train the neural network, and the result of pupil tracking using Kalman filter combined with BP neural network is shown in Fig.5(a), and Fig.5(b) is the tracking performance in different conditions. The experimental results are found to be rather robust which can meet the requirement of real-time and stability in system basically.
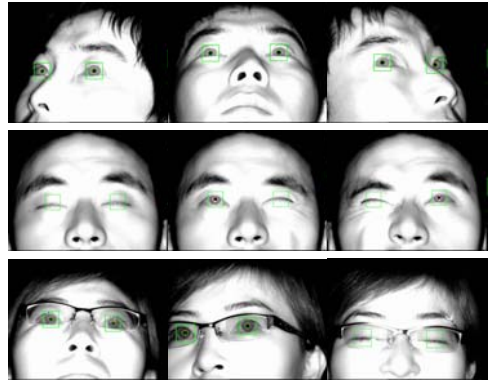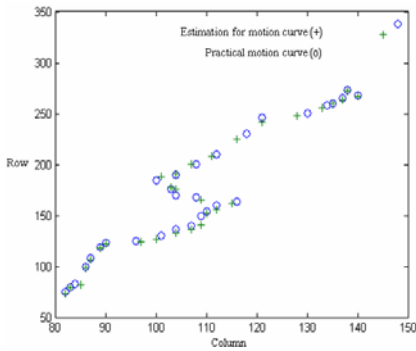
<div align="center">（a）　　　　　　　　　　　　　　（b）</div>

Fig. 5. (a) Kalman prediction of pupil moving based on BP network. (b)The performance of pupil tracking in different conditions.

4.2 Performance of Gaze Mapping Model

To improve mapping accuracy disturbed by head motion, we design the experiments of gazing multi-points in various postures. The data trained by support vector regression are collected on the same condition that the distance between users and the computer screen is about 300mm and users are asked to stare at nine predefined points on the screen in different head postures.

In total, 550 samples composed of the input gaze parameter vector X_g and its corresponding screen gaze point (s_x, s_y) are collected for training, and the user is asked to gaze at nine predefined points one by one under different head positions.

Given sufficient training samples, we believe that a unique nonlinear function that maps the input gaze parameter vectors to the screen gaze points hidden in the training samples can be captured by the SVR model. Since we don't know how complicated the mapping function will be, different kernels and their associated parameter settings need to be manually tested so that the optimal kernel and its parameter settings can be selected to estimate the gaze mapping function accurately. The comparative result of estimative gazing points and practical gazing points is shown in Fig.6.
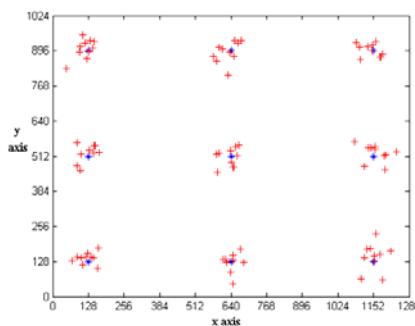
Fig. 6. Estimative gazing points and practical gazing points

Table 1 shows seven users are tested in the experiments with a statistical result of the horizontal average accuracy 1.5°and the vertical average accuracy 1.9°. λ is the precision of plane displacement for the same gaze point. σ is the precision of angle variance for the same gaze point, and the corresponding angle value are given.

Table 1. Results of gazing estimation for different users

User	Horizontal accuracy(λ, σ)	Vertical accuracy(λ, σ)
1	12.00±7.54mm (0.86° ±0.54°)	14.24±8.65mm (1.02° ±0.62°)
2	26.11±10.19mm (1.87° ±0.73°)	32.83±12.42mm (2.35° ±0.89°)
3	26.81±5.58mm (1.92° ±0.40°)	33.80±6.28mm (2.42° ±0.45°)
4	16.89±6.42mm (1.21° ±0.46°)	23.46±7.40mm (1.68° ±0.53°)
5	18.43±4.18mm (1.32° ±0.30°)	24.02±5.72mm (1.72° ±0.41°)
6	28.49±9.21mm (2.04° ±0.66°)	37.44±12.98mm (2.68° ±0.93°)
7	24.30±4.74mm (1.74° ±0.34°)	26.81±7.26mm (1.92° ±0.52°)

Table 2 shows the comparison of accuracy and allowable head movements among several practically working gaze tracking systems that allow natural head movements. For simplicity, only the depth or *Z* direction of the allowed head movement is illustrated, as shown in the second column of Table 2.

Table 2. Comparison with other system

Method	Head movement volume	Average accuracy	Configuration
[6]	<70 mm	0.8°	Multi-cameras, Multi-light sources.
[7]	>70 mm	0.6°	Four cameras, Multi-light sources.
[8]	Around 200 mm	1.6°	Two cameras, Two light sources.
[9]	Around 500 mm	5°	One cameras, One light sources.
[10]	Around 40 mm	0.9°	One cameras, Two light sources.
[11]	Around 100 mm	1.0°	One cameras, Four light sources.
Ours	Around 300 mm	1.8°	Two cameras, Two light sources.

We can see that our proposed technique can provide a competitive gaze accuracy as well as a large head movement volume with only one stereo camera system and without the help of a face tracking system.

5   Conclusion

Through research presented in this paper, we designed an eye gaze tracking system for real-time tracking user's gaze point. We focus on developing the necessary hardware and imaging algorithms that can simultaneously extract multiple visual cues typically characterized user's eye features. The main components of this system consist of a

hardware system for real-time acquisition of video images, various computer vision algorithms and their software implementations for real-time pupil detection, pupil tracking, Purkinje Image localization and gaze estimation.

Each part of our eye gaze tracking system has been tested in a simulating environment with subjects of different backgrounds, genders, ages, and in varied illumination conditions. Experiment results show that the system is very robust and reliable with the high gaze accuracy of 1.8°under natural head movement via proposed methods.

Acknowledgment

References

1. Morimoto, C.H., Mimica, R.M.: Eye Gaze Tracking Techniques for Interactive Applications. J. Computer Vision and Image Understanding 98(1), 4–24 (2005)
2. Hennessey, C., Noureddin, B., Lawrence, P.: Fixation Precision in High-Speed Noncontact Eye-Gaze Tracking. J. IEEE Transactions on Systems, Man, and Cybernetics 38(2), 289–298 (2008)
3. Wang, J.G., Sung, E.: Study on Eye Gaze Estimation. J. IEEE Transactions on Systems, Man, and Cybernetics 32(3), 332–350 (2002)
4. Noureddin, B., Lawrence, P.D., Man, C.F.: A Non-contact Device for Tracking Gaze in A Human Computer Interface. J. Computer Vision and Image Understanding 98(1), 52–82 (2005)
5. Kim, S.C., Sked, M., Ji, Q.: Non-intrusive Eye Gaze Tracking Under Natural Head Movements. In: Proceedings of Engineering in Medicine and Biology Society, pp. 2271–2274. IEEE Press, New York (2004)
6. Shih, S.W., Liu, J.: A Novel Approach to 3-D Gaze Tracking Using Stereo Cameras. J. IEEE Transactions on Systems, Man, and Cybernetics 34(1), 234–245 (2004)
7. Beymer, D., Flickner, M.: Eye Gaze Tracking Using an Active Stereo Head. In: Proceedings of Computer Vision and Pattern Recognition, pp. 451–458. Elsevier Science press, New York (2003)
8. Zhu, Z., Ji, Q.: Novel Eye Gaze Tracking Techniques Under Natural Head Movement. J. IEEE Transactions on Biomedical Engineering 54(12), 2246–2260 (2007)
9. Zhu, Z., Ji, Q.: Eye and Gaze Tracking for Interactive Graphic Display. J. Machine Vision and Applications 15(3), 139–148 (2004)
10. Guestrin, D., Eizenman, M.: General Theory of Remote Gaze Estimation Using the Pupil Center and Corneal Reflections. J. IEEE Transactions on Biomedical Engineering 53(6), 1124–1133 (2006)
11. Villanueva, A., Cabeza, R.: A Novel Gaze Estimation System With One Calibration Point. J. IEEE Transactions on Systems, Man, and Cybernetics 38(4), 1123–1138 (2008)

Computing Parallel Speeded-Up Robust Features (P-SURF) via POSIX Threads

Nan Zhang

Department of Computer Science and Software Engineering,
Xi'an Jiaotong-Liverpool University
nan.zhang@xjtlu.edu.cn

Abstract. Speeded-Up Robust Features (SURF), an image local feature extracting and describing method, finds and describes point correspondences between images with different viewing conditions. Despite the fact that it has recently been developed, SURF has already successfully found its applications in the area of computer vision, and was reported to be more appealing than the earlier Scale-Invariant Feature Transform (SIFT) in terms of robustness and performance. This paper presents a multi-threaded algorithm and its implementation that computes the same SURF. The algorithm parallelises several stages of computations in the original, sequential design. The main benefit brought about is the acceleration in computing the descriptor. Tests have been performed to show that the parallel SURF (P-SURF) generally shortened the computation time by a factor of 2 to 6 than the original, sequential method when running on multi-core processors.

Keywords: Parallel programming, Multi-threaded computing, Image processing, Local feature extraction, Machine intelligence.

1 Introduction

Speeded-Up Robust Features (SURF), proposed in [1], is a local image feature descriptor inspired by the Scale Invariant Feature Transform (SIFT) [2]. Unlike those global feature descriptors, such as MPEG-7 Scalable Colour for colour and MPEG-7 edge histogram for texture [3], which use a single global feature vector to represent an entire image, local descriptors, such as SURF and SIFT, search for distinctive locations, that is, the so called interest points, in an image and then generate vectors to represent the interest points. It is common to find hundreds or thousands of interest points in a medium sized image and the feature vectors generated are, therefore, of hundreds or thousands. It is because of this ability to represent local features, local descriptors, such as SURF and SIFT, are able to find correspondences between two images in the cases of scaling, rotation and view point changing, whereas, in such circumstances, global descriptors will generally fail. Hence, those local descriptors have already found their applications in the areas of object recognition [4], 3D reconstruction [5], sub-image detection [6] and content-based image retrieval [7].

Table 1. Comparisons between the computation times of the MPEG-7 edge histogram and SURF. The parameters were reasonably set. The programs were compiled by g++ 4.3.2 with the -O0 option, and were run on Ubuntu Linux 8.10 (kernel 2.6.27-9). The time durations were measured in milliseconds.

Image	Dimension (pixels)	Edge (ms)	SURF (ms)
test1.pgm	300×300	9.7	263.2
test2.pgm	500×392	11.5	413.8
test3.pgm	640×480	13.6	724.6
test4.pgm	900×600	17.6	1439.5
test5.pgm	1024×768	22.1	1907.3

However, it is also because those algorithms of local descriptors generate hundreds or thousands of feature vectors for one image, they are considerably slower than those of global descriptors. SURF gained much performance increase than the earlier SIFT because the computations are made on the basis of an intermediate image representation known as the integral image [8]. Despite of that, it is still considered slow when applied to time demanding applications. The data in Table 1 shows a group of performance comparisons between implementations of SURF and the MPEG-7 edge histogram. The test were made on an Intel Core Duo P8600@2.4G, although both the algorithms are sequential and therefore only made use of one of the two cores.

For certain applications in the area of computer vision which are not demanding for speed the performance of SURF may suffice. But if applied to other applications, such as those found in large scale content-based image retrieval, where millions of images are to be indexed, the performance of SURF definitely needs to be further speeded up.

To pursue this aim, the current author has designed and developed a parallel algorithm to compute the same SURF through multi-threaded programming via POSIX Thread libraries. This is to take the advantages of the nature of the SURF algorithm as well as the advantage brought about by the multi-core processors widely available in the market. For in computing the SURF, at several stages, it is natural to think about developing parallel solutions to compute and obtain the same results. The current author has therefore taken the step and made parallel several stages in the original, sequential algorithm. Tests later made showed that the acceleration ratios of the parallel SURF against the original were about 2 to 6 times on dual-core or quad-core processors.

Organisation of the paper. A brief explanation of the SURF algorithm is presented in Section 2. However for more detailed discussions the reader should refer to [1] and [9]. Section 3 contains a presentation of the parallel algorithm. Tests and results are reported in Section 4. Conclusions are drawn in Section 5, together with possible future work.

2   Computing SURF

The algorithm that computes SURF, as it is described in [1], consists of two major stages – interest point detection and interest point description – each of which, in turn, takes several steps. In the stage of the interest point detection, the image is first integrated. The output of the integration, known as the integral image, is then used as the basis of the subsequent scale-space analysis. The responses obtained from the scale-space analysis are used to localise the interest points. In the stage of the interest point description, the localised interest points are assigned orientations and then have their features described. This work-flow is shown in Figure 1.
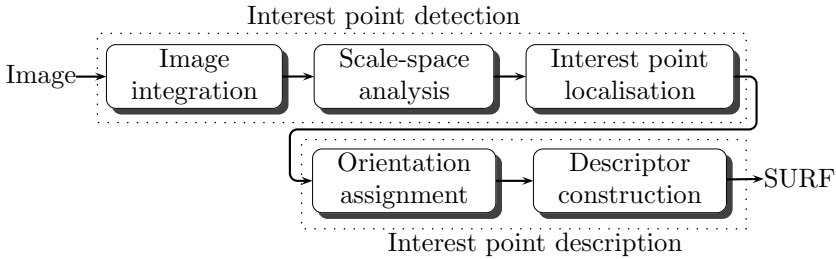
Fig. 1. The work-flow of the algorithm computing SURF

An integral image $I_\Sigma(\boldsymbol{x})$ at a location $\boldsymbol{x} = (x, y)$ is to sum up all the pixels within a rectangular region formed by the origin and $\boldsymbol{x}$. Formally, this can be defined by the formula:

$$I_\Sigma(\boldsymbol{x}) = \sum_{i=0}^{i \le x} \sum_{j=0}^{j \le y} I(i, j). \tag{1}$$

This representation of the original image makes acceleration for the convolutions taking place in the scale-space analysis because of the using of the box filters which approximate the second order Gaussian partial derivatives.

In the subsequent scale-space analysis, the scale space is divided into octaves, each of which is sub-divided into a constant number of intervals. Each of the intervals represents a scale level. The input image is repeatedly convolved by the box filters of increasing sizes, corresponding to the increasing of the scale at which the samplings take place. The box filters are approximations of the second order Gaussian derivatives in the x, y and xy directions. The initial size of the filter is 9×9 approximating a Gaussian with standard deviation 1.2, and represents the lowest scale. The subsequent filter sizes are increased to 15×15, 21×21, 27×27 in the first octave, and 15×15, 27×27, 39×39, 51×51 in the second, ... (see [1]).

Now, if the responses of the input image I at pixel (x, y) under the convolution at scale s in x, y and xy directions are denoted as $D_{xx}(\boldsymbol{x})$, $D_{yy}(\boldsymbol{x})$ and $D_{xy}(\boldsymbol{x})$ respectively, with $\boldsymbol{x} = (x, y, s)$, the following Hessian matrix of approximation is

constructed and its normalised (by the factor 0.9) determinant is calculated as the response $f(\boldsymbol{x})$ of the image I at point (x, y, s).

$$H_{approx} = \begin{bmatrix} D_{xx}(\boldsymbol{x}) & D_{xy}(\boldsymbol{x}) \\ D_{xy}(\boldsymbol{x}) & D_{yy}(\boldsymbol{x}) \end{bmatrix}. \tag{2}$$

$$f(\boldsymbol{x}) = \det(H_{approx}) = D_{xx}(\boldsymbol{x})D_{xy}(\boldsymbol{x}) - (0.9D_{xy}(\boldsymbol{x}))^2, \tag{3}$$

where f denotes the scale-space function.

For an image of w-wide and h-high, totally $w \times h \times o \times i$ of determinants will be calculated, where o is the number of the octaves and i the number of intervals in each octave, although, in practice, not all pixels in the image are sampled. (The step of the sampling is jointly determined by a parameter called `initial sample` and the current octave.) And only the determinants exceeding a certain threshold are saved while the others are set to zeros. The values of the determinants are used for the following interest point localisation.

To localise the interest points, a non-maximum suppression is applied in the image and over the scales, where the Hessian determinant of each sampled pixel is compared with that of all the other pixels in a $3 \times 3 \times 3$ neighbourhood, 8 in its native scale interval, and 9 in each of the intervals above and below, totally 26. The pixel is a local maximum if the value of its determinant surpasses that of all the other 26 surrounding pixel's.

Once a set of local maxima is found, the last step in the interest point detection is to interpolate the nearby data to determine the interpolated location of the pixel in the scale space to sub-pixel accuracy, a method proposed in [10] and explained in [2], but more thoroughly detailed by what follows. This method is to fit a 3D quadratic function to the local maximal points. The scale-space function f at a point $\boldsymbol{x} = (x, y, s)$ in the scale space is expressed as a Taylor expansion up to the quadratic terms and is shifted so that the origin is at a detected maximum $\boldsymbol{x_0} = (x_0, y_0, s_0)$.

$$f(\boldsymbol{x}) = f(\boldsymbol{x_0}) + f'(\boldsymbol{x_0})^{\mathrm{T}} \Delta \boldsymbol{x} + \frac{1}{2} \Delta \boldsymbol{x}^{\mathrm{T}} f''(\boldsymbol{x_0}) \Delta \boldsymbol{x}, \tag{4}$$

where $\Delta \boldsymbol{x} = (\boldsymbol{x} - \boldsymbol{x_0})$ is the offset from the maximum. The location of the extremum, $\hat{\Delta} \boldsymbol{x}$, is found by taking the derivative of this function with respect to $\Delta \boldsymbol{x}$ and setting it to zero

$$\hat{\Delta} \boldsymbol{x} = -f''(\boldsymbol{x_0})^{-1} f'(\boldsymbol{x_0}) = - \begin{bmatrix} \frac{\partial^2 f}{\partial x^2}(\boldsymbol{x_0}) & \frac{\partial^2 f}{\partial x \partial y}(\boldsymbol{x_0}) & \frac{\partial^2 f}{\partial x \partial s}(\boldsymbol{x_0}) \\ \frac{\partial^2 f}{\partial x \partial y}(\boldsymbol{x_0}) & \frac{\partial^2 f}{\partial y^2}(\boldsymbol{x_0}) & \frac{\partial^2 f}{\partial y \partial s}(\boldsymbol{x_0}) \\ \frac{\partial^2 f}{\partial x \partial s}(\boldsymbol{x_0}) & \frac{\partial^2 f}{\partial y \partial s}(\boldsymbol{x_0}) & \frac{\partial^2 f}{\partial s^2}(\boldsymbol{x_0}) \end{bmatrix}^{-1} \begin{bmatrix} \frac{\partial f}{\partial x}(\boldsymbol{x_0}) \\ \frac{\partial f}{\partial y}(\boldsymbol{x_0}) \\ \frac{\partial f}{\partial s}(\boldsymbol{x_0}) \end{bmatrix}. \tag{5}$$

Let (x_n, y_n, s_n) denote the sampled point whose position is n sampling steps from $\boldsymbol{x_0}$ along the x direction, n sampling steps from $\boldsymbol{x_0}$ along the y direction,

and n scale levels away from $\boldsymbol{x_0}$. The partial derivatives, for example, $\frac{\partial f}{\partial x}(\boldsymbol{x_0})$, $\frac{\partial^2 f}{\partial x^2}(\boldsymbol{x_0})$, $\frac{\partial^2 f}{\partial x \partial y}(\boldsymbol{x_0})$, can be evaluated by

$$
\begin{aligned}
\frac{\partial f}{\partial x}(\boldsymbol{x_0}) &= \frac{f(x_1,y_0,s_0)-f(x_{-1},y_0,s_0)}{2} \\
\frac{\partial^2 f}{\partial x^2}(\boldsymbol{x_0}) &= f(x_1,y_0,s_0)+f(x_{-1},y_0,s_0)-2f(x_0,y_0,s_0) \\
\frac{\partial^2 f}{\partial x \partial y}(\boldsymbol{x_0}) &= \frac{f(x_1,y_1,s_0)-f(x_{-1},y_1,s_0)-f(x_1,y_{-1},s_0)+f(x_{-1},y_{-1},s_0)}{4}.
\end{aligned}
\tag{6}
$$

The located maximum is accepted if the extremum $\hat{\Delta}\boldsymbol{x}$ of the offset is less than 0.5 in all the dimensions. The final interpolated estimate for the location of the maximum is obtained by adding the extremum offset $\hat{\Delta}\boldsymbol{x}$ to the location of the maximum.

After the interest points are detected and their locations (in x, y directions and scale) in the scale space are interpolated, each interest point is assigned a reproducible orientation by which a square region is constructed, and the SURF descriptor is extracted from it. The orientation is assigned based on the information from a circular region around the interest point. The radius of the region is $6s$, s being the scale at which the interest point was detected. The Haar wavelet responses in x and y directions are calculated for each point in the circle region, with sampling step s. The side length of the wavelet is set to $4s$. The responses of the wavelets are then weighted with a Gaussian (standard deviation $2.5s$) centred at the interest point. The results are interpreted as points in a vector space. Then a sliding window of size $\frac{\pi}{3}$ is rotated around the origin. Within the area covered by the sliding window, the x and y responses are summed to yield a local orientation vector. The orientation of the detected interest point is the orientation of the longest among all such vectors (see Figure 2).

The final step to extract the SURF descriptor is to construct a square window around the interest point. The size of the window is $20s$, s being the scale, and it is aligned with the orientation of the interest point. The window is divided
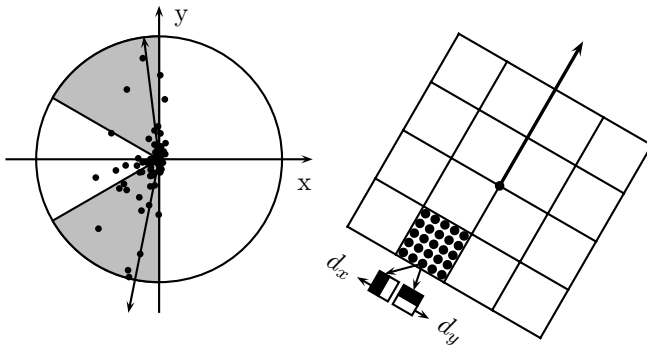
Fig. 2. Left: Assign orientation to the interest point. The data fall into the range $(-0.1, 0.1)$. The detected interest point was at $(133.0, 211.6, 1.92)$ in the scale space. Right: Extract the SURF descriptor from the square window constructed and aligned with the orientation of the interest point.

into 4×4 regular sub-regions, from each of which 5×5 regular spaced points are sampled. For each sampled point the Haar wavelet (side $2s$) responses are calculated and then weighted by a Gaussian centred at the sampled point with standard deviation $3.3s$. If d_x and d_y are used to denote the Gaussian weighted Haar wavelet responses in x and y directions, for each of the 4×4 sub-regions, $\sum d_x, \sum |d_x|, \sum d_y, \sum |d_y|$ are used to denote the sums of the x and y responses of all the 5×5 sampled points. And for all the 4×4 sub-regions, there are totally 64 entries which form the SURF descriptor (see Figure 2).

3   The Parallel Algorithm

The multi-threaded algorithm parallelises the computations in the scale-space analysis, the localisation of the interest points, the interpolation of the interest points, and the description, including the orientation assignment and the extraction of the SURF descriptor, shown in Figure 3.
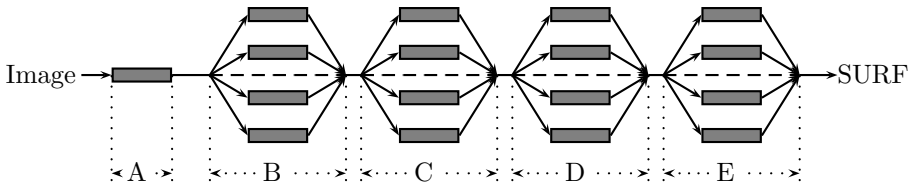
Fig. 3. The parallelised stages of the algorithm, where A stands for the integration, B for the scale-space analysis, C for the localisation, D for the interpolation, and E for the description

For a scale-space analysis with o octaves and i intervals in each octave, it is natural to think about creating $o \times i$ threads and assigning each thread a scale level (o_j, i_k), $j \in [0, o)$ and $k \in [0, i)$, to work with. But that will fix the number of the threads to be created. Instead, to make the system more flexible, the program allows the user to specify the maximum number, $t_{\max}$, of threads that will be created in the computations from the command line, and the program will dynamically adjust the number of threads to be created in each parallelised stage of the execution. The number of threads, t_s, to be created for the scale-space analysis is determined by Equation 7.

$$t_s = \begin{cases} o \times i & : & t_{\max} > o \times i \\ t_{\max} & : & t_{\max} \leq o \times i \end{cases} \tag{7}$$

Before creating the working threads, an array of $o \times i$ places is set up, so that each place of the array holds a pointer pointing to a group of arguments, including, among others, a scale level (o_j, i_k), $j \in [0, o)$ and $k \in [0, i)$. The ith created working thread initially will find the ith place of the array, get the group of arguments and do the corresponding computation at the assigned scale level. Once it is finished, the thread will get to the $(i + t_s)$th place of the array
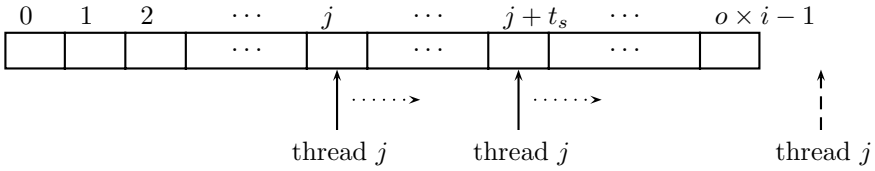
Fig. 4. Thread j will get arguments from the places j, $j + t_s$, $j + 2t_s$, ..., and do the corresponding computations, until the index moves out of the array

and do likewise until the updated index exceeds the length of the array (see Figure 4). The scale-space analysis is done when all the threads have finished their computations and joined the main thread.

To localise the maxima, searchings are made at all the scale levels except the bottom and top levels in each octave. This means the total number of the scale levels at which the searchings will be made is $o \times (i - 2)$. The number of working threads, t_l, to be created for the searchings is likewise determined by Equation 7, except substituting $o \times (i - 2)$ for $o \times i$.

It can be that each working thread is assigned a scale level and made searching for the maxima, and once finished, doing the interpolation subsequent to it. But that will make the work load for each thread unbalancing, because the number of maxima detected decreases rapidly as the scale level increases.

Therefore, in the design, the task of interpolation has been separated from the localisation. Only after all the maxima have been found and collected does the interpolation begin so that the maxima are distributed evenly among working threads.

In preparing for creating the threads for the localisation, like the preparation for the scale-space analysis, an argument array of length $o \times (i - 2)$ is set up. Each place of the array holds a group of arguments for a thread to be working at a specific scale level (o_j, i_k), $j \in [0, o)$ and $k \in [1, i - 1)$. The ith thread created finds the ith, $(i + t_l)$th, $(i + 2t_l)$th, ... places of the array and does the corresponding computations until the updated index exceeds the length of the array. Among the arguments passed to a thread, there is a `head` and a `tail` pointer. The maxima found by the thread are joined into a linked list. The `head` pointer points to the head of the list and `tail` points to the tail. After all the working threads finished their job, the main thread will scan through the argument array, joining the `tail` held by the i place to the `head` held by the $(i + 1)$ place, by-passing any place where the pointers are null, to form a whole list of all the detected maxima. The `head` pointer held by the first non-null place becomes the head of the maximum list. This is illustrated in Figure 5.

After the list of all the detected maxima is formed, the maxima are distributed evenly among the threads that is going to be created for the interpolation. The number of threads, t_i, to be created for the interpolation is jointly determined by $t_{\max}$ and the number of the maxima, m, in the list, with m replacing $o \times i$ in Equation 7. Each thread will get $\frac{m}{t_i}$ maxima to interpolate except the last created thread. Like before, an argument array is set up, but with t_i places, and
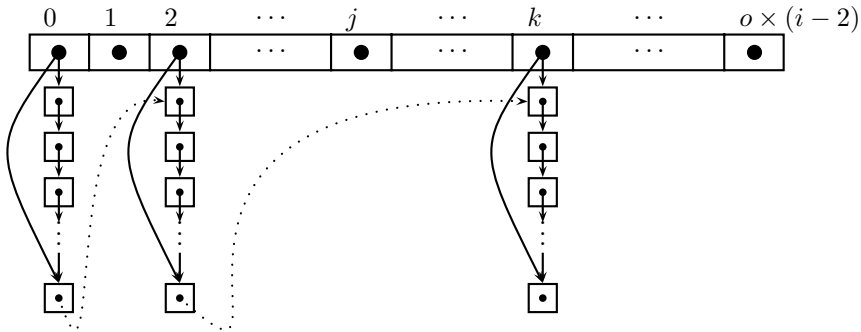
Fig. 5. An illustration of joining the lists of maxima found by the threads. The `head` and the `tail` pointer in the places 1, j and $o \times (i-2)$ are null.

each place holds the group of arguments, including a `head` and a `tail` pointer. The ith created thread, after collecting the arguments, will work from the $i \times \frac{m}{t_i}$ position of the list of the maxima to the $(i+1) \times \frac{m}{t_i}$ position if the thread is not last created. The thread created last will have to work to the end of the list. Each thread will form a linked list of interpolated interest points, and after they all have finished, the main thread will join the lists to form a list containing all the detected and interpolated interest points.

The list of the interest points is then passed to the threads created for the description, including the orientation assignment and the extraction of SURF, with the same manner as in the stage of the interpolation.

4   Tests and Results

Tests have been made to compare the performance of the implementation of P-SURF[1] with that of the sequential SURF[2]. The original implementation[3] of the SURF was not used as the reference for it was close sourced.

The performance comparisons were made on three systems. Relevant configurations of the three systems are summarised in Table 2.

Nine randomly selected PGM images were used for the tests, whose dimensions are reported in Table 1. In all the tests the number of octaves was set to 3, the number of intervals 4, the threshold 0.0004, the initial sample 1, the maximum number of threads to be created 12. All the time durations were measured in milliseconds (ms). The compiler used was g++ 4.3.2. The programs were compiled with the -O3 option. The results are reported in Table 3.

From the data it can be seen that the speed-up ratios that P-SURF gained over the sequential SURF were roughly 2 to 6 times on multi-core processors, depending on the configuration of the system and the dimension of the image.

---

[1] P-SURF 0.1, found at `http://www.xjtlu.edu.cn/depts/csse/csse1511`

[2] OpenSURF-1.2, found at `http://www.jmkirby.co.uk/`

[3] SURF-1.0.9, found at `http://www.vision.ee.ethz.ch/~surf/index.html`

Table 2. The testing systems

System	Processor model	#Processors	Hyper-threading	Kernel
1	Core Duo P8600@2.4G	1(dual-core)	No	2.6.27
2	Xeon MV5063@3.2G	1(dual-core)	Yes	2.6.9
3	Xeon E5405@2.0G	2(quad-core)	No	2.6.27

Table 3. Performance comparisons of P-SURF and SURF

Image	#Points	System 1		System 2		System 3	
		P-SURF	SURF	P-SURF	SURF	P-SURF	SURF
1	375	89	124	80	168	60	155
2	342	122	156	116	220	77	195
3	728	203	287	197	406	126	360
4	1676	376	612	378	853	208	750
5	1727	488	741	490	1069	275	937
6	936	482	713	545	1056	259	900
7	165	532	825	700	1245	308	1046
8	1728	1174	1836	1351	2673	611	2320
9	30152	5334	9487	5642	13535	2109	11954

For large images the ratios tended to be higher than those gained for small images.

5   Conclusion and Future Work

This paper presented a parallel algorithm which computes the SURF descriptor, and two groups of test results. Generally, the computational costs of local descriptors, such as SURF and SIFT, are much higher than that of global descriptors. However, local descriptors are able to capture correspondences between images even facing resizing and rotation, when global descriptors will fail to do. The improvement gained by SURF over SIFT is the speed-up of the computation. The work of the design of the P-SURF and the implementation further advanced this trend of research by computing the SURF parallelly via POSIX threads. Tests have been made to show that the speed-up ratios of P-SURF over SURF were 2 to 6 times on multi-core processors. The ratios tended to be higher in the cases of large images than in the cases of small images. The number of available computing units in the system also has a major influence on the ratio.

On the basis of P-SURF, further speed-up can be gained by applying SSE instructions to the computations in the various stages, however, in the cost of sacrificing the readability of the program. The maximum number $t_{\max}$ of threads to be created also affects the computational time of P-SURF. The optimal value of $t_{\max}$, in terms of getting the shortest computational time, depends on several factors, such as the configuration of the system and the dimension of the image. For a particular system how to determine the optimal value of this number is an issue that has not been explored, which can also be a line of future work.

References

1. Bay, H., Ess, A., Tuytelaars, T., Gool, L.V.: Speeded-Up Robust Features (SURF). Computer Vision and Image Understanding (CVIU) 110(3), 346–359 (2008)
2. Lowe, D.G.: Distinctive Image Features from Scale-Invariant Keypoints. International Journal of Computer Vision 60(2), 91–110 (2004)
3. ISO/IEC/JTC1/SC29/WG11: CD 15938-3 MPEG-7 Multimedia Content Description Interface - Part 3. In: MPEG Document W3703 (2000)
4. Bay, H., Fasel, B., Gool, L.V.: Interactive Museum Guide: Fast and Robust Recognition of Museum Objects. In: The First International Workshop on Mobile Vision (May 2006)
5. Vergauwen, M., Gool, L.V.: Web-based 3D Reconstruction Service. Machine Vision and Applications 17(6), 411–426 (2006)
6. Ke, Y., Sukthankar, R., Huston, L.: An Efficient Parts-based Near-duplicate and Sub-image Retrieval System. In: Proceedings of the 12th Annual ACM International Conference on Multimedia, pp. 869–876. ACM, New York (2004)
7. Jing, Y., Baluja, S.: VisualRank: Applying Pagerank to Large-Scale Image Search. IEEE Transactions on Pattern Analysis and Machine Intelligence 30(11), 1877–1890 (2008)
8. Viola, P., Jones, M.: Rapid Object Detection using a Boosted Cascade of Simple Features. In: The 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, vol. 1, pp. I–511–I–518 (2001)
9. Evans, C.: Notes on the OpenSURF Library. Technical report, University of Bristol (January 2009), http://www.cs.bris.ac.uk/Publications/Papers/2000970.pdf
10. Brown, M., Lowe, D.G.: Invariant Features from Interest Point Groups. In: BMVC, British Machine Vision Association (2002)

A Heuristic Optimization Algorithm for Panoramic Image Generation Problem from Multiple Cameras

Megumi Isogai, Nobuo Funabiki, and Toru Nakanishi

Department of Communication Network Engineering, Okayama University
3-1-1 Tsushimanaka, Okayama 700-8530, Japan
{funabiki,nakanisi}@cne.okayama-u.ac.jp

Abstract. Recently, a panoramic image has been expected in various applications due to the advantage of expressing a wide range of scenes by one image. In this paper, we propose a heuristic optimization algorithm for the panoramic image generation problem from multiple cameras. Our three-stage algorithm composed of the approximate calibration, the detailed calibration, and the image synthesis, transforms the images of the side cameras to be fit to the image of the central camera as best as possible. The image parameters are optimized by a local search method with a Tabu period as a typical heuristic optimization method. Through experiments, we show the effectiveness of our proposal.

1 Introduction

With the advantage of expressing a wide range of scenes by one image, the *panoramic image* has the great potential of increasing the realistic sensation [1,2,3,4]. The panoramic image has been expected in various applications such as the monitoring system and the virtual reality due to the recent advancement of small inexpensive computers and the multimedia technology. Traditionally, the panoramic image has been used in the air photograph and the remote sensing.

The methods of generating panoramic images can be roughly classified into two groups. The first group uses one specialized camera with a wide-angle lens and/or one active camera, and the second one uses multiple conventional cameras with normal lens. For the first group method, because images taken by such cameras are usually distorted, algorithms for transforming them into normal images have widely been studied [5,6,7]. However, this group method has several drawbacks of the insufficient resolution due to the use of a single lens, and the nonuniformity of the resolution depending on the camera angle. Besides, the wide-angle lens and the active camera are usually very expensive, so that the use in homes is hard.

For the second group method, multiple images taken by multiple conventional cameras are synthesized into one image by applying the projective transformation to them [8,9,10,11,12]. As the number of cameras increases, the resolution increases and the resolution nonuniformity decreases, whereas the system becomes more complex. In order to obtain a reasonable panoramic image from

multiple images, algorithms of applying both the coordinate and optical trans-
formations precisely are very important. Thus, a number of studies have been
reported for improving these synthesis algorithms. To synthesize two adjacent
images into one, most of the algorithms apply the coordinate transformation to
match the coordinates of the feature points extracted from the overlapped areas
among two images. Because only the coordinates of limited points are matched,
the quality of the synthesized image can be insufficient, and the area other than
the feature points can be unclear [9].

In this paper, we propose a heuristic optimization algorithm to generate a
panoramic image from two adjacent images taken by two conventional cameras.
By increasing the number of synthesized image pairs, our algorithm can generate
a large panoramic image. Our algorithm consists of the *approximate calibration
stage*, the *detailed calibration stage*, and the *image synthesis stage*. The first stage
applies the *coordinate transformation* composed of the *rotation*, the *zooming*, and
the *translation*, so that the coordinates of the feature points selected from the
overlapped areas among two images are matched. The difference between the
coordinates comes from the difference of positions and directions of cameras.
The second stage applies the *coordinate adjustment* and the *color adjustment* so
that every pixel in the overlapped areas has the same color density. Here, by
concerning every pixel in the overlapped areas, our algorithm achieves the high
quality. For both stages, a *local search method with a Tabu period*, is used to
optimize the transformation parameters. The last stage synthesizes two images
into one rectangular panoramic image with the *graduation* and the *trimming*.

The rest of this paper is organized as follows: Section 2 describes the outline
of the panoramic image system of this paper. Sections 3-5 present three stages
of our algorithm respectively. Section 6 shows the evaluation of our proposal.
Section 7 provides the concluding remark with future works.

2 Overview of Panoramic Image System

In our panoramic image system, three cameras are set up horizontally so that
it can generate one panoramic image with a wide horizontal range. The image
of the center camera is regarded as the standard image, and the two images
of the side cameras are transformed by the algorithm so that one panoramic
image is generated from them. Here, we note that if four or more cameras are
used, distortions of the images from the end cameras become too large for a high
quality panoramic image.

Figure 1 illustrates the overview of the panoramic image system in this paper.
In addition to three cameras, one personal computer (PC) is used to synthesize
the panoramic image. For convenience, the center camera is called *camera A*,
the right one is *camera B*, and the left one is *camera C*, where the image of each
camera is called *image A*, *image B*, and *image C* respectively. In the coordinate
system, the x-axis corresponds to the horizontal line of the photographed facet
of *image A*, the y-axis does to the vertical line, and the z-axis does to the
perpendicular line. The left-upper corner represents the origin.
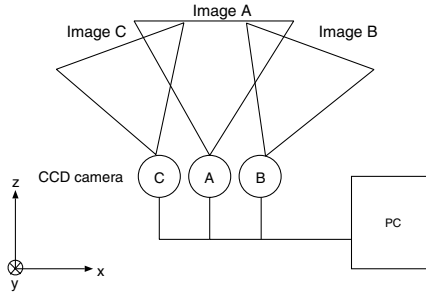
Fig. 1. Overview of panoramic image system

For the prototype system, we use PC cameras by Orange Micro Corp. [13]. This camera has the feature of small lens distortion, and can take $30fps$ images with 320×240 pixels. The cameras are fixed at the aluminum bar on top of a tripod. The direction of the side cameras is rotated by about 20 degrees from the center camera, so that they can cover a wide range of scenes horizontally.

Our algorithm is applied to *image B* and *image C* to be synthesized with *image A*. The following three sections describe the details of the three stages in our algorithm respectively. These descriptions assume the synthesis of *image A* and *image B* for simplification.

3   Approximation Calibration

The first stage transforms the coordinates of *image B* so that the feature points in the overlapped areas between *image A* and *image B* have the same coordinates.

From Figure 1, the *rotation* around the y-axis, the *zooming*, and the *translation* should be applied to match the coordinates of the feature points in *image B* to the corresponding ones in *image A*. The rotation can compensate the difference of camera directions. The zooming can compensate the difference of distances from the two cameras to objects. The translation can compensate the difference of the coordinate systems so that the origin in *image A* becomes the origin of the synthesized image. Thus, the following equation [14] is used for this stage:

$$\begin{bmatrix} x' \\ y' \end{bmatrix} = \frac{r_1 - r_2}{-(x - \frac{W}{2}) \tan \alpha - r_2} \begin{bmatrix} x - \frac{W}{2} \\ y - \frac{H}{2} \end{bmatrix} + \begin{bmatrix} \Delta x \\ \Delta y \end{bmatrix}. \tag{1}$$

where r_1 represents the distance between the photographed facet for *image A* and that for *image B*, r_2 does the vanishing point, α does the rotation angle around the y-axis, Δx and Δy does the translation distance along the x-axis and the y-axis respectively, and W and H does the width (= 320 pixels) and the height (= 240 pixels) of the image.

A calibration board with five feature points is used to extract the feature points for the approximation calibration. Up to three calibration boards are set
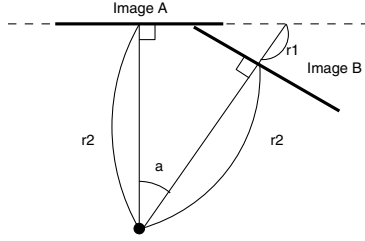
Fig. 2. Relative positions between *image A* and *image B*

around the overlapped area between the two images in our experiments. Here, we note that if the calibration board cannot be installed, the feature points should be extracted from the overlapped areas manually.

The parameters for the approximation calibration in (1) are optimized through a heuristic optimization algorithm. Thus, the corresponding optimization problem is defined:

Input: The coordinates of the feature points in *image A* and *image B*.
Output: r_1, α, Δx, and Δy in (1), where r_2 is fixed to 250.
Objective: To minimize the following cost function E:

$$E = \sum_i \left((x_i^A - x_i^{B'})^2 + (y_i^A - y_i^{B'})^2 \right), \tag{2}$$

where (x_i^A, y_i^A) represents the coordinates of the ith feature point in *image A*, and $(x_i^{B'}, y_i^{B'})$ represents the coordinates of the ith feature point in *image B'* that is transformed from *image B* by the approximation calibration.

This paper adopts a local search method with a Tabu period as a high quality heuristic algorithm [15]. The procedure of this algorithm is described in Appendix A. The initial values for the parameters are set as follows:

α: 20 is used as the camera setup.
r_1: The difference of distances to the photographed facets between *image A* and *image B*, which can be obtained from the following equation with $r_2 = 250$:

$$r_1 = r_2 - \frac{r_2}{\cos \alpha}. \tag{3}$$

$\Delta x, \Delta y$: The difference of the coordinates of the center feature point, called the *standard feature point*, is used.

$$\begin{bmatrix} \Delta x \\ \Delta y \end{bmatrix} = \begin{bmatrix} x_c^A \\ y_c^A \end{bmatrix} - \begin{bmatrix} x_c^B \\ y_c^B \end{bmatrix} \tag{4}$$

where (x_c^A, y_c^A) represents the coordinate of the standard feature point in *image A* and (x_c^B, y_c^B) does the one in *image B*.

Using the coordinate transformation equation in (1) with optimized parameters by our algorithm, the color density of each pixel in the transformed image (*image B'*) is calculated from the color density of the corresponding pixel in *image B*. Although every pixel must have integer coordinates in any digital image, the equation (1) usually outputs real coordinates. Thus, we apply the following *bi-linear interpolation* using the four neighbor pixels:

$$\boldsymbol{f^{B'}}(x', y') = (1-\tau)(1-\upsilon)\boldsymbol{f^B}(\lfloor x \rfloor, \lfloor y \rfloor) + \tau(1-\upsilon)\boldsymbol{f^B}(\lfloor x \rfloor + 1, \lfloor y \rfloor) \quad (5)$$
$$+ (1-\tau)\upsilon\boldsymbol{f^B}(\lfloor x \rfloor, \lfloor y \rfloor + 1) + \tau\upsilon\boldsymbol{f^B}(\lfloor x \rfloor + 1, \lfloor y \rfloor + 1).$$

where (x', y') represents a pixel coordinate in *image B'*, (x, y) does the corresponding one in *image B*, $\boldsymbol{f^{B'}}(x', y')$ does the color density after interpolation, $\boldsymbol{f^B}(\lfloor x \rfloor, \lfloor y \rfloor), \boldsymbol{f^B}(\lfloor x \rfloor + 1, \lfloor y \rfloor), \boldsymbol{f^B}(\lfloor x \rfloor, \lfloor y \rfloor + 1), \boldsymbol{f^B}(\lfloor x \rfloor + 1, \lfloor y \rfloor + 1)$ does the color density of the left upper pixel, the right upper, the left lower, and the right lower of the pixel with (x, y) respectively, τ, υ does the distance along the x-axis and the y-axis between the left upper pixel with $(\lfloor x \rfloor, \lfloor y \rfloor)$ and the pixel with (x, y) respectively.

4   Detailed Calibration

The second stage adjusts the coordinates and the color densities so that every pixel in the overlapped areas between *image A* and the transformed image from *image B'* (let *image B"*) has the same color density.

The coordinate transformation in the detailed calibration applies the rotation around the z-axis, and the adjustments of the zooming and the translation. For higher quality, the zooming adjustment is applied along the x-axis and the y-axis independently. Then, the transformation equation is given by:

$$\begin{bmatrix} x' \\ y' \end{bmatrix} = \begin{bmatrix} k_x & 0 \\ 0 & k_y \end{bmatrix} \left\{ \begin{bmatrix} \cos\beta & \sin\beta \\ -\sin\beta & \cos\beta \end{bmatrix} \begin{bmatrix} x - x_0 \\ y - y_0 \end{bmatrix} + \begin{bmatrix} x_0 \\ y_0 \end{bmatrix} \right\} + \begin{bmatrix} \delta x \\ \delta y \end{bmatrix}, \quad (6)$$

where β represents the rotation angle around the z-axis, k_x, k_y does the zooming rate along the x-axis and the y-axis respectively, and $\delta x, \delta y$ does the translation distance along the x-axis and the y-axis respectively.

This paper adopts the *RGB color system* to represent the color density. Thus, each color density of R, G, and B should be adjusted independently. For this purpose, we adopt the following equation for this color density adjustment:

$$\boldsymbol{f^{B''}}(x, y) = K \cdot \boldsymbol{f^{B'}}(x, y), \quad (7)$$

where $\boldsymbol{f^{B'}}(x, y), \boldsymbol{f^{B''}}(x, y)$ represents the color density of the pixel with (x, y) in *image B'* and *image B"*, and K does the zooming coefficient matrix given by:

$$K = \begin{bmatrix} K_r & 0 & 0 \\ 0 & K_g & 0 \\ 0 & 0 & K_b \end{bmatrix}, \quad (8)$$

where K_r, K_g, K_b represents the zooming coefficient for R, G, and B respectively.

Although the shape of the overlapped area between two images becomes trapezoid due to the transformation in the first stage, we regard it as rectangular to speed up the calculation in the following parameter optimization algorithm. Then, the range of this area can be given by:

$$\Delta x + \delta x \leq x \leq W, y_u^{B''} \leq y \leq y_b^{B''}, \tag{9}$$

where $y_u^{B''}$ and $y_b^{B''}$ represents the y-coordinate of the left upper corner and the left lower corner in *image B"*. We note that the cameras are set up so that these points are located inside *image A*.

In order to optimize the parameters in (6) and (8), the corresponding optimization problem is defined:

Input: Color densities of *image A* and *image B'*: $\boldsymbol{f^A}(x, y), \boldsymbol{f^{B'}}(x, y)$.
Output: The eight parameters in (6) and (8).
Objective: To minimize the following cost function:

$$E = \sum_{(x,y) \in S} \left(\boldsymbol{f^A}(x, y) - \boldsymbol{f^{B''}}(x, y) \right)^2 / |S|, \tag{10}$$

where $|S|$ represents the size of the overlapped area S.

The same method as the first stage is used to optimize the parameters. The initial values for the rotation and the translation are set 0, and those for the zooming and the color density adjustments are set 1. Then optimized, the bilinear interpolation is applied to obtain *image B"* from *image B'*.

5   Image Synthesis

After the two-stage calibrations, the obtained *image B"* and *image A* are combined together into one panoramic image (let *image D*) by applying the *graduation* and the *trimming*.

The boundary between *image A* and *image B"* should disappear in the combined image as much as possible. Thus, the following graduation is applied to obtain the combined image:

$$\boldsymbol{f^D}(x, y) = (1 - \kappa(x)) \cdot \boldsymbol{f^A}(x, y) + \kappa(x) \cdot \boldsymbol{f^{B''}}(x, y), \tag{11}$$

where $\boldsymbol{f^D}(x, y)$ represents the color density of the pixel with (x, y) in *image D*, and $\kappa(x)$ does the ratio of combining two color densities of the corresponding pixels in *image B"* and *image A*. $\kappa(x)$ is proportional to the distance from the boundary that is given by:

$$\kappa(x) = \frac{1}{W - \Delta x - \delta x} \{ x - (\Delta x + \delta x) \}. \tag{12}$$

The final image should be rectangular, assuming it is obtained from one wide angle camera. Thus, we remove unnecessary parts as trimming.

6   Evaluations

In order to evaluate our proposal, we have implemented the three-stage algorithm on a PC with WindowsXP, Pentium4 (3GHz), and 2Gbyte memory, using C++ and Microsoft VisualStudio.NET Professional.

The three cameras with $100mm$ intervals are fixed on a tripod at the height of about $1m$ from the ground using an aluminum bar. The direction of each side camera is differentiated outward by about 20 degrees from the center one. Then, about 1/6 of each image becomes the overlapped area between two adjacent images. Three calibration boards are put inside of each overlapped area to extract the feature points.

The experimented day had the fine weather. Starting from the noon, we took the images of the same scene in our office in every two-hour. The calibration parameters are optimized using the images taken at noon. Then, the same set of parameters are used in the following panoramic image generations.

Effectiveness of Two-Stage Calibration. Figure 3 shows the input images for the algorithm and its output images. Among three calibration boards, we used only the center one at this time. Tables 1 and 2 summarize the obtained parameter values by our algorithm. The CPU time is 56 seconds for the approximation calibration, and 127 seconds for the detailed calibration. Because these calibrations should be finished before starting the system, the CPU time is acceptable.

The image after the approximation calibration in Fig. 3 (b) has discontinuity of objects, such as the ceiling and the fan on the wall, at the boundary of the two input images at its upper or lower sides. Note that the feature points are matched with each other, which means that this stage can well match the coordinates of feature points on calibration boards. Besides, the color density is also inconsistent at the image boundary. Then, the synthesized image after the detailed calibration and the image synthesis in Fig. 3 (c) has resolved these problems and become very smooth, by matching the color density of every pixel in the overlapped areas. Therefore, our three-stage algorithm is very effective to generate a high quality panoramic image.

Effect of Calibration Board. In order to evaluate the effect of the number and positions of calibration boards, we change the use of calibration boards in

Table 1. Parameters for approximate calibration

α	r_1	Δx	Δy
-31	-48	288	4

Table 2. Parameters for detailed calibration

k_x	k_y	δx	δy	β	K_r	K_g	K_b
1.39	1.14	-6	-2	0	0.89	0.87	0.92

(a) Input images

(b) Synthesized image after approximate calibration

(c) Synthesized image after detailed calibration
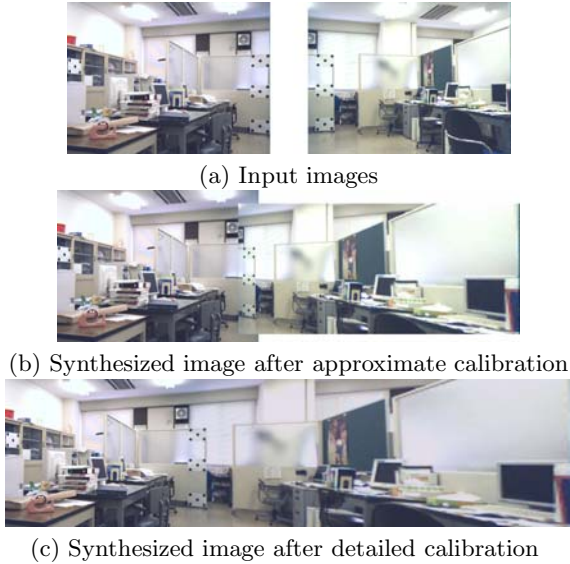
Fig. 3. Effectiveness of two-stage calibrations

the algorithm; the upper board only, the lower board only, the upper and lower boards, and the three boards. Then, we found that the synthesized image quality after the detailed calibration is not much different among them, except for the case of using the lower board only, and the CPU time is similar. Therefore, the center calibration board is enough to optimize the calibration parameters.

Validity of Calibration Parameters by Elapsed Time. Figure 4 the panoramic images synthesized from three input images taken at noon, and 6pm, using the same set of parameters. Although the outdoor becomes dark as the time elapses, any panoramic image keeps the integrity in terms of object

(a) Image for calibration (at noon)

(d) Image after 6 hours (at 6pm)
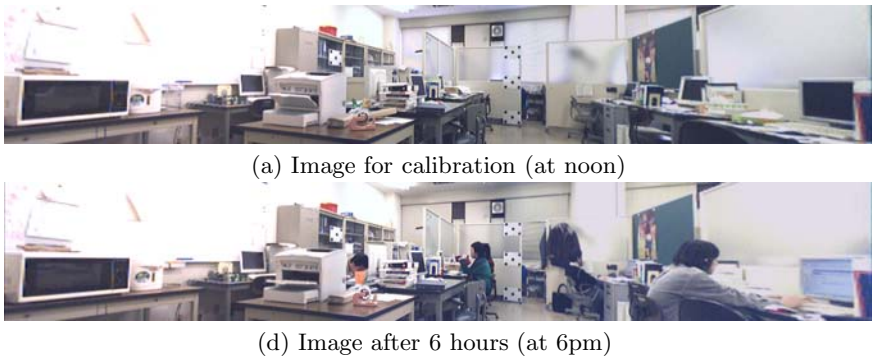
Fig. 4. Effectiveness of calibration parameters with time elapse

positions and color densities. Therefore, our algorithm is effective with the same set of parameters for the change of illumination.

7   Conclusion

This paper has presented a heuristic optimization algorithm for the panoramic image generation problem from multiple cameras arranged horizontally. This algorithm consists of three stages; the approximation calibration, the detailed calibration, and the image synthesis, to obtain a high quality panoramic image. The parameters for both calibrations are optimized through a local search method with a Tabu period. The experimental results confirm the effectiveness of our proposal. Our future works may include the expansion of our approach to combine multiple images arranged vertically.

References

1. Aizawa, K., Sakaue, K., Suenaga, Y.: Image Processing Technologies: Algorithms, Sensors, and Applications. CRC Press, Boca Raton (2004)
2. Benosman, R., Kang, S.B., Faugeras, O.: Panoramic Vision: Sensors, Theory, and Applications. Springer, Heidelberg (2001)
3. Kang, S.B.: A Survey of Image-based Rendering Techniques. Tech. Paper. Cambridge Research Lab (1997)
4. Gledhilla, D., Tian, G.Y., Taylora, D., Clarke, D.: Panoramic Imaging: a Review. Computers & Graphics 27(3), 435–445 (2003)
5. Kawanishi, T., Yamazawa, K., Iwasa, H., Takemura, H., Yokoya, N.: Generation of High-resolution Stereo Panoramic Images by Omnidirectional Imaging Sensor Using Hexagonal Pyramidal Mirrors. In: Proc. Int. Conf. Pattern Recognition, vol. 1, pp. 485–489 (1998)
6. Yagi, Y., Yachida, M.: Real-time Omnidirectional Image Sensors. Int. J. Comput. Vision, 173–207 (2004)
7. Tokai, S., Yumiba, R., Wu, X., Matsuyama, T.: Dynamic Wide-area Scene Visualization Using an Active Camera. Systems and Computers in Japan 37(8), 101–112 (2006)
8. Szeliski, R., Shum, H.Y.: Creating Full View Panoramic Image Mosaics and Environment Maps. In: Proc. Comput. Graphics Annu. Conf. Ser., pp. 251–259 (1997)
9. Laar, V., Aizawa, K.: Capturing Wide-view Images with Uncalibrated Cameras. IEICE Trans. Inf.& Syst. E83-D(4), 895–903 (2000)
10. Nielsen, F.: Randomized Adaptive Algorithm for Mosaicing System. IEICE Trans. Inf.& Syst. E83-D(7), 1386–1394 (2000)
11. Ikeda, S., Sato, T., Yokoya, N.: High-resolution Panoramic Movie Generation from Video Streams Acquired by an Omnidirectional Multi-camera System. In: Proc. Int. Conf. Multisensor Fusion and Integration for Intelligent Systems (MFI 2003), pp. 155–160 (2003)
12. Guandong, G., Kebin, J.: A New Image Mosaics Algorithm Based on Feature Points Matching. In: Proc. ICICIC (2007)
13. http://www.orangemicro.com/ibot2.html
14. Ozaki, R., Sato, Y., Sugiyama, Y., Okamoto, T.: A Fast Method for Coordinate Calculation in Projective Transformations. IEICE Trans. J88-D-II(1), 1–7 (2005)

15. Ibaraki, T., Nonobe, K., Yagiura, M.: Metaheuristics, Progress as Real Problem Solvers. Springer, Heidelberg (2005)
16. Sekioka, T., Yokogawa, Y., Funabiki, N., Higashino, T., Yamada, T., Mori, Y.: A Proposal of a Lip Contour Approximation Method Using the Function Synthesis. IEICE Trans. J84-D-II(3), 459–470 (2001)

A   Parameter Optimization Algorithm

This appendix describes the procedure of the local search method with a Tabu period [16] in this paper.

(1) Initialize the parameters by their defined ones, and the *Tabu period* for every parameter and the *iteration* by 0.
(2) Calculate the cost function E using the initial parameters, and set the cost function of the *interim solution* by $E_{best} = E$.
(3) Select one parameter with the *zero Tabu period*, randomly.
(4) Calculate the cost functions E_+ and E_- using the parameters when only the parameter selected in (3) is increased or decreased by its defined *variation width*.
(5) Change the parameter to the one that minimizes the cost function among E, E_+, and E_-. If E is minimum, the defined *Tabu length* is set to the *Tabu period* for this parameter, because its change cannot improve the solution currently. Otherwise, the minimum cost function is copied to E.
(6) Apply the following hill-climbing procedure if every parameter has the *non− zero Tabu period*:
  1) If $E < E_{best}$, update E_{best} by E, and memorize the parameters.
  2) Increase the *iteration* by 1.
    (a) If the *iteration* does not exceed its defined upper limit, increase or decrease randomly the value of one randomly selected parameter with twice of its variation width, and initialize the *Tabu period* for every parameter by 0.
    (b) Otherwise, output the interim parameters, and terminate the algorithm.
(7) Decrease any positive *Tabu period* by 1, and return to (3).

The *iteration* is set 5 in the approximation calibration, and 10 in the detailed calibration. The *Tabu length* is set 5. The *variation width* is set 1 for any parameter in the approximation calibration, and is set 0.01 for k_x, k_y, K_r, K_g, K_b and 1 for $\delta x, \delta y, \beta$ in the approximation calibration.

A New Method for Iris Recognition Based on Contourlet Transform and Non Linear Approximation Coefficients

Amir Azizi[1] and Hamid Reza Pourreza[2]

[1] Islamic Azad University Mashhad Branch,
[2] Ferdowsi University of Mashhad
Amirazizi_b@Yahoo.com, hpourreza@um.ac.ir

Abstract. In different methods of Biometrics, recognition by iris images in recent years has been taken into consideration by researchers as one of the common methods of identification like passwords, credit cards or keys. Iris recognition a new biometric technology has great advantages such as variability, stability and security. In this paper we propose a new feature extraction method for iris recognition based on contourlet transform. Contourlet transform captures the intrinsic geometrical structures of iris image. It decomposes the iris image into a set of directional sub-bands with texture details captured in different orientations at various scales so for reducing the feature vector dimensions we use the method for extract only significant bit and information from normalized iris images. In this method we ignore fragile bits. At last the feature vector is approximated by non linear approximation coefficient. Experimental results show that the proposed method reduces processing time and increase the classification accuracy and outperforms the wavelet based method.

Keywords: Biometric, Iris Recognition, Contourlet Transform, Wavelet Transform.

1   Introduction

The purpose of 'Iris Recognition', a biometrical based technology for personal identification and verification, is to recognize a person from his/her iris prints. In fact, iris patterns are characterized by high levels of stability and distinctiveness. Each individual has a unique iris (see Figure 1); the difference even exists between identical twins and between the left and right eye of the same person. Various iris recognition methods have been proposed for automatic personal identification and verification. In Fig.2 you can see the typical stages of Iris Recognition system. Daugman first presented a prototype system [1, 2] for iris recognition based on multi-scale Gabor wavelets. Wildes presented another iris recognition system [3] in which the iris pattern was decomposed into multiresolution pyramid layers using wavelet transform. Both systems of Daugman and Wildes employed carefully designed image acquisition devices to get equal high quality iris images. Tan et al. presented a shift-invariant method [4] which decomposed the iris pattern into multiple bands using a two-dimensional Gabor filter. Boles decomposed one-dimensional intensity signals computed on circles [5] in

the iris and use zero-crossings of the decomposed signals for the feature representation. The number of zero-crossings can differ among iris image samples of an identical iris due to noises. This method was improved[6,7] in which it was assumed that if two samples were acquired from an identical iris the distances between corresponding pairs of zero-crossing in one sample and another were less than given threshold value. However, the spurious zero-crossing points could degrade the performance.
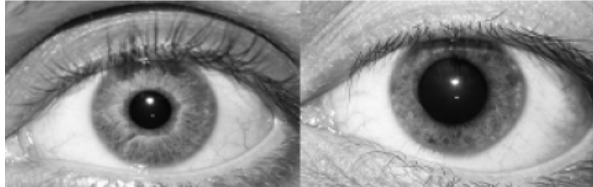
Fig. 1. Distinctiveness of human iris

A well-established fact that the usual two-dimensional tensor product wavelet bases are not optimal for representing images consisting of different regions of smoothly varying grey-values separated by smooth boundaries. This issue is addressed by the directional transforms such as contourlets, which have the property of preserving edges. The contourlet transform is an efficient directional multiresolution image representation which differs from the wavelet transform. The contourlet transform uses non-separable filter banks developed in the discrete form; thus it is a true 2D transform, and overcomes the difficulty in exploring the geometry in digital images due to the discrete nature of the image data. The remainder of this paper is organized as follows: Section 2 deals with Iris Recognition System overview. Section 3 deals with Experimental results and discussion. Section 4 concludes this paper.

An iris image, as shown in Fig.3a, contains not only the region of interest (iris) but also some 'unuseful' parts (e.g. eyelid, pupil etc.). In addition, a change in the camera-to-eye distance may result in the possible variation in the size of the same iris. Furthermore, the brightness is not uniformly distributed because of non-uniform
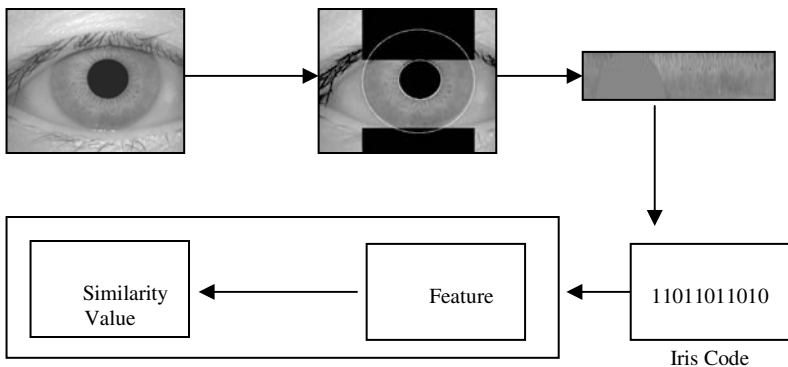
Fig. 2. Typical stages of iris recognition

illumination. Before extracting features from the original image, the image needs to be preprocessed to localize iris, normalize iris, and reduce the influence of the factors mentioned above. Such preprocessing is described in the following subsections.

2   Iris Recognition System Overview and Proposed Method

In this section we first overview the iris recognition system and then describe proposed method for iris feature extraction.

2.1   Iris Localization (Segmentation)

Both the inner boundary and the outer boundary of a typical iris can approximately be taken as circles. However, the two circles are usually not co-centric. The iris is localized in two steps: (1) approximate region of iris in an image can be found by projecting iris image in horizontal and vertical direction. (2) The exact parameters of these two circles are obtained by using edge detection and Hough transform in a certain region determined in the first step. An example of iris localization is shown in Fig.3b.

2.2   Iris Normalization

Irises from different people may be captured in different size, and even for the iris from the same person, the size may change because of the variation of the illumination and

(a)   (b)

(c)
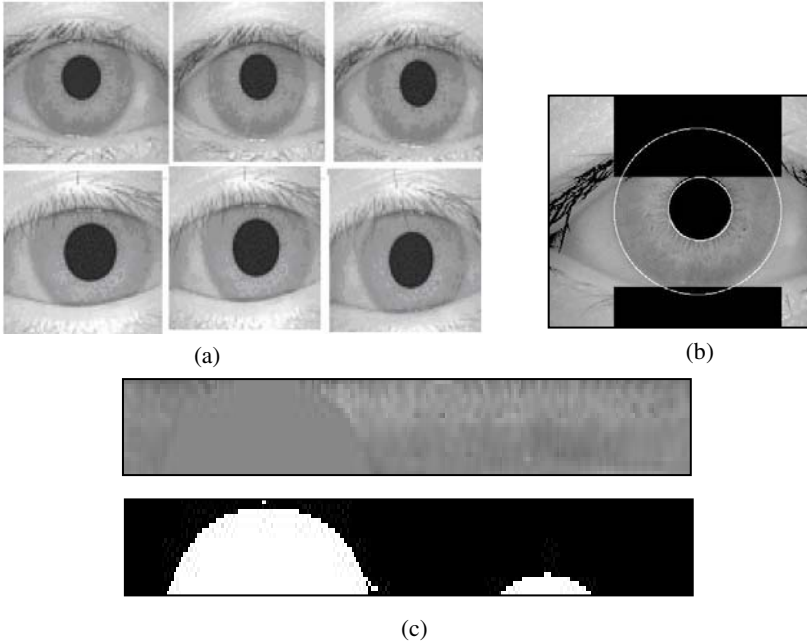
Fig. 3. (a): Iris image sample (CASIA version 1), (b): Localized (Segmented) iris by Duagman method and (c): Normalized Iris

other factors. Such elastic deformations in iris texture affect the results of iris matching. For the purpose of achieving more accurate recognition results, it is necessary to compensate for these deformations. Here, we anti-clockwise unwrap the iris ring to a rectangular block of texture of a fixed size (20x240) by piecewise linear mapping. The distortion of the iris caused by pupil dilation can thus be reduced. The result after iris normalization is shown in Fig.3c.

2.3   Feature Extraction in the Proposed Method

Only the significant features of the iris must be encoded so that comparisons between templates can be made. Gabor filter and wavelet are the well-known techniques in texture analysis [1], [2], [8], [9], [10]. In wavelet family, Haar wavelet [11] was applied by Jafer Ali to iris image and they extracted an 87-length binary feature vector. The major drawback of wavelets in two-dimensions is their limited ability in capturing directional information. The contourlet transform is a new extension of the wavelet transform in two dimensions using multiscale and directional filter banks. The feature representation should have information enough to classify various irises and be less sensitive to noises. Also in the most appropriate feature extraction we attempt to extract only significant information, more over reducing feature vector dimensions , the proceed lessened and enough information to introduce iris feature vectors classification.

2.3.1.1   Contourlet Transform. Contourlet transform (CT) allows for different and flexible number of directions at each scale. CT is constructed by combining two distinct decomposition stages [12], a multiscale decomposition followed by directional decomposition. The grouping of wavelet coefficients suggests that one can obtain a sparse image expansion by applying a multi-scale transform followed by a local directional transform. It gathers the nearby basis functions at the same scale into linear structures. In essence, a wavelet-like transform is used for edge (points) detection, and then a local directional transform for contour segments detection. A double filter bank structure is used in CT in which the Laplacian pyramid (LP) [13] is used to capture the point discontinuities, and a directional filter bank (DFB) [14] to link point discontinuities into linear structures. The combination of this double filter bank is named pyramidal directional filter bank (PDFB) as shown in Fig. 4.
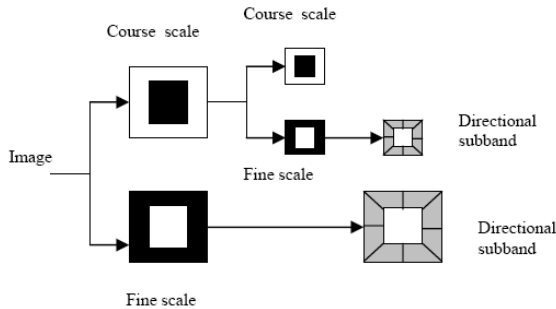
Fig. 4. Two level contourlet decomposition

2.3.1.2   Powers of Contourlet Transform in the Iris Feature Extraction. To capture smooth contours in images, the representation should contain basis functions with variety of shapes, in particular with different aspect ratios. A major challenge in capturing geometry and directionality in images comes from the discrete nature of the data, the input is typically sampled images defined on rectangular grids. Because of pixelization, the smooth contours on sampled images are not obvious. For these reasons, unlike other transforms that were initially developed in the continuous domain and then discretized for sampled data, the new approach starts with a discrete-domain construction and then investigate its convergence to an expansion in the continuous-domain. This construction results in a flexible multi-resolution, local, and directional image expansion using contour segments. Directionality and anisotropy are the important characteristics of contourlet transform. Directionality indicates that having basis function in many directions, only three direction in wavelet. The anisotropy property means the basis functions appear at various aspect ratios where as wavelets are separable functions and thus their aspect ratio is one. Due to this properties CT can efficiently handle 2D singularities, edges in an image. This property is utilized in this paper for extracting directional features for various pyramidal and directional filters.

2.3.1.3   The Best Bits in an IRIS Code. Iris biometric systems apply filters to iris images to extract information about iris texture. Daugman's approach maps the filter output to a binary iris code. The fractional Hamming distance between two iris codes is computed and decisions about the identity of a person are based on the computed distance. The fractional Hamming distance weights all bits in an iris code equally. However, not all the bits in an iris code are equally useful. For a given iris image, a bit in its corresponding iris code is defined as "fragile" if there is any substantial probability of it ending up a 0 for some images of the iris and a 1 for other images of the same iris. According to [15] the percent of fragile bits in each row of the iris code. Rows in the middle of the iris code (rows 5 through 12) are the most consistent (See Fig. 5.)
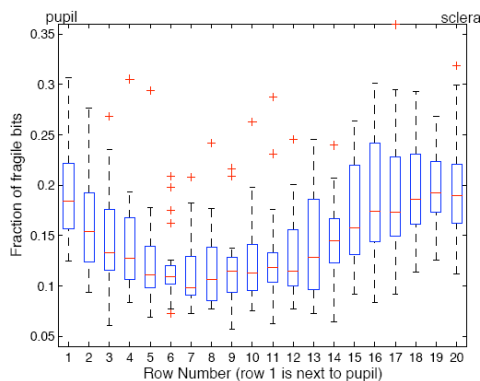
Fig. 5. Percent of Fragile Bit in Iris Pattern

2.3.2   Feature Vector in Proposed Method

As stated in the previous section level 2 sub bands are extracted and according to the Following Rule are modified into binary mode:

If Coeff (i)>=0 then NewCoeff (i) =1
Else NewCoeff (i) =0

According to the method mentioned in section 2.3.1.3, we concluded the middle band of iris normalized images have more important information and less affected by fragile bits, so for introducing iris feature vector based on contourlet transform the rows between 5 and 12 in iris normalize image are decomposed into eight directional sub-band outputs using the DFB at three different scales and extract their coefficients. in our method we use non linear approximation coefficients for select the significant coefficient .for this purpose use the following formula :

$$Nsignif = Round (Npixel*2.5/100) \tag{1}$$

Where npixel is the number of pixel in iris normalized image and nsignif is the number of significant coefficient. In other words it is proved that [16] only bye having 2.5% of coefficients can reconstruct the image. The feature vector in our method has only 48 elements.

2.4   Classifier

We determine whether two irises belong to same class by viewing the similarity of their feature vectors. Rotational alignment is achieved by generating cyclical input feature vectors and matching input feature vectors with template feature vectors. Since the proposed feature extraction process is performed on a block-by-block basis the proposed method is robust to small angular deviations, even without rotation compensation. Two types of measures such as weighted hamming distance and Euclidian distance are used for classification. Comparing the feature vectors X_j and Y_j, the weighted Hamming distance *WHD* is defined as

$$WHD = \frac{1}{N} \sum_{j=1}^{N} X_J (XOR) Y_j \tag{2}$$

Where X_j jth component of the sample feature vector is Y_j is jth component of template feature vector and N is the dimension of input feature vector. If the result of the XOR is zero means the jth component of sample feature vector and template feature vector are the same. Euclidean distance between the template and test image is measured with the following equation.

$$D = \sqrt{\sum_{i=1}^{N} (X_j - Y_j)^2} \tag{3}$$

N is the dimension of the feature vector, X_j the component of the sample feature vector and Y_j is the jth component of the template feature vector.

3   Experimental Results

To evaluate the performance of this proposed system we use "CASIA"[17] iris image database (version 1) created by National Laboratory of pattern recognition, Institute of Automation, Chinese Academy of Science that consists of 108 subjects with 7 sample each. Images of "CASIA" iris image database are mainly from Asians. For each iris class, images are captured in two different sessions. The interval between two sessions is one month. There is no overlap between the training and test samples.
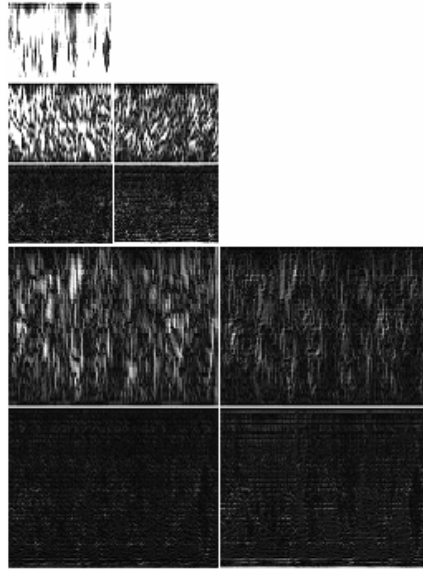
Fig. 6. Contourlet Coefficient

Table 1. Comparison of recognition rate (%) for wavelet and contourlet

(HD=Hamming Distance, ED= Euclidian Distance)

No of Subject	Harr Wavelet		Pfilter:Harr Dfilter:pkva		Pfilter:Harr Dfilter:9/7		Pfilter:Harr Dfilter:5/3	
	HD	ED	HD	ED	HD	ED	HD	ED
20	96	94	97.2	95	98	97	96.5	93
40	94	92.1	95	92	97.8	95.3	93.5	91
60	93	90	92	90	95.8	94.1	92.5	90
80	91	89	90	89	94.6	89	91.7	87
100	89	87	88	86	89	87	86.4	84.3
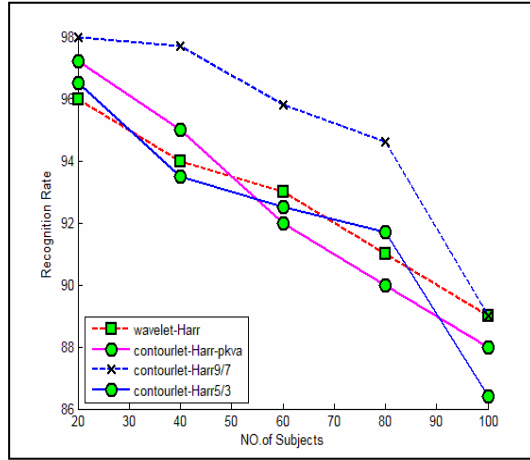
Fig. 7. Comparisons of recognition rate for Wavelet and Contourlet
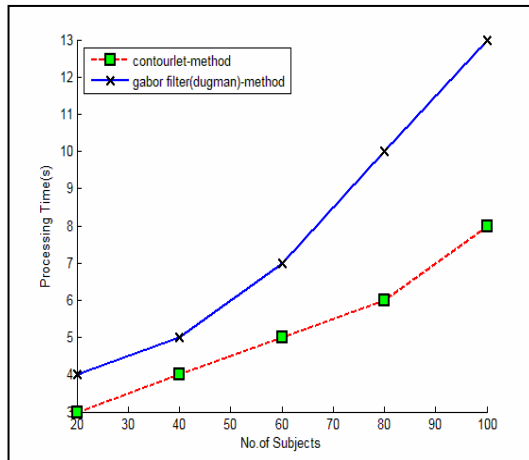
Fig. 8. Comparison of processing Time with wavelet based method (Daugman method)

In our experiments, three-level contourlet decomposition is adopted. The above experiments are performed in Matlab 7.0. The normalized iris image obtained from the localized iris image is segmented by Daugman method. The contourlet transform of the image is shown in Fig. 6. We have used the filters designed by A. Cohen, I. Daubechies, and J.-C. Feauveau. For the quincunx filter banks in the DFB stage.

Table 1 gives the result of the proposed scheme against wavelet based iris image. We compare the proposed system with three filter combinations of contourlet transform and the corresponding plot is shown in Fig.7. Fig.8 shows Comparison of processing Time.

4   Conclusion

A new algorithm for iris recognition using contourlet transform and non linear approximation coefficients has been presented. For Segmentation and normalization we use Dugman methods. Then, each iris image is decomposed with pyramidal directional filter bank and then a fixed length feature vector is obtained. For feature extraction according to [15] only the middle band of normalized iris image decomposed by contourlet transform An extensive result has been taken with different filters. Here, only three filter combinations are considered. Compared with wavelet transform method, contourlet-based method achieves a higher accuracy because contourlet transform has capacity to capture comparatively richer directional information. Experimental results reveal that our algorithm reduces the processing time.

References

1. Daugman, J.: High confidence visual recognition of persons by a test of statistical independence. IEEE Transactions on Pattern Analysis and Machine Intelligence 15(11), 1148–1160 (1993)
2. Daugman, J.: How Iris Recognition works. IEEE Transactions on Circuits and systems for video Technology 14(1), 21–30 (2004)
3. Wildes, R.P., et al.: A machine-vision system for iris recognition. In: Machine Vision and Applications. Springer, Heidelberg (1996)
4. Sun, Z.N., Wang, Y.H., Tan, T., Cui, J.: Improving Iris Recognition Accuracy Via Cascaded Classifier. IEEE Transactions on systems, Man and Cybernetics-part C: Applications reviews 35(3) (2008)
5. Boles, W.W., Boashash, B.: A Human Identification Technique Using Images of the Iris and Wavelet Transform. IEEE Trans. Signal Processing 46(4), 1185–1188 (1998)
6. Sanchez-Avila, C., et al.: Iris-based Biometric Recognition Using Dyadic Wavelet Transform. IEEE Aerospace and Electronics Systems Magazine 17(10), 3–6 (2002)
7. De Martin-Roche, D., Sanchez-Avila, C., Sanchez-Reillo, R.: Iris recognition for biometric identification using dyadic wavelet transform zero crossing. In: Security Technology, 2001 IEEE Int. Carnahan Conference, pp. 272–277 (2001)
8. Ma, L., Tan, T., Wang, Y., Zhang, D.: Personal identification based on iris texture analysis. IEEE Trans. On Pattern Analysis and Machine Intelligence 25(12), 1519–1533 (2003)
9. Zhu, Y., Tan, T., Wang, Y.: Biometric Personal Identification Based on Iris Patterns. Pattern Recognition 2, 801–805 (2000)
10. Ma, L., Wang, Y., Tan, T.: Iris Recognition Based on Multichannel Gabor Filtering. In: Proc. Fifth Asian Conf. Computer Vision, vol. 1, pp. 279–283 (2002)
11. Ali, J.M.H., Hussanien, A.E.: An Iris Recognition System to Enhance E-security Environment Based on Wavelet Theory. In: AMO - Advanced Modeling and Optimization, vol. 5(2) (2003)
12. Do, M.N., Vetterli, M.: The contourlet transform: An efficient directional multiresolution image representation. In: IEEE Trans. Image Proc. (2005)

13. Burt, P.J., Adelson, E.H.: The Laplacian pyramid as a compact image code. IEEE Trans. Commun. 31(4), 532–540 (1983)
14. Bamberger, R.H., Smith, M.J.T.: A filter bank for the directional decomposition of images: Theory and design. IEEE Trans. Signal Proc. 40(4), 882–893 (1992)
15. Hollingsworth, K.P., et al.: The Best Bits in an Iris Code. IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI) 31(6), 964–973 (2008)
16. Contourlet Toolbox (version 2.0) (November 2003), `http://www.ifp.uiuc.edu/~minhdo/software/`
17. Chinese Academy of Sciences – Institute of Automation. Database of 756 Grayscale Eye Images. Version 1.0 (2003), `http://www.sinobiometrics.com`

An Effective Edge-Adaptive Color Demosaicking Algorithm for Single Sensor Digital Camera Images

Md. Foisal Hossain[1], Mohammad Reza Alsharif[1], and Katsumi Yamashita[2]

[1] Department of Information Engineering, Univeristy of the Ryukyus, Okinawa, Japan
foisalkuet@yahoo.com, asharif@ie.u.ryukyu.ac.jp
[2] Graduate School of Engineering, Osaka Prefecture University, Osaka, Japan
yamashita@eis.osakafu-u.ac.jp

Abstract. Most digital cameras use a color filter array of mosaic pattern to capture the colors of the scene. To render a full-resolution color image using a single-chip camera, the missing information must be estimated from the surrounding pixels. In this paper, we present an edge-adaptive demosaicking method for color demosaicking. The algorithm first estimate missing green samples in red and blue position by determining direction of edge in horizontal, vertical and diagonal directions. After determining the green samples, missing red and blue samples estimated using color differences along the direction of edges. Experimental results demonstrate that the proposed method produces visually pleasing images and significantly outperforms existing demosaicking methods in terms peak signal-to-noise ratio (PSNR) and ΔE_{ab}, which is a measure for the average color distance between original and demosaicked images in the CIELAB color space.

Keywords: Color demosaicking, Bayer color filter array, Edge-adaptive demosaicking, Diagonal edge direction, PSNR.

1 Introduction

Now days, most digital cameras capture images using a single sensor (CCD or CMOS) array to reduce cost and size. At each pixel, the sensor detects a particular color channel from the three primary colors (red, green and blue). This kind of sensor is called Color Filter Array (CFA). The most popular CFA pattern is Bayer pattern [1], as shown in Fig. 1. It samples the green band using a quincunx grid, while red and blue are obtained by a rectangular grid. In this way, the density of the green samples is twice than that of the red and blue channels. In order to reconstruct a full color image, the missing color samples need to be interpolated by a process called demosaicking. The quality of the reconstructed color images depends on the image contents and the employed demosaicking algorithms [2].

The early demosaicking methods include nearest neighbor replication [3], bilinear interpolation [4], and cubic B-spline interpolation [5]. These methods can be simply implemented. However, these methods suffer from many drawbacks such as blocking, blurring, false colors, and zipper effect. Specially, many artifacts are produced in the edge region. These effects are caused due to improper fusion of the surrounding color values.
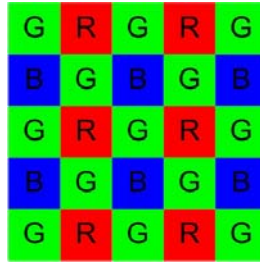
Fig. 1. The Bayer CFA mosaic pattern

In order to improve performance, demosaicking methods combining information from each color plane have been proposed [6], [7], [8], [9]. These methods take advantage of high correlations between the red, green and blue planes in the local regions of natural scene images. However, they still tend to generate zipper effects along edges.

Edge adaptive demosaicking techniques generally use horizontal and vertical edge parameters. The edge parameters of edge adaptive methods represent edge directions at missing pixels. In these methods, interpolation is carried out along edges rather than across them. An edge-adaptive demosaicking method was proposed by Hamilton et al. [10]. This method uses edge indicators to find an edge direction and interpolates missing color pixels along the detected edge direction. Various demosaicking methods using gradients [11], [12], directional information [13] and weighted edge interpolation [14], [15], [16] also use edge information to reduce artifacts along edges. As these methods use only two directions of edges, sometimes they detect wrong edges and results in artifacts along edges.

In this paper, we introduce an innovative solution for demosaicking of single sensor camera images. In the proposed method, we determine direction of edges in four directions instead of two directions. In addition of horizontal and vertical direction, we determine edges in two diagonal directions also. This gives the exact direction of edge of each sample. Then the missing green pixels are estimated by using edge-adaptive interpolation method and the red and the blue channels are interpolated using the green information.

The rest of the paper is organized as follows. In section 2, conventional method of edge adaptive demosaicking has been presented. Section 3 describes the proposed method in detail. Experimental results and performance of the proposed approach are discussed in section 4 and compared to those of the other existing methods. Finally, in section 5, we report the conclusions of the presented work.

2   Conventional Method

In this section, we simply introduce the conventional edge adaptive demosaicking methods. Edge adaptive interpolation method uses the edge direction information to estimate missing green channel pixels. To determine direction of interpolation, the method compares horizontal and vertical gradients. The direction that shows lower gradient is chosen as direction of interpolation. To determine edge direction of G_{13} in Fig. 2, horizontal and vertical gradients are defined as follows:

R1	G2	R3	G4	R5
G6	B7 (G7)	G8	B9 (G9)	G10
R11	G12	R13	G14	R15
G16	B17	G18	B19	G20
R21	G22	R23	G24	R25

Fig. 2. Reference samples in CFA. The green and blue pixels at R_{13} are to be estimated. The pixels in parentheses are estimated pixels.

$$\Delta H = \left| G_{12} - G_{14} \right|$$
$$\Delta V = \left| G_8 - G_{18} \right| . \tag{1}$$

The missing green pixels can be estimated by comparing the two gradients as follows:

$$G_{13} = \begin{cases} \dfrac{G_{12} + G_{14}}{2} & if \quad \Delta H < \Delta V \\[2mm] \dfrac{G_8 + G_{18}}{2} & if \quad \Delta H > \Delta V . \\[2mm] \dfrac{G_8 + G_{12} + G_{14} + G_{18}}{4} & if \quad \Delta H = \Delta V \end{cases} \tag{2}$$

After the green channel is interpolated, the other channels can be estimated by interpolating the color difference plane.

3   The Proposed Method of Demosaicking

In this section, we present the proposed method using information of edge direction at the missing pixel location. The proposed method has two steps. In the first step, we determine the direction of edge of each missing pixels from the four directions. Fig. 3
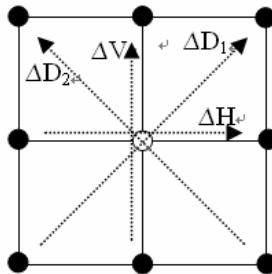
Fig. 3. Four possible directions of edge

shows the possible four directions of edges. We interpolated the green plane along the determined edge direction. In the second step, we interpolate the missing red and blue channels.

3.1   Green Channel Interpolation

In this step, we interpolate the missing green channel using edge information. We concentrate on reconstructing missing green values first, since the performance of reconstructing red and blue values depends thoroughly on how well green values are reconstructed. There are some other reasons also for this. First, the green channel has twice as many samples as the other two channels in the Bayer mosaic pattern. Second, the sensitivity of the human visual system peaks at the green wavelength. Third, the green is closer to red and to blue than the difference between red and blue in wavelength. For these reasons, green channel plays a key role in the estimation of missing color samples.

In order to determine the direction of edge, we calculate four edge indicators, i.e.: horizontal, vertical, diagonal (45^0), and diagonal (-45^0) respectively as shown in Fig. 3. The edge indicators can correctly detect the edge direction as we are using four directions instead of conventionally used two directions. The direction that gives the minimum value of edge indicator is chosen as the direction of edge. Then the missing green sample is interpolated along that direction.

To determine the edge direction of G_{13} in Fig. 2, the horizontal, vertical and diagonal edge indicators are defined as follows:

$$
\begin{aligned}
\Delta H &= |R_{11} - R_{13}| + |R_{13} - R_{15}| + |G_{12} - G_{14}| \\
\Delta V &= |R_3 - R_{13}| + |R_{13} - R_{23}| + |G_8 - G_{18}| \\
\Delta D_1 &= |R_5 - R_{13}| + |R_{13} - R_{21}| + |G_9 - G_{17}| \\
\Delta D_2 &= |R_1 - R_{13}| + |R_{13} - R_{25}| + |G_7 - G_{19}|
\end{aligned}
\tag{3}
$$

where ΔH represents a horizontal difference, ΔH represents a vertical difference, ΔD_1 represents a diagonal (45^0) difference and ΔD_2 represents a diagonal (-45^0) difference respectively, and G_7, G_9, G_{17} and G_{19} are initially calculated using bilinear interpolation.

We interpolate a missing pixel of green channel G_{13} along the edge direction based on the values of the indicators using the following equation:

$$
G_{13} = \begin{cases}
\dfrac{(G_{12} + G_{14})}{2} + \dfrac{(2 \times R_{13} - R_{11} - R_{15})}{4} & \text{if } \quad \Delta H = \delta \\[2mm]
\dfrac{(G_8 + G_{18})}{2} + \dfrac{(2 \times R_{13} - R_3 - R_{23})}{4} & \text{if } \quad \Delta V = \delta \\[2mm]
\dfrac{(G_{10} + G_{14} + G_{18} + G_{22} + G_4 + G_8 + G_{12} + G_{16})}{8} & \text{if } \quad \Delta D_1 = \delta \\[2mm]
\dfrac{(G_2 + G_6 + G_8 + G_{12} + G_{14} + G_{18} + G_{20} + G_{24})}{8} & \text{if } \quad \Delta D_2 = \delta
\end{cases}
\tag{4}
$$

Where $\delta = \min(\Delta H, \Delta V, \Delta D_1, \Delta D_2)$ and G_{13} is the missing green pixels to be esti-
mated. We estimate all the missing green pixels using the same way.

3.2   Red and Blue Channel Interpolation

There are two interpolation cases in the second step of the proposed method. One is
the interpolation of red and blue at the location green pixel. The other is the interpola-
tion of the opposite missing color red or blue at each blue or red pixel.

   Since the performance of interpolating the missing color components depends on
the result of green reconstruction, it is unlikely that any interpolation methods to
estimate red and blue greatly affect the reconstruction stage.

   The average mean-square reconstruction error (MSE) for the red component is
smaller in the green pixels than in the blue ones [17]. Also, the average MSE for the
blue component is smaller in the green pixels than in the red ones. For this reason, we
apply the bilinear interpolation of the color differences in the green locations since it
gives good performance with a low computational cost.

   We applied a different approach to interpolate the red values in the blue pixels and
the blue values in the red ones. We use all the estimated samples at the green posi-
tions to interpolate red and blue pixels. Now each blue position has four estimated red
neighbors placed in the cardinal directions and four original red neighbors placed in
the diagonal direction. Therefore, we choose to perform an edge-directed interpola-
tion using the estimated red samples in the green location and the original red samples
in diagonal directions.

   It is well known that red and blue are less correlated than red and green. In the
blue location of the Bayer pattern the interpolation using the difference R-B gives an
estimate of the red values more representative than using the difference R-G. For this
reason we interpolate the color difference R-B instead of R-G. In a blue position, the
red sample is estimated as follows:

$$
\hat{R}_{i,j} = \begin{cases}
B_{i,j} + \dfrac{1}{2}\left(\hat{R}_{i-1,j} - \hat{B}_{i-1,j} + \hat{R}_{i+1,j} - \hat{B}_{i+1,j}\right) & \text{if} \quad \Delta H = \delta \\[2ex]
B_{i,j} + \dfrac{1}{2}\left(\hat{R}_{i,j-1} - \hat{B}_{i,j-1} + \hat{R}_{i,j+1} - \hat{B}_{i,j+1}\right) & \text{if} \quad \Delta V = \delta \\[2ex]
B_{i,j} + \dfrac{1}{2}\left(R_{i-1,j+1} - \hat{B}_{i-1,j+1} + R_{i+1,j-1} - \hat{B}_{i+1,j-1}\right) & \text{if} \quad \Delta D_1 = \delta \\[2ex]
B_{i,j} + \dfrac{1}{2}\left(R_{i-1,j-1} - \hat{B}_{i-1,j-1} + R_{i+1,j+1} - \hat{B}_{i+1,j+1}\right) & \text{if} \quad \Delta D_2 = \delta
\end{cases} \tag{5}
$$

where $\delta = \min(\Delta H, \Delta V, \Delta D_1, \Delta D_2)$, $\hat{R}_{i,j}$ is the red sample in the blue position to be
estimated, and R is the original red pixel in the diagonal directions of position (i,j).

   The procedure to interpolate the missing blue values in the red pixels follows the
same way.

4   Experimental Results

In order to evaluate the performance of the proposed method, experiments were conducted using some digital color images from the Kodak photo sampler [18]. We compared the proposed method with four conventional demosaicking methods: bilinear interpolation [3], adaptive color plane interpolation [10], homogeneity directed method [13], and edge directed interpolation [16] method.

We evaluate the performance of our proposed method by using two measures. The first measure is the PSNR, and the other one is the CIELAB $\Delta E_{a,b}^{*}$. We also compared average of the above measures over 20 images. We calculate PSNR using the following equation:

$$PSNR = 10\log_{10}\left(255^{2}/MSE\right),\qquad(6)$$

where MSE is the mean square error between the original image and reconstructed image. The $\Delta E_{a,b}^{*}$, which measures the Euclidian distance between the original and reconstructed images in the CIELAB color space, can better capture the color difference perceived by human observers. It is given by

$$\Delta E_{a,b}^{*} = \frac{1}{N}\sum_{1\leq n\leq N}\left\|O(n)_{Lab} - R(n)_{Lab}\right\|,\qquad(7)$$

where $O(n)_{Lab}$ and $R(n)_{Lab}$ are the CIELAB color values of the n'th pixel in the original and reconstructed images, respectively.

Fig. 4 and Fig. 5 show the demosaicked results over image hats and window images respectively for the visual evaluation of the performance of the proposed method as compares to other methods. Observing the figures, it can be noticed that the proposed method produce highest visual quality image and does not produce zipper effect near the edges.
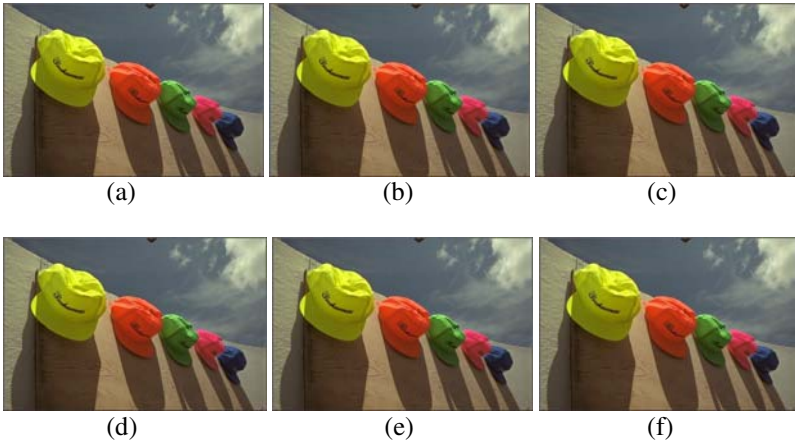
Fig. 4. Demosaicked results over image hats (a) Original; (b) Bilinear; (c) Method in [10]; (d) Method in [13]; (e) Method in [16]; (f) The proposed method
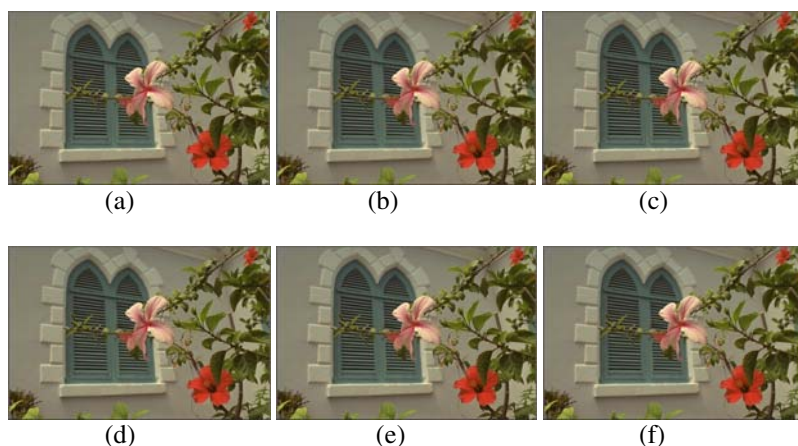
Fig. 5. Demosaicked results over image window (a) Original; (b) Bilinear; (c) Method in [10]; (d) Method in [13]; (e) Method in [16]; (f) The proposed method.

Fig. 6 and Fig. 7 visually compare the performance of the proposed method with other method using cropped regions of images statue and lighthouse respectively. It can be seen that the proposed method gives better performance than other methods for edge regions.

Table 1 shows the PSNR comparison of some output images using five methods. We have also presented the average PSNR of 20 images in this table. From Table 1, it can be seen that the proposed method outperforms the other demosaicking methods in respect to PSNR.
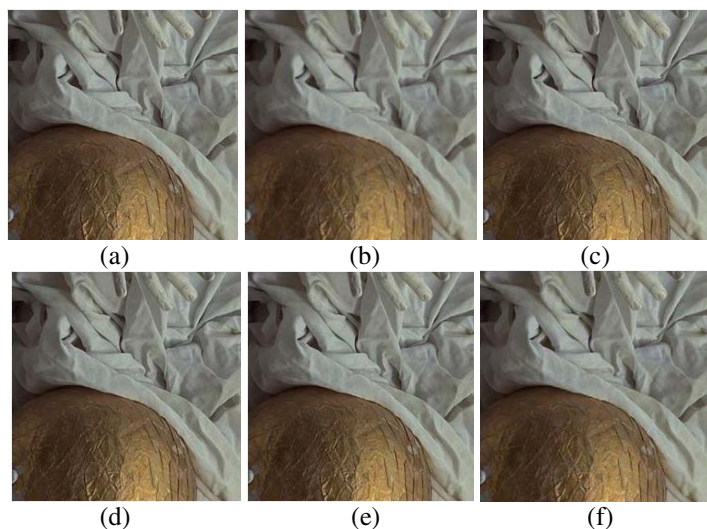
Fig. 6. Demosaicked results over cropped region of image statue. (a) Original; (b) Bilinear; (c) Method in [10]; (d) Method in [13]; (e) Method in [16]; (f) The proposed method.
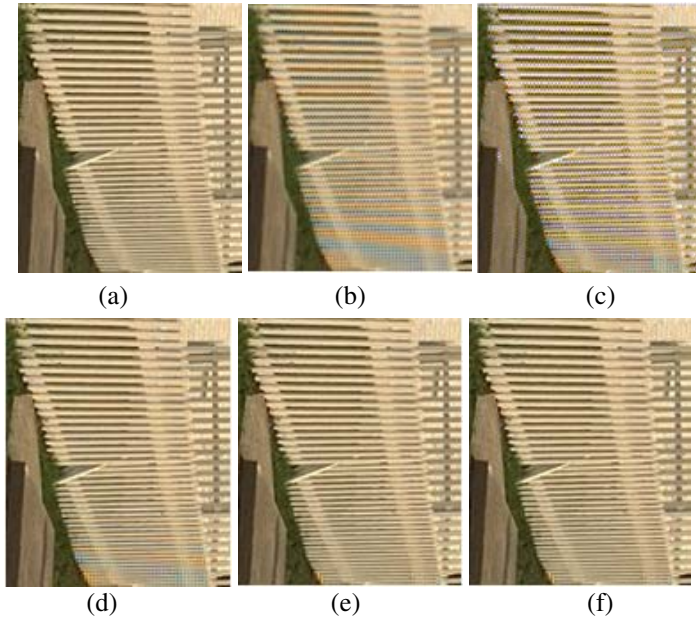
Fig. 7. Demosaicked results over cropped region of image lighthouse (a) Original; (b) Bilinear; (c) Method in [10]; (d) Method in [13]; (e) Method in [16]; (f) The proposed method

Table 1. PSNR(dB) comparison of different methods over different images

Image/method	Bilinear	Method in [10]	Method in [13]	Method in [16]	Proposed method
lighthouse	28.22	37.36	36.12	34.07	38.09
sail	32.52	37.51	37.47	36.18	42.61
statue	32.16	35.54	38.44	34.987	41.72
window	33.50	37.78	40.74	35.459	42.15
hats	34.50	37.703	35.13	36.33	41.58
Average of 20 images	28.91	34.58	37.13	37.24	38.49

Table 2. Performance comparison-CIELAB $\Delta E_{a,b}^*$ over different images

Image/method	Bilinear	Method in [10]	Method in [13]	Method in [16]	Proposed method
lighthouse	3.58	1.683	1.82	2.096	1.57
sail	2.024	0.946	0.96	1.159	0.825
statue	2.92	1.674	1.16	2.049	1.03
window	2.37	1.186	0.93	1.393	0.84
hats	1.99	1.048	1.72	1.255	0.983
Average of 20 images	3.609	1.78	2.21	2.243	1.52

The measure of color reproduction errors of digital images by taking human visual systems, $\Delta E_{a,b}^*$ is given in Table 2 over five images. The average color reproduction error of 20 images has also been compared in this table. Investigating Table 2, it can easily be seen that the proposed method outperforms other methods in respect to $\Delta E_{a,b}^*$ also.

5   Conclusion

In this paper, we presented a new approach of CFA demosaicking based on adaptive edge direction. The direction of edge information is perfectly determined by considering horizontal, vertical, and also diagonal direction. The proposed method generates images with highest visual quality. Experimental result ascertains that the proposed method gives better result than other conventional methods. The proposed demosaicking method outperforms other conventional methods with respect to $\Delta E_{a,b}^*$, and PSNR. The proposed method also significantly reduces unwanted artifacts such as aliasing and zippering near edges.

References

1. Bayer, B.E.: Eastman Kodak Company: Color Imaging Arrays. US patent 3 971 065 (1975)
2. Longere, P., Zhang, X.M., Delahunt, P.B., Brainard, D.H.: Perceptual Assessment of Demosaicking Algorithm Performance. Proc. of IEEE 90, 123–132 (2002)
3. Adams, J.E.: Intersections Between Color Plane Interpolation and Other Image Processing Functions in Electronic Photography. In: Proceedings of SPIE, vol. 2416, pp. 144–151 (1995)
4. Adams, J., Parulski, K., Spaulding, K.: Color Processing in Digital Cameras. IEEE Micro 18(6), 20–30 (1998)
5. Hou, H., Andrews, H.: Cubic Splines for Image Interpolation and Digital Filtering. IEEE Trans. Acoustis, Speech and Signal Processing 26, 508–517 (1987)
6. Pei, S.C., Tam, K.I.: Effective Color Interpolation in CCD Color Filter Array Using Signal Correlation. IEEE Trans. Circuits Syst. Video Tech. 13(6), 503–513 (2003)
7. Moriya, S., Makita, J., Kuno, T., Matoba, N., Sugiura, H.: Advanced Demosaicking Methods Based on the Changes of Colors in a Local Region. IEEE Trans. Consumer Electron 52(1), 206–214 (2006)
8. Lukac, R., Martin, K., Plataniotis, K.N.: Demosaicked Image Postprocessing Using Local Color Ratios. IEEE Trans. Circuits Syst. Video Tech. 14(6), 914–920 (2004)
9. Li, X.: Demosaicking by Successive Approximation. IEEE Trans. Image Proc. 14(3), 370–379 (2005)
10. Hamilton Jr., J.F.: Adaptive Color Plane Interpolation in single Sensor Solor Electronic Camera. U. S. Patent 5, 629, 734 (1997)
11. Kimmel, R.: Demosaicking: Image Reconstruction from Color CCD Samples. IEEE Trans. Image Proc. 8, 1221–1228 (1999)
12. Chang, E., Cheung, S., Pan, D.: Color Filter Array Recovery Using a Threshold-based Variable Number of Gradients. In: Proceedings of SPIE, vol. 14, pp. 36–43 (1999)

13. Hirakawa, K., Parks, T.W.: Adaptive Homogeneity Directed Demosaicking Algorithm. IEEE Trans. Image Proc. 14(3), 360–369 (2005)
14. Chang, L., Tan, Y.P.: Effective Use of Spatial and Spectral Correlations for Color Filter Array Demosaicking. IEEE Trans. Consumer Electron. 50, 355–365 (2004)
15. Zhang, L., Wu, X.: Color Demosaicking via Directional Lineal Minimum Mean Square Error Estimation. IEEE Trans. Image Proc. 14(12), 2167–2178 (2005)
16. Laroche, C.A., Prescott, M.A.: Apparatus and Method for Adaptively Interpolating a Full Color Image Utilizing Chrominance Gradients. U. S. Patent 5 373 322 (1994)
17. Menon, D., Andriani, S., Calvagno, G.: Demosaicing with Directional Filtering and a Posteriori Decision. IEEE Trans. Image Proc. 16(1) (2007)
18. Images from KODAK Photo CD Photo Sampler (1991),
    `http://www.site.uottawa.ca/~eduboid/demosaicking/`

Ship Classification by Superstructure Moment Invariants

Prashan Premaratne and Farzad Safaei

School of Electrical, Computer and Telecommunications Engineering,
University of Wollongong, North Wollongong, NSW, Australia
prashan@uow.edu.au

Abstract. Direct observation using satellites and long range video surveillance is not possible for ship classification in adverse weather and during night. Radar and more specifically radar imaging offers a solution for the above adverse conditions. Ship Classification using radar is of utmost important to defense of any country to manage vast naval resources and to tell the friend from foe. Automatic ship classification based on radar images has been very successful in determining the ship class as well as other details to reliably recognize a ship type using machine vision. Inverse Synthetic Aperture Radar (ISAR) Imaging which relies on a stationary radar and a moving object with preferably super-structure will result in an image that is somewhat unique to a particular ship class. There have been many attempts to classify these ISAR images automatically with varying degree of success. The results we present here using Moment Invariants (Hu Moments) are indeed superior to many other feature-based classification approaches as they have strong invariant properties.

Keywords: Ship Classification, Superstructures, Hu Moments; Radar Imaging.

1   Introduction

Inverse SAR imaging refers to the use of target motion alone to generate a synthetic aperture for azimuth resolution [1]. In practice, ISAR often includes antenna platform motion. Images of the target region produced by ISAR can be a useful tool in locating scattering regions on the target. ISAR images are produced by rotating the target and processing the resultant Doppler histories of the scattering centers. If the target rotates in azimuth at a constant rate through a small angle, scatters will approach or recede from the radar at a rate depending only on the cross range position- the distance normal to the radar line of sight with the origin at the target axis of rotation. The rotation will result in the generation of cross range dependent Doppler frequencies which can be sorted by a Fourier transform [1]. This operation is equivalent to the generation of a large synthetic aperture phased array antenna formed by the coherent summation of the receiver outputs for varying target/antenna geometries. For small angles, an ISAR image is the two-dimensional Fourier transform of the received signal as a function of frequency and target aspect angle.

ISAR is utilized in maritime surveillance for the classification of ships and other objects. In these applications the motion of the object due to wave action often plays a

greater role than object rotation. For instance a feature which extends far over the surface of a ship such as a mast will provide a high sinusoidal response which is clearly identifiable in a two dimensional image. Images sometimes produce an uncanny similarity to a visual profile with the interesting effect that as the object rocks towards or away from the receiver the alternating doppler returns cause the profile to cycle between upright and inverted. ISAR for maritime surveillance was pioneered by Texas Instruments in collaboration with the Naval Research Laboratory and became an important capability of the P-3 Orion and the S-3B Viking US Navy aircraft. Research has also been done with land based ISAR. The difficulty in utilizing this capability is that the object motion is far less in magnitude and usually less periodic than in the maritime case.

2   Ship Classification

ISAR images are obtained from the correlation of radar returns and the subsequent inverse Fourier Transformation. Any superstructure that undergoes rolling, pitch or yaw will result in strong radar returns that will be useful for identifying these superstructures in the ISAR image (see Fig. 4). There have been many approaches to identify the physical features such as length, superstructure spacing and features from the ISAR images such as geometric moments to classify ships effectively. This paper is about our research into Moments Invariants derived from ISAR images to classify ships.

Moment Invariants are a feature type that has been successfully used in various image classification problems [2-6]. Moment Invariants are a reliable and versatile way to construct a feature vector of low dimension as the basis for many classifiers. Moments have been used as pattern features in number of applications [7-11]. Furthermore, moment invariants are properties of connected regions in binary/boundary/intensity images that are invariant to

- Translation
- Rotation
- Scale

They are useful because they define a simply calculated set of region properties that can be used for shape classification and recognition.

As described in Zhongliang *et al*. [8], moment invariants of superstructure appear promising for automatic ship classification. Superstructure invariants are different from the general moment invariants for the entire ship. Zhongliang *et al*. reported that
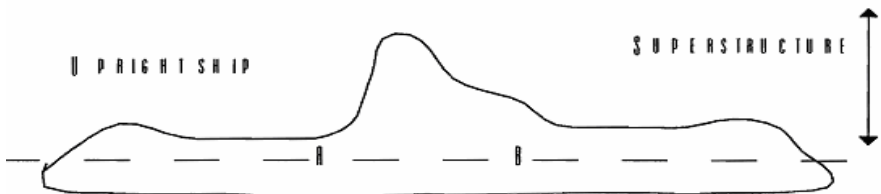
Fig. 1. Diagram depicting an ISAR image of a ship and the A-B line denotes the superstructure above the hull

analysis of theory and statistics of experimental results had shown that the classification accuracy using superstructure moment invariants for ship is higher than that of general ones.

2.1 Superstructure Extraction

The key to represent and recognize ships effectively lies in separating the superstructure from the ship images suitably and consistently. The following steps highlight superstructure separation from the ship silhouette by using geometry moments [8].

1. Calculate the centroid of the ship silhouette. The centroid is C(Ix, Jy), where $Ix = m_{10}/m_{00}$, $Jy = m_{01}/m_{00}$, m_{00}, m_{01} and m_{10} are geometry moments of the ship.
2. Separate the ship image into two parts by the vertical line, which goes across the centroid of the ship.
3. Calculate the centroids of the two parts of the ship. The centroid of the left section is CL(ILx, JLy), where $ILx = mL_{10}/mL_{00}$, $JLy = mL_{01}/mL_{00}$, and mL_{10}, mL_{01}, mL_{00} are geometry moments obtained from the left section of the ship, the centroid of the right section is CR(IRx, JRy), where $IRx=mR_{10}/mR_{00}$, and $JRy = mR_{01}/mR_{00}$, and mR_{10}, mR_{01}, mR_{00} are geometry moments of the right section.
4. The section above the line that goes across CL and CR is considered to be the superstructure according to Zhongliang *et al*. However, this is not entirely true for our data set and there are other issues as described in the section below.

2.2 Definition of Hu Moments

Moment invariants algorithm has been known as one of the most effective methods to extract descriptive feature for object recognition applications. The algorithm has been widely applied in classification of aircrafts, ships, ground targets, etc [8]. Essentially, the algorithm derives a number of self-characteristic properties from a binary image of an object. These properties are invariant to rotation, scale and translation. Let $f(i,j)$ be a point of a digital image of size $M{\times}N$ ($i = 1,2, …, M$ and $j = 1,2, …, N$). The two dimensional moments and central moments of order ($p + q$) of $f(i,j)$, are defined as:

$$m_{pq} = \sum_{i=1}^{M} \sum_{j=1}^{N} i^{p} j^{q} f(i, j), \qquad (1)$$

$$U_{pq} = \sum_{i=1}^{M} \sum_{j=1}^{N} (i - \bar{i})^{p} (j - \bar{j})^{q} f(i, j), \qquad (2)$$

Where

$$\bar{i} = \frac{m_{10}}{m_{00}} \qquad \bar{j} = \frac{m_{01}}{m_{00}}$$

From the second order and third order moments, a set of seven moment invariants are derived as follows [9]:

$$\phi_1 = \eta_{20} + \eta_{02}, \tag{3}$$

$$\phi_2 = (\eta_{20} - \eta_{02})^2 + 4\eta_{11}^2, \tag{4}$$

$$\phi_3 = (\eta_{30} - 3\eta_{12})^2 + (3\eta_{21} - \eta_{03})^2, \tag{5}$$

$$\phi_4 = (\eta_{30} + \eta_{12})^2 + (\eta_{21} + \eta_{03})^2, \tag{6}$$

$$\phi_5 = (\eta_{30} - 3\eta_{12})(\eta_{30} + \eta_{12})\left[(\eta_{30} + \eta_{12})^2 - 3(\eta_{21} + \eta_{03})^2\right]$$
$$+ (3\eta_{21} - \eta_{03})(\eta_{21} + \eta_{03})\left[3(\eta_{30} + \eta_{12})^2 - (\eta_{21} + \eta_{03})^2\right] \tag{7}$$

$$\phi_6 = (\eta_{20} - \eta_{02})\left[(\eta_{30} + \eta_{12})^2 - (\eta_{21} + \eta_{03})^2\right] + 4\eta_{11}(\eta_{30} + \eta_{12})(\eta_{21} + \eta_{03}), \tag{8}$$

$$\phi_7 = (3\eta_{21} - \eta_{03})(\eta_{30} + \eta_{12})\left[(\eta_{30} + \eta_{12})^2 - 3(\eta_{21} + \eta_{03})^2\right]$$
$$- (\eta_{30} - 3\eta_{12})(\eta_{21} + \eta_{03})\left[3(\eta_{30} + \eta_{12})^2 - (\eta_{21} + \eta_{03})^2\right] \tag{9}$$

Where η_{pq} is the normalised central moments defined by:

$$\eta_{pq} = U_{pq} \big/ U_{00}^r$$

where $r = [(p+q)/2] + 1$ and $p + q = 2, 3, ...$

2.3 Examples

Figure 2a and 2b show two sizes and orientations of letter 'A'. The natural logarithm of seven moment invariants described in the previous section is shown in the Table 1. These values are extremely close as would be expected for moment invariants. Figures 3a and 3b also show another two sizes and orientations of letter 'L' and their moment invariants are shown in Table 2. The extreme closeness of these vales confirms that moment invariants are invariant to size, translation and rotation of the object. Figures 5, 6a and 6b show true ISAR images of the same ship taken from different angles. Table 3 lists the Moment Invariant values for these images and we can conclude that Moment Invariants offer us a practical way to find the similarity of these ISAR images even though they look different to human subject.
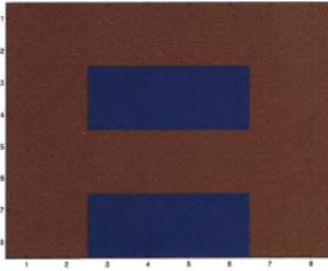
Fig. 2a. Upright large letter 'A'   **Fig. 2b.** Small rotated 'A'
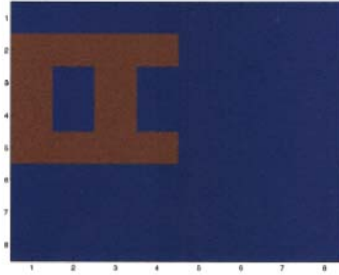
Table 1. Moment Invariants for above images

Moment Invariants	Figure 2a	Figure 2b
ϕ_1	1.40971463532799	1.45330507140285
ϕ_2	7.00695246156515	7.00695246156514
ϕ_3	8.50121623689311	8.50121623689315
ϕ_4	9.88751059801298	9.88751059801296
ϕ_5	19.08187401546602	19.08187401546602
ϕ_6	13.39098682879555	13.39098682879553
ϕ_7	50.03793106935080	49.34478388879084
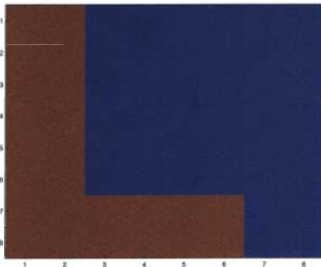
Fig. 3a. Upright large letter 'L'   **Fig. 3b.** Small rotated 'L'
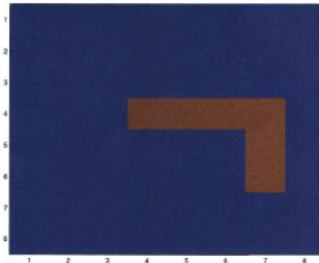
Table 2. Moment Invariants of Fig. 3a and 3b

Moment Invariants	Fig. 3a	Fig. 3b
ϕ_1	1.07799300146537	1.14117190308691
ϕ_2	3.13723183582769	3.13723183582769
ϕ_3	3.76584049525006	3.76584049525006
ϕ_4	5.96306507258628	5.96306507258628
ϕ_5	12.10048353231735	12.10048353231735
ϕ_6	8.80464666631302	8.80464666631302
ϕ_7	10.86833985102471	10.86833985102471
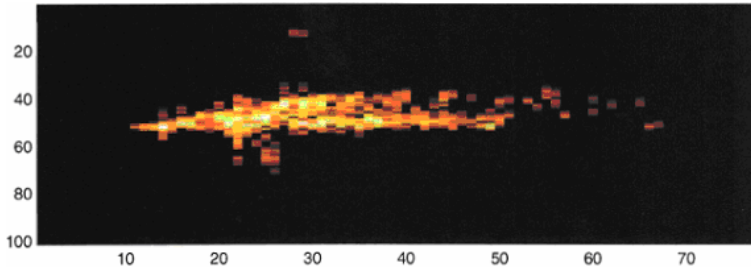
Fig. 4. ISAR ship image
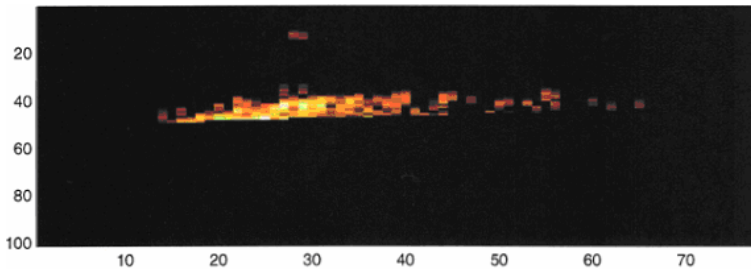
Fig. 5. Superstructure of the image in Fig. 4
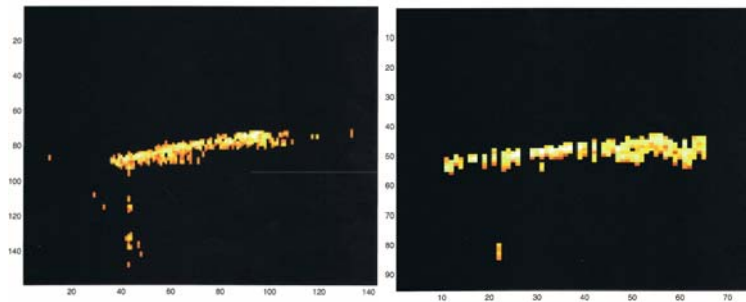
Fig. 6a (left) and **Fig. 6b** (right). Two ISAR images from different angles.

Table 3. Moment Invariants of the same ship with different angles and distances

	Fig. 5	Fig. 6a	Fig. 6b
Φ_1	0.00454532651883	0.00400145239121	0.00415891638507
Φ_2	0.00000935525509	0.00000913787402	0.00001335968243
Φ_3	0.00000007392946	0.00000002985004	0.00000000600204
Φ_4	0.00000005526468	0.00000001919806	0.00000000291252
Φ_5	0.00000000000000	0.00000000000000	0.00000000000000
Φ_6	0.00000000015326	0.00000000005616	0.00000000000969
Φ_7	0.00000000000000	0.00000000000000	0.00000000000000

3   Experimental Results

We used set of 350 ISAR images of 5 ship classes and classified them using a vector quantization classifier. The objective of the experiments was to assess the effectiveness of Moment Invariants as a feature against other physical features such as length estimates, cos aspects, geometric moments, fractal dimensions, etc.. Figure 7 depicts the effectiveness of Moment Invariants as set of features in classifying ships.
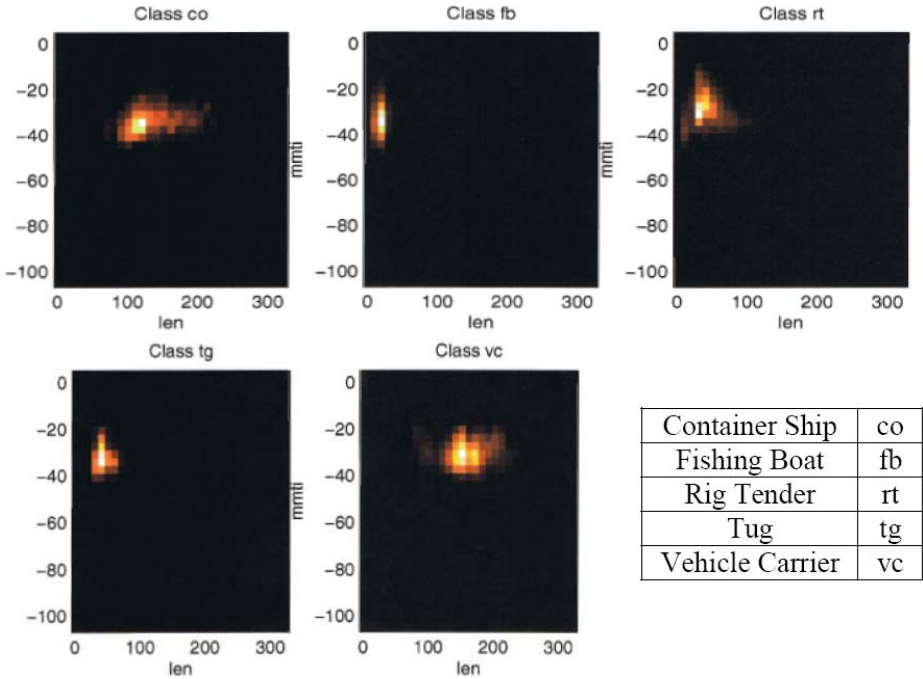
Fig. 7. Class separation with Moment Invariants

Container Ship	co
Fishing Boat	fb
Rig Tender	rt
Tug	tg
Vehicle Carrier	vc

4   Summary

It is possible to observe that Fig. 8 classification result is superior to Fig. 9 which uses only Moment Invariants. Even though, there is significant misclassification of container class as vehicle carrier and Rig Tender as Tugs, the overall performance compared to classification with other physical features can not be disregarded as poor performance. In fact, in practice, it is quite difficult to estimate the true length of ships due to water acting as a scattering agent behind ships. It is also very difficult and time consuming to calculate many physical quantities which usually lead to errors and subsequent misclassifications. Overall, Hu Moments show a great promise in automatic ships classification with reduced complexity in the overall classification system.
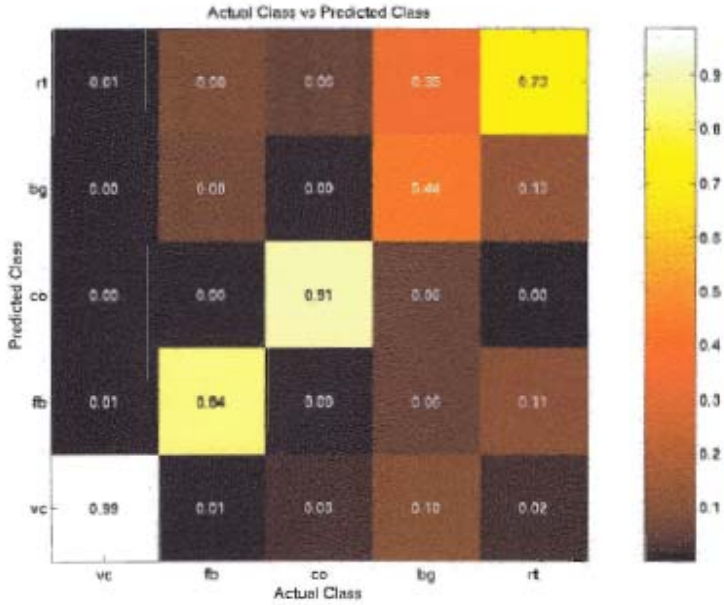
Fig. 8. Confusion Matrix with features such as Lengths, geometric moment ratios, cos(aspect), height profiles and fractal dimensions
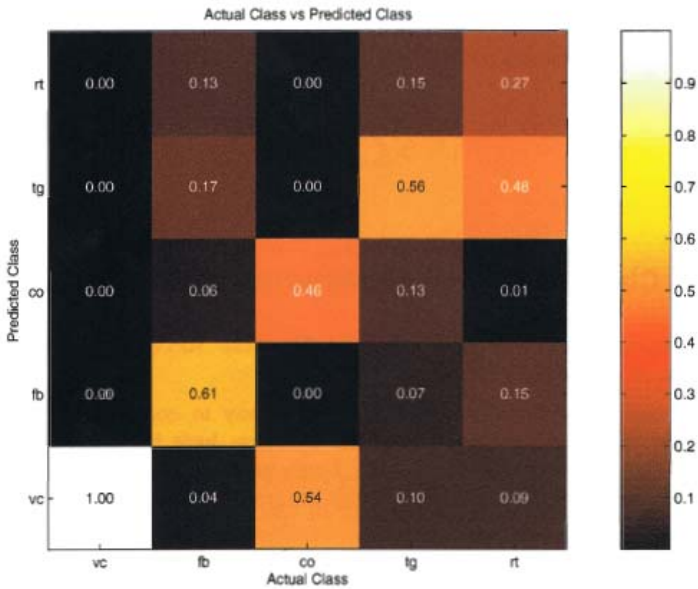
Fig. 9. Confusion Matrix with length and Hu Moments 1,2,6 & 7

References

1. Musman, S., Kerr, D., Bachmann, C.: Automatic recognition of ISAR ship images. IEEE Transactions on Aerospace and Electronic Systems 32(4), 1392–1404 (1996)
2. Dudani, S.A., Breeding, K.J., McGhee, R.B.: Aircraft Identification by Moment Invariants. IEEE Trans. on Computers c-26-1, 39–45 (1997)
3. Maitra, S.: Moment Invariants. Proc. IEEE 67(4), 697–699 (1979)
4. Belksim, S.O., Sridhar, M., Ahamad, M.: Pattern Recognition with Moment Invariants: A Comparative Study and New Results. Pattern Recognition 24(12), 1117–1138 (1991)
5. Pejnovic, P., Markovic, M., Stankovic, S.: Shape Classification by Moments and Autoregressive Invariants. Int. Journal of Pattern Recognition and Artificial Intelligence 15(2), 311–327 (2001)
6. Flussser, J., Suk, T.: Affine Moment Invariants: A New Tool for Character Recognition. Pattern Recognition Letters 15, 433–436 (1994)
7. Schweitzer, H., Stranch, J.: Utilizing Mount Invatiants and Grobner Bases to Reason about Shapes. www
8. Zhongliang, Q., Wenjung, W.: Automatic Ship Classification by Superstructure Moments Invariants and Two-Stage Classifier. In: ICCS/ISITA, pp. 544–547 (1992)
9. Hu, M.K.: Visual Pattern Recognition by Moments Invariants. IRE Trans. on Information Theory 8, 179–187 (1962)
10. Li, Y.: Reforming the Theory of Invariant Moments for Pattern Recognition. Pattern Recognition 25(7), 723–730 (1992)
11. Gouaillier, V., Gagnon, L.: Ship Silhouette Recognition Using Principal Component Analysis. In: SPIE Proc. Conference Application of Digital Image Processing (1997)

Segmentation of Blood and Bone Marrow Cell Images via Learning by Sampling

Chen Pan[1], Huijuan Lu[1], and Feilong Cao[2]

[1] College of Information Engineering, China Jiliang University, Hangzhou, Zhejiang, 310018, P.R. China
`pc916@cjlu.edu.cn, hjlu@cjlu.edu.cn`
[2] College of Science, China Jiliang University, Hangzhou, Zhejiang, 310018, P.R. China
`flcao@cjlu.edu.cn`

Abstract. This paper presents an automatic machine-learning method to segment blood and bone marrow cell images. Different from traditional methods, we focus on a few significant samples rather than all of them. Firstly, three mean-shift procedures are used to seek the local clustering modes corresponding to the regions of nuclei, mature erythrocytes and background respectively. And then a SVM is trained by uniform sampling from three modes in order to find more nuclei pixels. So we could dilate the nuclei regions only in high gradient pixels to get the part pixels of cytoplasm. Finally, we train a new SVM by a training set sampling from cytoplasm and three modes to extract the whole leukocytes. SVM with fixed parameters is used here to yield two classification models via learning by sampling on-line. The segmentation results of the new method are closer to the human visual perception. It can achieve higher accuracy of segmentation in complex scenes and more robust to color confusion and changes. Experiments have demonstrated the validity of the new method compared with the thresholding and the watershed algorithm.

Keywords: Color image segmentation, support vector machine, mean-shift, blood and bone marrow cell.

1 Introduction

For the diagnosis of leukemia, the microscopic investigation of the blood and bone marrow cells is an essential method. In order to improving the reliability of analysis and diagnosis, computer-based automatic "brain-eye system" would have important significance. However, image segmentation is one of the most basic and key step in the automatic system. Its accuracy and stability greatly affect the operating speed and recognition accuracy of the whole system. It's also a difficult and challenging problem owing to the complexity and the uncertainty of the microscopic images[1].

Machine learning algorithms such as artificial neural networks (ANN) and support vector machine (SVM) have been widely and successfully used in medical image analysis, including image segmentation [1,2,3], due to their good performance in classification and regression. However, most of ANN and SVM need training. Training speed is usually slow in large-scale data. So ANN and SVM are seldom used in

on-line training cases. Since their training sets are often fixed, the models are fixed too. Such immobilized solution cannot be very good to deal with uncertainty and change that often faced in the microscopic images.

Recently, we had presented an effective strategy to speed up the SVM training and avoid overfitting[4,5]. The main idea is to prepare pure training samples without over-lapping in probability distribution, which may belong to each clustering modes with high frequency. So hard margin criterion instead of soft margin one in SVM training. We can select training points actively and control the size of the training set less than 1000 samples. So the training speed using fixed parameters is nearly in real-time.

In this paper, we use on-line classifier of SVM continuously. The procedure is de-scribed as follows: (1)Firstly, we use three mean-shift procedures to search the clus-tering modes (local probability density maximum in RGB color space) corresponding to the color of nuclei, mature erythrocytes and background respectively. Then we can prepare the training samples by uniform sampling from three modes, and train a SVM to extract more nuclei pixels. (2)The nuclei regions are morphological dilated condi-tionally according to the gradient of pixels in order to get parts of cytoplasm. Then we train the second SVM with new samples sampling from cytoplasm and the old train-ing samples used in the first SVM. Color and spatial information could be utilized together in this time. (3)Finally, the last SVM is used to segment the image to extract whole white cells. Since the size of training set could be controlled easily in sampling, SVM training is fast.

Due to mean-shift procedure has robust performance to feature space analysis and SVM has good generalization performance and nonlinear modeling capabilities, our method shows three advantages: (1)It is an adaptive segmentation method to complex scenes. (2)It can overcome the effect of color confusion. (3)It more close to human visual perception.

The rest of the paper is organized as follows. Section 2 introduces mean-shift and SVM training briefly. Section 3 presents how to train SVM online. Section 4 shows some experimental results and comparisons with thresholding and watershed algorithm. Discussion and conclusion are given in section 5.

2   Mean-Shift and SVM Training

2.1 Mean-Shift Procedure

Mean-shift is a simple, nonparametric technique for estimation of the density gradi-ent, which was proposed by Fukunaga and Hostetler and recently generalized by Cheng, Comaniciu and Meer[6].

Define a d-variate kernel function $K(\mathbf{y})$ is a search window. If it is symmetric,

$$K(\mathbf{y}) = ck(\|\mathbf{y}\|^2) \tag{1}$$

where c is normalization constant, $k(s)$ is a symmetric univariate kernel which is called the profile of $K(\mathbf{y})$, only for $s \geq 0$; $\mathbf{y}$ is the center location of the search window.

Given n data points $\mathbf{x}_i, i = 1...n$ in the d-dimensional space R^d, we can compute the location of the centroid of a search window, which is:

$$\mathbf{y}_{centroid} = \frac{\sum_{i=1}^{n} \mathbf{x}_i k\left(\left\|\frac{\mathbf{y}-\mathbf{x}_i}{h}\right\|^2\right)}{\sum_{i=1}^{n} k\left(\left\|\frac{\mathbf{y}-\mathbf{x}_i}{h}\right\|^2\right)} \qquad (2)$$

where h is the bandwidth of kernel.

If we move the search window from $\mathbf{y}$ to $\mathbf{y}_{centroid}$, the multivariate shift vector denoted by $m_K(\mathbf{y})$ means that the vector of difference between the local mean and the center of the window, which is:

$$m_K(\mathbf{y}) = \mathbf{y}_{centroid} - \mathbf{y} \qquad (3)$$

If $\left\|m_K(\mathbf{y})\right\|^2 \geq \xi$, ξ is an arbitrary small positive value, we could move the window iteratively. Denote by $\{\mathbf{y}_j\}$, $j = 1,2,...$ the sequence of successive location of the kernel $K(\mathbf{y}_j)$. If $\mathbf{y}_j$ is the present center position, the next center of the search window $\mathbf{y}_{j+1}$ is:

$$\mathbf{y}_{j+1} = \frac{\sum_{i=1}^{n} \mathbf{x}_i k\left(\left\|\frac{\mathbf{y}_j-\mathbf{x}_i}{h}\right\|^2\right)}{\sum_{i=1}^{n} k\left(\left\|\frac{\mathbf{y}_j-\mathbf{x}_i}{h}\right\|^2\right)} \qquad (4)$$

It demonstrated that the mean shift vector $m_K(\mathbf{y}_j)$ is proportional to the gradient of the probability density at $\mathbf{y}_j$. Low-density regions correspond to large mean shift, and the shift is always in the direction of the probability density maximum (the mode or peak). At the mode, the mean shift is close to zero. This property can be exploited in a simple, adaptive steepest ascent algorithm.

In this paper, RGB space is the mean shift workspace. Mean shift algorithm is described as the following:

1. Choose the bandwidth h of the search window. It is a key parameter that controls the resolution of the algorithm in the feature space.
2. Choose the initial location p of the window.
3. Compute the mean shift vector and translate the search window by that amount.
4. Repeat till convergence.

The fixed point found by a mean shift is the local highest density that corresponds to clustering mode in color space.

It is worth to note that p is an initial location, its corresponding mode can tolerate p up to the edge of the attraction field. That is to say the changes in p do not affect the results of clustering as long as it is in the attraction domain of the mode. So the mean-shift procedure used here could overcome color variation such as in the staining process of the slides and in illumination caused by thickness of the smear.

2.2 SVM Training

Let the training set D be $\{(\mathbf{x}_i, l_i)\}_{i=1}^{N}$, with input $\mathbf{x}_i$ and $l_i = \{\pm 1\}$. SVM [7] first maps $\mathbf{x}$ to $\mathbf{z} = \phi(\mathbf{x}) \in F$. When the data is linearly separable in F, SVM constructs a hyperplane $\mathbf{w}^T \mathbf{z} + b$ for which separation between the positive and negative examples

is maximized. It can be shown that $\mathbf{w} = \sum_{i=1}^{N} \alpha_i l_i \mathbf{z}_i$, where $\boldsymbol{\alpha} = (\alpha_1,...,\alpha_N)$ can be found by solving the following quadratic programming problem:

$$\min \frac{1}{2} \boldsymbol{\alpha}^T \mathbf{Q} \boldsymbol{\alpha} - \mathbf{e}^T \boldsymbol{\alpha} \tag{5}$$

subject to $\boldsymbol{\alpha} \geq 0$ and $\boldsymbol{\alpha}^T \mathbf{l} = 0$. Where $\mathbf{e}$ is the vector of all ones, $\mathbf{l} = (l_1,...,l_N)^T$ and $\mathbf{Q}$ has entries $l_i l_j \mathbf{z}_i^T \mathbf{z}_j = l_i l_j K(\mathbf{x}_i, \mathbf{x}_j)$, where $K(\mathbf{x}_i, \mathbf{x}_j)$ is called a kernel. When the training set is not separable in F , the SVM algorithm introduces non-negative slack variable $\xi_i \geq 0$. The resultant problem becomes

$$\min \frac{1}{2} \|\mathbf{w}\|^2 + C \sum_{i}^{n} \xi_i \tag{6}$$

subject to $l_i(\mathbf{w}^T \mathbf{z}_i + b) \geq 1 - \xi_i$. C is a regularization parameter controlling the trade-off between model complexity and training error.

The $\mathbf{x}_i$ for which $\alpha_i \neq 0$ are defined as the support vectors, since they determine the optimal hyperplane, the hyperplane with maximal margin. Geometrically, the support vectors correspond to the closest to the optimal hyperlane. The optimal decision functions is

$$f(\mathbf{x}) = sign(\sum_{i=1}^{N} \alpha_i l_i K(\mathbf{x}_i, \mathbf{x}) + b) \tag{7}$$

In this paper, radial basis function (RBF) is selected as the kernel of SVM, which is

$$K(\mathbf{x}_i, \mathbf{x}_j) = \exp(-q \|\mathbf{x}_i - \mathbf{x}_j\|^2) \tag{8}$$

The rest parameters in training are the kernel parameter q and the regularization parameter C of SVM. Because our training data in this paper has not intersection, it means that the training set is linear-separable in high-dimensional feature space. The regularization parameter C can be set a fixed value. So that only parameter q needs to be optimized in our method, that is often an off-line task in practice via multiple-fold cross-validation. Many literatures could be referred for the adjustment problem of SVM parameters.

We denote q^* as optimal value. It worth to note that $q^* \in [1 \ \ 100]$ in most cases for cell image segmentation. Once the set of parameters are fixed, the performance of SVM could be controlled by the training samples. In experiments [4,5], the size of the training points could be set less than 1000 by uniform sampling to guarantee a satisfactory level for segmentation.

3   Methods

In our research, blood and bone marrow smears are conventionally prepared with Wright-Giemsa stain. From the view of the observer, the bright or white regions correspond to background, the dark regions correspond to nuclei, the regions that

have intermediate brightness may correspond to mature erythrocytes and other regions may be cytoplasm. Cytoplasm regions always surround nucleus[8].

According to the prior knowledge mentioned above, three mean shift procedures can be used to find clustering modes of nuclei, mature erythrocytes and background. However, no all images contain nucleated cells and mature erythrocytes simultaneously in practice. Three common situations exist in a scene: (1)no nucleated cell; (2) only nucleated cell; (3) nucleated cell and red blood cell both in an image.

3.1   Selected the Image Contains the Nucleated Cells Using SVM

In order to guarantee the success of our segmentation method, we had presented a preprocessing procedure to classify the image into three types according to above analysis. We had acquired over thousands of blood and bone marrow cell images, and represented them with their hue histograms. Then we trained a three-classification model via SVM (off-line). The model classifies the microscopic images into three types. The flow chart is shown in Fig.1.
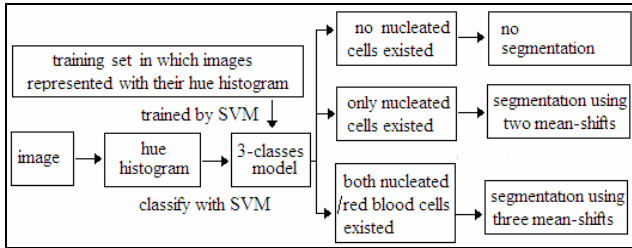
Fig. 1. The cell images could be divided into three types via a preprocessing model

From the Fig.1, we execute segmentation to those images in which the nucleated cells existed. In particular, in the scene of only the nucleated cells existed, the segmentation of the nucleated cell may be regarded as a special case where the mature erythrocytes are not existed, so we could use two mean-shifts to segment the image.

3.2   Preparing Training Samples for SVM to Extract Nuclei

Thus in this paper, we usually focus on the complexity case using three mean-shifts to extract the nucleated cells from the image. The corresponding initial locations of search window are defined as p_n, p_e and p_b, their bandwidths are h_n, h_e and h_b respectively.

Since background pixels have highest color value and nuclei have smallest one in color space, we set $p_b(R,G,B)=(255,255,255)$ and $p_n(R,G,B)=(0,0,0)$. The color of mature erythrocytes is often various in different cases, so p_e could be set by experience or designated manually by mouse clicking.

To guarantee mean shift procedure convergence to the local mode, h_n , h_e and h_b could self increase to a limit before the search window first shift. And then those bandwidth parameters are fixed in the rest shift procedures.
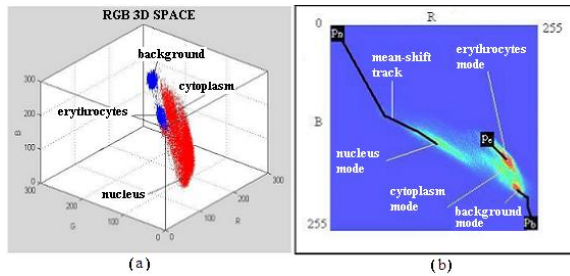
Fig. 2. (a) Color vectors distribution in RGB space. (b) Three clustering modes and corresponding mean shift tracks be illustrated in RB color plane.

However, fixed key parameter could not deal with the various color or uncertainty in the microscopic images. For examples, it is hard to take the whole nuclei regions when the nuclei colors are various, i.e. belong to different local modes. So selecting suitable bandwidth parameters is a critical work in this case. So we have an idea to solve this problem via the generalization performance of SVM. In following, we present an approach to extract the most nuclei regions using small nuclei samples.

When three clustering modes have been found, the corresponding pixels around the modes are mapped back to image domain. Thus, nuclei pixels are marked with I^+, and mature erythrocytes or background pixels are marked with I^-. In order to reduce the size of training set, a sampling approach is described as follows:

① Set the size of the training set is *N*. It's an experiential value by which the performance of SVM should not decrease.

② Uniform sampling in I^+ and I^- regions, *N/2* samples in each class.

Let the training set $(\mathbf{x}_i, l_i)$, $\mathbf{x}_i \in I^+ \cup I^-$, $l_i = \begin{cases} +1, if & \mathbf{x}_i \in I^+ \\ -1, if & \mathbf{x}_i \in I^- \end{cases}$. $\mathbf{x}_i = (R, G, B)^T$,

where (R,G,B) is color value of the pixel, $i \leq N$. In general, *N=1000* is suitable in most cases. So we can train a SVM model on-line to extract the nuclei regions.

3.3 Preparing Training Samples for SVM to Extract Whole Nucleated Cells

It well known that the pixels on the edge not only have higher gradient value, but also have same color features on both sides of the edge. In this section, the nuclei regions extracted by SVM are dilated conditionally if the pixel gradient is higher than a gradient threshold. So the new dilation regions (parts of cytoplasm) are limited in the boundary. By this way, we could take some useful pixels belong to nucleated cells, but also reduce noise component effectively in dilation procedure.

We continually use SVM to play the role of the cytoplasm of small samples. Specific approach is to train second SVM with four types samples, which are sampled from the part of cytoplasm and three old training samples (pure samples) used in the first SVM, respectively. The each size of them is N/4. It means the weights of the four type pixels are equal, even if the sizes of them are different. Color and spatial information could be utilized together in this time.

Thus, cytoplasm and nuclei pixels are marked with I^+, and mature erythrocytes and background pixels are marked with I^-. In order to avoid uncertainty, we set $I^+ \cap I^- = \Phi$ (empty).

Let the training set $(\mathbf{x}_i, l_i)$, $\mathbf{x}_i \in I^+ \cup I^-$, $l_i = \begin{cases} +1, if & \mathbf{x}_i \in I^+ \\ -1, if & \mathbf{x}_i \in I^- \end{cases}$.

$\mathbf{x}_i = (R, G, B, x, y)^T$, where (x, y) is the coordinate of pixel. Since $N<1000$, and the training data are separable, training a SVM on-line is very fast. Finally, we use the second model to segment the cell image.
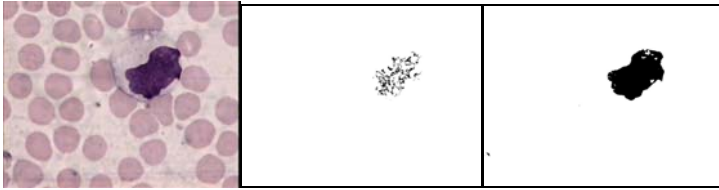
(1) Original cell image (2)Nuclei found by mean-shift (3)Nuclei found by SVM
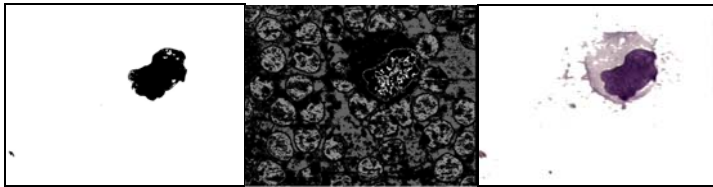
(4) Region after dilation (5)The candidates of training samples (6)Final segmentation result

Fig. 3. shows a white cell is extracted with our method (See sub-figures from 2 to 6). Obviously, the new nuclei regions found by SVM contain most of the nucleus pixels. Our method focuses on some nucleus and cytoplasm pixels rather than all of the pixels in the image.

4 Experiments and Comparisons

In our method, the most parameters could be fixed. $N=1000$, $C=100$, $q^*=1$, $p_n(R,G,B)=(0,0,0)$, $p_b(R,G,B)=(255,255,255)$, $h_n=h_e=h_b=5$. Experimental environment is follows: Hardware is P4_2.7GHz cpu, 512MB memory. Software is C++ program, Windows XP. The SVM tool is Libsvm 2.81[9].

As a comparison, thresholding and watershed based methods also perform the same segmentation task. More than 100 images (768*570) and 600 nucleated cells had been tested in our work. Some typical experiment results are shown in Fig.4. The (a1),(a2) and (a3) show three original images. Their preparation and illumination are large varied. They are hardly segmented by traditional methods since the colors of the nucleus, cytoplasm and background are often similar or confused.

Because two-class classifier finally used to achieve image segmentation, we took advantage of confusion matrix (number of false positives and false negatives) to evaluate the effect of segmentation. We first manually draw the contours of nucleated

cells in each test image, and then use the area of object (total pixels of nucleated cells) and no-object as the evaluation criteria. Comparing the results of automatic segmentation with the manual segmentation, the confusion matrix can be achieved. Table 1 shows three confusion matrixes of the compared methods average from 100 images. Here we denote the number of object pixels with NC and the number of no-object pixels with NNC, which are normalized (divided by the total number of pixels of 100 images).

Table 1. Comparison of average confusion matrix and time cost

Method	Watershed-based			Thresholding-based			SVM-based		
Confusion Matrix		NC	NNC		NC	NNC		NC	NNC
	NC	0.15	0.09	NC	0.13	0.11	NC	0.23	0.01
	NNC	0.14	0.62	NNC	0.33	0.43	NNC	0.03	0.73
Time cost	5.1 seconds			0.5 seconds			3.0 seconds		

Table 1 shows the largest confusion and faster speed appeared in thresholding-based method, the middle of the performance in the watershed-based one, and the smallest confusion and the middle speed in SVM-based one. Fig.4 (b1),(b2) and (b3) depict the corresponding segmentation result by watershed algorithm, (c1),(c2) and (c3) by thresholding method, and the (d1),(d2) and (d3) by our method.

Watershed-based algorithm needs seeds and gradient information as a reference. But in most complex scenes, since the gradient of pixels easily affected by noise, it is difficult to obtain good segmentation results. In (b1) of Fig.4, some non-object pixels are bring into the segmentation of object because of poor color contrast. A white cell with rod-shaped nuclei in right of the (b2) lost its cytoplasm since its color is very closer to the color of mature erythrocyte.

Thresholding-based method usually ignores the space information, over-segmentation often occurs (see (c2)). In most cases, multi-thresholding in a histogram is also a difficult task, yet single-thresholding on each dimension may be too coarse to distinct the similar color in color space. So over-segmentation and under-segmentation may appears simultaneously (see (c1),(c3)).

Our method seldom shows over-segmentation or under-segmentation (see (d1~d3)). Since learning by sampling strategy and more spatial information are considered in our method, whole white cells could be extracted even if poor color contrast and color confusion occurs (see Fig.4 (d1) and (d2)). It can differentiate similar color (see the cytoplasm in (d3)). Obviously, according to the comparison of the confusion matrix and subjective observation, the segmentation results of our method are closer to human visual perception.

5 Discussion and Conclusion

In this paper, we present a blood and bone marrow cell images segmentation method, which combines the mean-shift algorithm with the support vector machine. The mean-shift procedures are firstly used to overcome the change of color and select pure candidate samples. And then a SVM is trained by uniform sampling from three modes

in order to find whole nuclei regions. So we could dilate the nuclei regions only in high gradient pixels to get the part pixels of cytoplasm. Finally, we train a new SVM by a training set sampling from limited cytoplasm and three modes to extract the whole leukocytes. Due to robust and good generalization performance inherited from the mean-shift algorithm and SVM, this learning by sampling based method brings more robust and accurate segmentation.
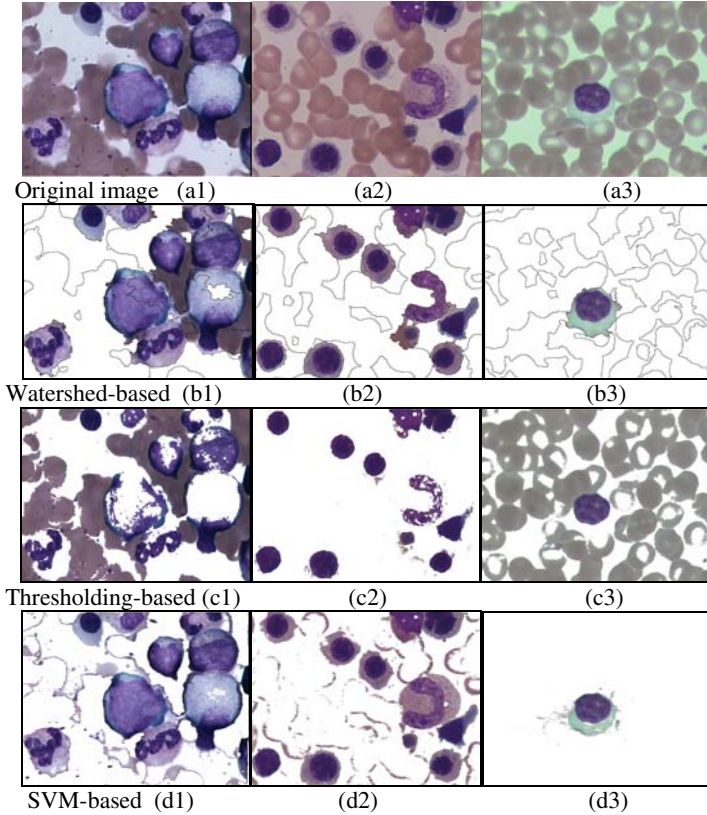
Fig. 4. Color image segmentation results achieved by the watershed-based (b1~b3), thresholding-based (c1~c3) and SVM-based methods (d1~d3)

As mentioned above, prior knowledge of cell image, color and spatial information are availably integrated into a new method. The SVM-based segmentation in joint domain succeeds in overcoming the inherent limitations of methods based only on gray-level or color.

Although there are many parameters should be set by experience, such as the three initial locations and bandwidths of mean-shift windows, the parameters of SVM and the size of the training set, those parameters are not strict. They are insensitive in our algorithm.

It is known that the performance of SVM is determined not only by its parameters, but also by its training set. In this paper, the on-line SVM models are used twice in segmentation, in which the basic parameters of SVM are fixed commonly. We only focus on finding effective small samples to training SVM according our need. With SVM, we eliminated the sensitivity of the bandwidth of mean-shift, and extracted whole nucleated cells via small samples and space information.

Acknowledgments

References

1. Wang, S.T., Wang, M.: A New Algorithm NDA Based on Fuzzy Cellular Neural Networks for White Blood Cell Detection. IEEE Trans. on Information Technology in Biomecicine 10(1), 5–10 (2006)
2. Adollah, R., Mashor, M.Y., Mohd Nasir, N.F., Rosline, H., Mahsin, H., Adilah, H.: Blood Cell Image Segmentation: A Review. In: 4th Kuala Lumpur International Conference on Biomedical Engineering, vol. 21, pp. 141–144 (2008)
3. Villegas, R., Bosnjak, A., Chumbimuni, R., Flores, E., López, C., Montilla, G.: Detection of Basal Nuclei on Magnetic Resonance Images using Support Vector Machines, vol. 22, pp. 421–424. Springer, Heidelberg (2009)
4. Chen, P., Yan, X.G., Zheng, C.X.: Hard margin SVM for biomedical image segmentation. In: Wang, J., Liao, X.-F., Yi, Z. (eds.) ISNN 2005. LNCS, vol. 3498, pp. 754–759. Springer, Heidelberg (2005)
5. Chen, P., Yan, X.G., Zheng, C.X.: Fast training of SVM for Color-based Image Segmentation. In: Proceedings of International Conference on Machine Learning and Cybernetics, vol. 6, pp. 3820–3825. IEEE, Los Alamitos (2004)
6. Comaniciu, D., Meer, P.: Mean Shift: A Robust Approach Toward Feature Space Analysis. IEEE Transactions on Pattern Analysis and Machine Intelligence 24(5), 603–619 (2002)
7. Vapnik, V.: Statistical Learning Theory. John Wiley & Sons, Chichester (1998)
8. Lezoray, O., Elmoataz, A., Cardot, H., et al.: Segmentation of Color Images from Serous Cytology for Automated Cell Classification. Analytical and Quantitative Cytology and Histology 22(4), 311–322 (2000)
9. Chang, C.C., Lin, C.J.: LIBSVM: A Library for Support Vector Machines, http://www.csie.ntu.edu.tw/~cjlin/papers/libsvm.ps.gz

Combination of Gabor Wavelets and Circular Gabor Filter for Finger-Vein Extraction

Jinfeng Yang[1], Jinli Yang[1], and Yihua Shi[2]

[1] Tianjin Key Lab for Advanced Signal Processing,
[2] College of Science Civil Aviation University of China,
P.O. Box 9, Tianjin, P.R. China, 300300
jfyang@cauc.edu.cn

Abstract. Recently, more attentions have been paid on finger-vein based personal identification. In real applications, finger-vein segmentation always is a crucial step to extract finger-vein features. Since finger-vein images usually are in low contrast, segmentation results often are the abridged versions of finger-vein networks. In this paper, we present a new method of finger-vein extraction based on combination of Gabor wavelets and a circular Gabor filter such that the finger-vein networks can be highlighted significantly as well as nonvascular region elimination. First, a family of Gabor wavelets is used to enhance vascular regions in an image. Then, image reconstruction is implemented using a combination rule. Finally, a circular Gabor filter is used for finger-vein extraction. Experimental results show that the proposed method is capable of extracting finger veins in an image reliably and effectively.

1   Introduction

As a new approach to personal identification, finger-vein recognition is becoming an active topic in biometrics. Compared with other traditional biometric characteristics (such as face, iris, fingerprints etc.), finger vein exhibits some excellent advantages in application. For instance, apart from uniqueness, universality, active liveness, permanence and measurability, finger-vein based personal identification systems are immune to counterfeit fingers and noninvasive to their users.

In anatomy, finger veins lie beneath epidermis, and form a network spreading along a finger in a high random manner. Since they are internal, visible lights usually are incapable of imaging them. Thus, illuminating the subcutaneous region of a finger properly is an important task of vein visualization. In medical applications, the NIR (near infrared) lights (760- 850nm) are often used to vein imaging because they can penetrate relatively deep into the skin as well as the radiation of lights can be absorbed greatly by the deoxyhemoglobin [1]. Hence, the NIR lights are usually used for finger-vein imaging in image acquisition system.

In our application, a finger-vein image acquisition system is designed and established as shown in Fig.1. The luminaire with 760 nm wavelength contains main NIR light-emitting diode (LED) array and two additional LED arrays, and a CCD sensor is placed underneath a finger. Here, the additional LEDs are used for enhancing the contrast between vascular and nonvascular regions. Furthermore, to reduce the

variations of finger poses, a position sensor is set to light an indicator lamp when a finger is available. In the proposed image acquisition system, an open window with a fixed size centered at the width of CCD image plane is set for finger-vein imaging. Therefore, it is convenient to chop a finger-vein image from the CCD image plane using a preset window (denoted by $W_{200\times80}$ in Fig.1).
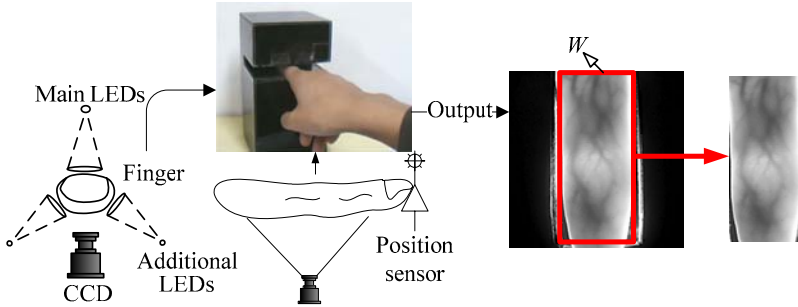
Fig. 1. The proposed principle of finger-vein acquisition system

For a finger, the network of veins is usually stable, since it can not be broken unless finger veins suffer rupture. Segmenting finger-vein networks for feature extraction therefore is an important task in finger-vein recognition [2], [3], [4], [5], [6], [7]. In practice, a finger-vein image usually appears low contrast due to a specific vein imaging mode (see Fig.1). Thus, the current methods of finger-vein image segmentation usually have poor performance in obtaining reliable finger-vein networks. This brings a big challenge for personal identification in real applications.

In computer vision, it has been demonstrated that Gabor filters were powerful in image feature analysis. And investigators have applied Gabor filters to exploit different biometric characteristics successfully, such as face [8], iris [9], [16], fingerprint [10], palmprint [11], retina [12], ear [13] etc. Considering the specific application in finger-vein image analysis, this paper proposes a finger-vein extraction method based on the combination of Gabor wavelets and a circular Gabor filter. Fig.2 illustrates the block diagram of the proposed method.
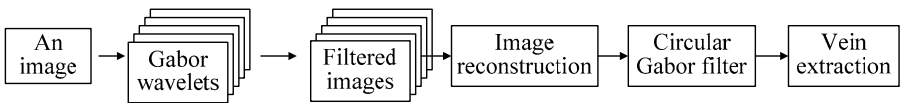
Fig. 2. Block diagram of finger-vein extraction

First, a family of Gabor wavelets is used to enhance veins in an image. Then, image reconstruction is implemented for finger-vein network regeneration. Finally, a circular Gabor filter is used for further highlighting finger-vein regions as well as nonvascular region elimination. Experimental results show that the proposed method is capable of extracting the finger veins in an image reliably and effectively, especially for images with low quality.

2   Finger-Vein Enhancement and Reconstruction

Governed by the anatomical structure of a finger, the veins often vary greatly in di-ameters and running manners. Hence, methods describing the finger-vein network textures more reliably often are desirable for finger-vein feature extraction. In the spatial domain, Gabor filters have been widely used for analyzing texture information, and have been demonstrated that they were powerful in capturing some specific local characteristics in an image. Therefore, a family of Gabor wavelets here is designed to exploit finger-vein texture information for image enhancement.

2.1   Vein Enhancement

The Gabor function proposed in [14] is inadmissible due to the d.c. response of its cosine component. This makes Gabor filters sensitive to illumination when imple-menting image operation. To make Gabor filters into admissible Gabor wavelets, the d.c. response should be compensated. Here, an admissible two-dimensional Gabor wavelet is defined as follows

$$\psi\left(x,y\right)=K\exp\left[-\pi(a^2x_r^2+b^2y_r^2)\right]\left[\exp\left(j2\pi f_0x_r\right)-\exp(-\pi f_0^2a^{-2})\right]$$

$$\begin{bmatrix}x_r\\y_r\end{bmatrix}=\begin{bmatrix}\cos\theta&\sin\theta\\-\sin\theta&\cos\theta\end{bmatrix}\begin{bmatrix}x\\y\end{bmatrix},\quad j=\sqrt{-1}\qquad\qquad,\qquad(1)$$

where K is the magnitude of the Gaussian envelop, θ is the orientation of a Gabor wavelet, f_0 denotes the filter center frequency, a and b are two space constants of the Gaussian envelope, x_r and y_r are rotated versions of two perpendicular axes of the Gaussian.

Using Euler formula, $\psi(x,y)$ can be decomposed into a real part and an imaginary part. The real part, usually called even-symmetric Gabor wavelet, is suitable for ridge detection in an image, while the imaginary part, usually called odd-symmetric Gabor wavelet, is beneficial to edge detection. Since the finger veins appear dark ridges in image plane, even-symmetric Gabor wavelets here are used to exploit the vein infor-mation in an image. The even-symmetric Gabor wavelet is represented as

$$\psi_k^e\left(x,y\right)=K\exp\left[-\pi(a^2x_r^2+b^2y_r^2)\right]\left[\cos\left(2\pi f_kx_r\right)-\exp(-\pi f_k^2a^{-2})\right],\qquad(2)$$

where k (=1, 2, ···, 8) is the channel index, θ_k (=$k\pi/8$) and f_k respectively denote the orientation and the center frequency of a even-symmetric Gabor wavelet in the kth channel. Assume that $I(x, y)$ denote a finger-vein image, $W_k(x, y)$ denote a filtered $I(x, y)$ in the kth channel, we can obtain

$$W_k\left(x,y\right)=\psi_k^e\left(x,y\right)*I(x,y),\qquad(3)$$

where $*$ denotes convolution in two dimensions. Thus, for a finger-vein image, eight filtered images are generated by Gabor wavelets.

For a specific application, the parameters f_0, a and b usually govern the optimal outputs of Gabor wavelets (see Eq.1).   Therefore, the three parameters should be determined specifically to the finger-vein textures. Considering that both the

diameters and the spread manners of veins hold high random characteristics, a here is set equal to b (i.e., Gaussian function is isotropic) for reducing diameter-deformation arising from elliptic Gaussian envelop, while the center frequency f_k is changeable with channels. To determine the relation of a and f_k, a scheme proposed in [14] is used accordingly, which is defined as follow

$$f_k / a = \sqrt{\ln 2 / \pi} \cdot (2^{\Delta F_k} + 1) / (2^{\Delta F_k} - 1) \tag{4}$$

where ΔF_k denotes a half-magnitude bandwidth in the kth channel. In our experiments, a is set to 0.2, and $\Delta F_1 < \Delta F_5 < \Delta F_2 = \Delta F_8 < \Delta F_3 = \Delta F_7 < \Delta F_4 = \Delta F_6$ ($\Delta F_\bullet \in [0.5, 2.5]$). Thus, a family of Gabor wavelets with 8 orientations and five center frequencies are designed for finger-vein information analysis. Some results are shown in Fig.3.
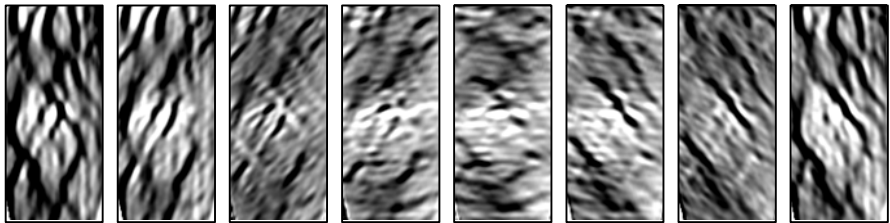
Fig. 3. The filtered images at 8 orientations 0°, 22.5°, 45°, 67.5°, 90°, 112.5°, 135°, 157.5°, $a = b = 0.2$. The original image used here corresponds to the right image in Fig.1.

From Fig.3, we can see that the veins are highlighted (darker than before) selectively and significantly at 8 orientations. Hence, an image reconstruction method is discussed subsequently to regenerate an integrated finger-vein network.

2.2 Image Reconstruction

Assume that $p(x_i, y_i)$ denote a point in $I(x, y)$, S_i denote a point set composed of 8 filtered versions of $p(x_i, y_i)$, and $R(x, y)$ denote a reconstructed image, the proposed image reconstruction rule here is defined as

$$\begin{cases} R(x_i, y_i) = \arg \min_{i \in N}(S_i) \\ S_i = \{W_1(x_i, y_i), \cdots, W_k(x_i, y_i), \cdots, W_8(x_i, y_i)\} \end{cases}, \tag{5}$$

where N is total number of pixels of a finger-vein image. The proposed reconstruction rule is effective to obtain the optimal response of Eq.3 to a pixel point in one specific orientation. This is illustrated in the middle of Fig.4, where the valleys in 3D plots correspond to the ridges in an image plane. Therefore, using Eq.5, we can obtain a reconstructed finger-vein image $R(x, y)$, as shown in the right of Fig.4. Moreover, the illumination of the reconstructed image appears more uniform than that of the original image, which illustrates that the d.c. compensation for Gabor filters can weaken illumination affect on their outputs.
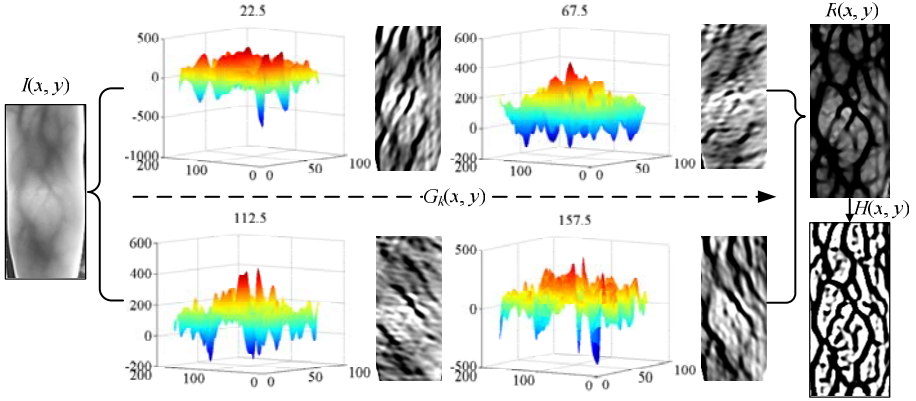
Fig. 4. The response of convolution between a finger-vein image and Gabor wavelets. The 3D plots correspond to 3D distributions of $W_k(x, y)$. Here, only four filtered results are illustrated for saving space.

Obviously, Eq.5 is able to prevent ridge information losing in the process of image reconstruction, but some noises arising from nonuniform illumination may also be strengthened in $R(x, y)$, as shown in the upper-right of Fig.4. Since noises usually hold random property, many generated false veins unlike capillary veins are often with no joints to real vein vessels. From the upper-right of Fig.4, we can also see that the contrast between capillary veins and nonvascular regions are still lower. Hence, to further explore finger-vein information, the capillary veins should be enhanced once again. Undoubtfully, spatial filter methods with orientation here are no longer suitable for re-enhancing the constructed image, since false dark regions also appear ridges. Hence, a circular Gabor filter is used subsequently to re-enhance veins contained in $R(x, y)$.

3   Finger-Vein Extraction

The circular Gabor filter proposed in [15] has a circular-symmetric sinusoidal function instead of an oriented sinusoidal function, it is defined as

$$g_c(x, y) = K_c \exp\left[-\pi(a_c^2 x^2 + b_c^2 y^2)\right] \exp\left(j2\pi f_c \sqrt{x^2 + y^2}\right) \qquad (6)$$

where K_c is the magnitude of the Gaussian envelop, a_c and b_c are the space constants that scale the two axis of the Gaussian envelop, f_c represents the center frequency of the sinusoidal function. Like Gabor filter proposed in [14], the real part of $g_c(x, y)$, called circular symmetric Gabor filter (CSF) [9], [16], is also capable of exploiting ridges in an image. The circular-symmetric Gabor filter is represented as

$$g_{CSF}(x, y) = K_c \exp\left[-\pi(a_c^2 x^2 + b_c^2 y^2)\right] \cos\left(2\pi f_c \sqrt{x^2 + y^2}\right), \qquad (7)$$

where the choice of parameters is similar to Eq.4 without channel selection.

Assume that $H(x, y)$ denote a filtered version of $R(x, y)$, we can obtain

$$H(x, y) = g_{CSF}(x, y) * R(x, y),\tag{8}$$

where $*$ denotes convolution in two dimensions. Using CSF, we can exploit the vein information again in an isotropic manner such that false dark regions appearing veins can be isolated, since they usually have no connection with positive veins. Moreover, CSF can further strengthen potential real veins as well as nonvascular region elimination according to its bandpass property. Thus, using Eq.8, the finger-vein network can be extracted from an image reliably and effectively, as shown in the right-bottom of Fig.4.

4 Experiments

We build a finger-vein image database which contains 2100 finger-vein images from 70 individuals. Each individual contributes 30 finger-vein images from three different fingers: forefinger, middle finger and ring finger (10 images per finger) of the right hand.

During the process of finger-vein collection, we find that fingers vary greatly in figuration not only from different people but also from an identical individual. So, finger-vein images vary greatly in quality though imaging environment is stable in practice. To test the performance of the proposed method in finger-vein extraction, one finger-vein image from each finger is selected as testing sample. Here, only some images with lower contrast (see Fig.6(a)) are shown for saving space. The background is not subtracted from images, since it affects little the results of finger-vein enhancement.

For image enhancement, histogram based transformation methods always are used to improve image quality. However, these methods often generate unsatisfying results when images are in lower contrast. The main reason is that gray-level transformation often is blind to regions of interest. Fig.6 (b), (c) and (d) respectively demonstrate that histogram equalization, gray-level grouping (GLG) method [17] and local histogram equalization are all not applicable for enhancing finger-vein regions.

(a)   (b)   (c)
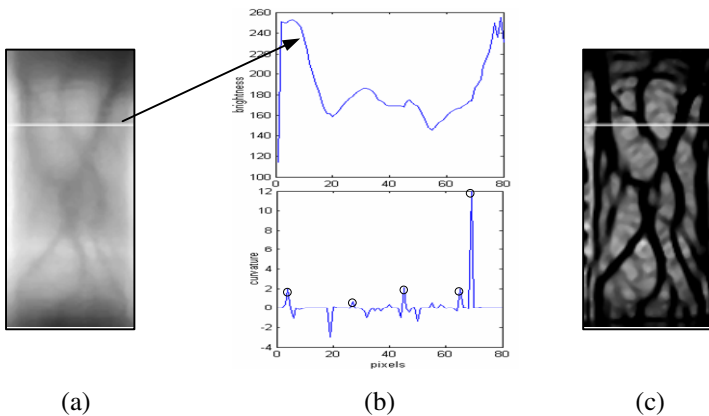
Fig. 5. Method proposed by Miura. (a) An original image. (b) A cross-sectional profile corresponding to a white line, and its curvature curve. (c) The reconstructed image.
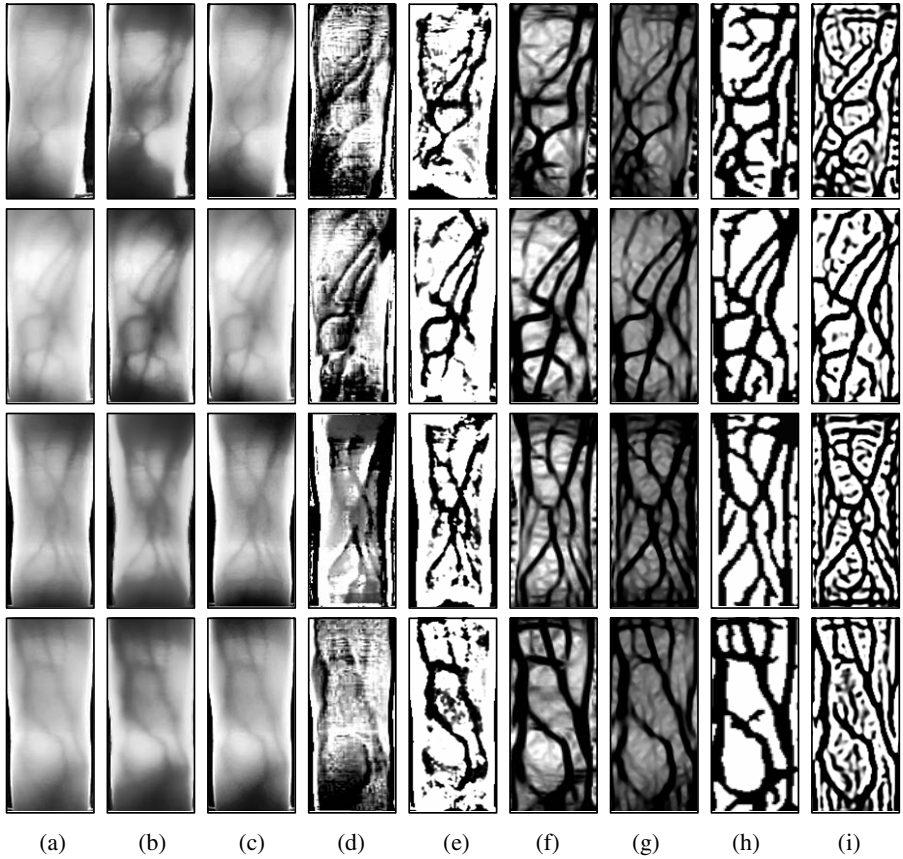
Fig. 6. Comparisons in image enhancement. (a) Some original images. (b) Histogram equalization. (c) GLG method. (d) Local histogram equalization. (e) Homomorphic filter. (f) Image reconstruction without d.c. compensation in Gabor filters. (g) Image reconstruction using the proposed method. (h) Image segmentation results based on threshold image proposed in [18]. (i) Finger-vein extraction using the proposed method.

To enhance image effectively, image filtering techniques are often used in practice. Here, homomorphic filter is an effective approach to improve image contrast, since it can improve image illumination as well as image reflectance. Compared to histogram based methods, Fig.6 (e) demonstrates that homomorphic filter is capable to enhance the finger veins greatly. However, from Fig.5 (e), we also can see that not only the information of some capillary veins is lost, but also many noises are generated accordingly.

Furthermore, Miura [3] had proposed a method of calculating local maximum curvatures in cross-sectional profiles of a vein image to extract a finger-vein network. In this method, the vein centerlines are detected by locating the positions where the curvatures of a cross-sectional profile are locally maximal. But our experiments illustrate that this method usually fails when an image with low quality. Fig.5(b) shows that five

points with local maximum value can be determined based on the principal in [3]. However, only four center points exist in this profile in practice, as show in Fig.5(c). Hence, this method is s not robust in vein network extraction in real applications.

Compared with the above methods, the proposed method exhibits an excellent performance in highlighting the underlying finger-vein information. Fig.6 (f) shows that Gabor filters without d.c. compensation are incapable of improving illumination uniformization. Moreover, the diameters of the reconstructed veins appear bigger than the originals. Based on admissible Gabor wavelets, the reconstructed results better approximate to finger-vein realities (see Fig.6(g)) in addition that many capillary veins are mined effectively.

To extract finger-vein network from an image, Wang had proposed a new method based on threshold image [18], and the segmentation results are illustrated in Fig.6(h). From Fig.6(h), we can see that this method not only makes capillary veins lost but also dilates the extracted veins, which is not beneficial to finger-vein recognition in practice. From Fig.6(i), we can see that using circular symmetric Gabor filters, the extracted finger-vein networks appear unabridged clearly. Hence, the experimental results show that the proposed method performs well in finger-vein enhancement and extraction. This is beneficial to finger-vein based personal identification.

5   Conclusion

This paper has proposed a new method based on Gabor wavelets and circular Gabor filter for finger-vein enhancement and extraction. First, even-symmetric Gabor wavelets with 8 orientations were used to enhance vein regions in an image. Then, using a fusion rule, a finger-vein image was reconstructed based on the filtered images. Finally, the finger veins were extracted using a circular Gabor filter. Experimental results have shown that the proposed method performed better in finger-vein extraction.

Acknowledgements

References

1. Zharov, V., Ferguson, S., Eidt, J., Howard, P., Fink, L., Waner, M.: Infrared Imaging of Subcutaneous Veins. Lasers in Surgery and Medicine 34(1), 56–61 (2004)
2. Miura, N., Nagasaka, A.: Feature Extraction of Finger-Vein Pattern Based on Repeated Line Tracking and Its Application to Personal Identification. Machine Vision and Applications 15(4), 194–203 (2004)
3. Miura, N., Nagasaka, A., Miyatake, T.: Extraction of Finger-Vein Patterns Using Maximum Curvature Points in Image Profiles. In: IEICE - Transactions on Information and Systems, pp. 1185–1194 (2007)

4. Lian, Z., Rui, Z., Yu, C.: Study on the Identity Authentication System on Finger Vein. In: International Conference on Bioinformatics and Biomedical Engineering, pp. 1905–1907 (2008)

5. Zhang, Z., Ma, S., Han, X.: Multiscale Feature Extraction of Finger-Vein Patterns Based on Curvelets and Local Interconnection Structure Neural Network. In: International Conference on Pattern Recognition, pp. 145–148 (2006)

6. Vlachos, M., Dermatas, E.: Vein segmentation in infrared images using compound enhancing and crisp clustering. In: Gasteratos, A., Vincze, M., Tsotsos, J.K. (eds.) ICVS 2008. LNCS, vol. 5008, pp. 393–402. Springer, Heidelberg (2008)

7. Zhang, L., Zhang, R., Yu, C.B.: Study on the Identity Authentication System on Finger Vein. In: International Conference on Bioinformatics and Biomedical Engineering, pp. 1905–1907 (2008)

8. Jie, Z., Ji, Q., Nagy, G.: A Comparative Study of Local Matching Approach for Face Recognition. IEEE Transactions on Image Processing 16(10), 2617–2628 (2007)

9. Ma, L., Tan, T., Wang, Y., Zhang, D.: Personal Identification Based on Iris Texture Analysis. IEEE Trans. on Pattern Analysis and Machine Intelligence 25(12), 1519–1533 (2003)

10. Jain, A.K., Chen, Y., Demirkus, M.: Pores and Ridges: High-Resolution Fingerprint Matching Using Level 3 Features. IEEE Trans. on Pattern Analysis and Machine Intelligence 29(1), 15–27 (2007)

11. Laadjel, M., Bouridane, A., Kurugollu, F., Boussakta, S.: Palmprint Recognition Using Fisher-Gabor Feature Extraction. In: IEEE International Conference on Acoustics, Speech and Signal Processing, pp. 1709–1712 (2008)

12. Daugman, J.G.: High Confidence Visual Recognition of Persons by A Test of Statistical Independence. IEEE Trans. on Pattern Analysis and Machine Intelligence 15(11), 1148–1160 (1993)

13. Arbab-Zavar, B., Nixon, M.S.: Robust Log-Gabor Filter for Ear Biometrics. In: International Conference on Pattern Recognition, pp. 1–4 (2008)

14. Daugman, J.G.: Uncertainty Relation for Resolution in Space, Spatial Frequency, and Orientation Optimized by 2D Visual Cortical Filters. Journal of the Optical Society of America 2(7), 1160–1169 (1985)

15. Zhang, J., Tan, T., Ma, L.: Invariant Texture Segmentation Via Circular Gabor Filters. In: International Conference on Pattern Recognition, vol. 2, pp. 901–904 (2002)

16. Ma, L., Wang, Y., Tan, T.: Iris Recognition Using Circular Symmetric Filters. In: International Conference on Pattern Recognition, vol. II, pp. 414–417 (2002)

17. Chen, Z., Abidi, B.R., Page, D.L., Abidi, M.A.: Gray-Level Grouping (GLG): An Automatic Method for Optimized Image Contrast Enhancement-Part I: The Basic Method. EEE Trans. on Image Processing 15(8), 2290–2302 (2006)

18. Wang, K., Yuan, Z.: Finger Vein Recognition Based on Wavelet Moment Fused with PCA Transform. Pattern Recognition and Artificial Intelligence 20(5), 692–697 (2007) (in Chinese)

Recent Progress of the Quasientropy Approach to Signal and Image Processing

Yang Chen and Zhimin Zeng

School of Information Science and Engineering, Southeast University,
Nanjing 210096, China
cheny@seu.edu.cn

Abstract. The quasientropy (QE) is a class of infinitely many functions of probabilities that is similar to the Shannon entropy. In this paper, we review the application of the QE approach to independent component analysis (ICA) and chaotic time series analysis. We also report the new progress of the QE approach to textural features extraction in image processing.

Keywords: Quasientropy, Signal processing, Image processing.

1   Introduction

The entropy was originally a concept in thermal dynamics and statistical physics. It is a measure of the degree of disorder [1]. The entropy as a measure of the amount of information, which has the same mathematical formulation as the entropy in thermal dynamics, was proposed by Shannon [2]. Although the mathematical expression that satisfies all the properties that entropy should possess outlined by Shannon is unique, as long as we modify the assumptions on entropy in some way, it is possible to propose other definitions of entropy. Ever since the 1960s, Renyi [3], Csiszar [4], Kapur [5], and some other researchers proposed various forms of non-Shannon entropies. For example, Renyi suggested some properties that the entropy should satisfy, such as the additivity, and proposed a class of entropies with a parameter. The Shannon entropy is the limit of these entropies when the parameter approaches 1 [3]. Nevertheless, up to now, when we talk about the (information) entropy, we routinely refer to the Shannon entropy. We follow this convention, too, hereinafter. In recent years, we have been focusing on the study of the quasientropy (QE) proposed in [6], which is a class of infinitely many functions of probabilities similar to the entropy. In this paper, we shall report the current progress of our study of the QE approach to signal and image processing.

The rest of the paper is organized as follows. The general principle of QE is explained in Section 2. The subsequent Sections 3 and 4 are devoted, respectively, to the applications of QE to independent component analysis (ICA) [6], [7], and chaotic time series analysis [8]. In Section 5, we describe the newly developed QE approach to textural features extraction in image processing. Finally, we summarize the paper in Section 6.

2   Principle of Quasientropy (QE)

Let us start from a general definition of QE. Suppose we have a set of possible events whose probabilities of occurrence are $p_1, p_2, \ldots, p_n$ with $\sum_{i=1}^{n} p_i = 1$. Then, the QE is defined as

$$\beta = \sum_{i=1}^{n} f(p_i), \tag{1}$$

where $f(\cdot)$ is a strictly convex function on $[0,1]$. Clearly, when $f(u) = u \log u$, up to a minus sign, QE degenerates to the entropy

$$H = -\sum_{i=1}^{n} p_i \log p_i. \tag{2}$$

The entropy and the QE share the three common properties as follows [6]. For comparison, we list the corresponding properties of entropy and QE in pairs.

Property 1 of entropy: If and only if the probabilities are uniform, the entropy is maximal.
Property 1 of QE: If and only if the probabilities are uniform, the QE is minimal.

Property 2 of entropy: If and only if the probabilities are the most concentrated, the entropy is minimal.
Property 2 of QE: If and only if the probabilities are the most concentrated, the QE is maximal.

Property 3 of entropy: For the uniform probabilities, the entropy is a monotonic increasing function of the number of events.
Property 3 of QE: For the uniform probabilities, the QE is a monotonic decreasing function of the number of events, provided that $f(0) = 0$.

Generally, the β defined in (1) does not satisfy the additivity property of entropy. Therefore, it is more appropriate to nominate β as quasientropy instead of some entropy, which means that β is something similar to, but not exactly is, the entropy. From the Properties 1 and 2 of entropy and QE, we immediately get the following theorem.

Theorem 1: The QE is minimal if and only if the entropy is maximal. The QE is maximal if and only if the entropy is minimal.

Theorem 1 shows that the entropy and the QE are equivalent in terms of maximization and minimization. The only difference is a negative sign. If a problem can be solved by maximizing (minimizing) the entropy, then this problem can also be solved by minimizing (maximizing) the QE.

Properties 1 and 2 of entropy and QE also show that both the entropy and the QE can be viewed as measures of uniformity or nonuniformity.

3 Independent Component Analysis (ICA)

Independent component analysis (ICA) [9], sometimes also referred to as blind source separation (BSS) [10], is the task of recovering unknown independent source signals from their mixtures without knowing the mixing coefficients. Suppose we have n mixture signals collected in $x(t) = \left[x_1(t), \ldots, x_n(t) \right]^T$ and described by

$$x(t) = As(t) \tag{3}$$

where, $s(t) = \left[s_1(t), \ldots, s_n(t) \right]^T$ is a vector of n unknown independent source signals, and A is an unknown $n \times n$ mixing matrix. The goal of ICA is to estimate $s(t)$ or A from $x(t)$.
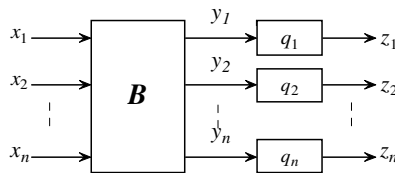
Fig. 1. Illustration of maximum entropy approach to ICA

A mainstream class of ICA algorithms is based on the maximum entropy [11], or infomax [12], approach, as illustrated in Fig. 1. $x_1, \ldots, x_n$ are the n mixture signals. B is the de-mixing matrix to be estimated. $y_1, \ldots, y_n$ are the output de-mixed signals that are expected to be a copy of the unknown independent source signals, which can be realized if and only if $y_1, \ldots, y_n$ are independent. $q_1, \ldots, q_n$ are the n hypothesized source cumulative distribution functions (CDFs). Suppose that $q_1, \ldots, q_n$ are exactly the CDFs of the recovered sources. Then, $(z_1, \ldots, z_n)$ will be uniformly distributed in $[0,1]^n$ when $y_1, \ldots, y_n$ are independent, and at this time, $H(z_1, \ldots, z_n)$, the joint (differential) entropy of $z_1, \ldots, z_n$, is maximized.

Suppose that we uniformly quantize $z_1, \ldots, z_n$, respectively, using l quantization levels to get $k_1, \ldots, k_n \in \{1, \ldots, l\}$. Then, when $y_1, \ldots, y_n$ are independent, $(z_1, \ldots, z_n)$ will be uniformly distributed in $[0,1]^n$, and thus, $(k_1, \ldots, k_n)$ will be uniformly distributed in $\{1, \ldots, l\}^n$. Therefore, the QE of the joint probabilities of $k_1, \ldots, k_n$, which is defined as the QE of $y_1, \ldots, y_n$ in [6], will be minimized according to Property 1 of QE. This leads to the minimum QE approach to ICA. Actually, the QE based ICA algorithm

is implemented in a pairwise iterative scheme [6]. Besides, there is a subclass of QE for measuring independence that can be realized by simply counting the number of occupied grid-boxes, which is called grid occupancy (GO) [7]. Because we utilize order statistics to directly estimate QE and GO from observations, the QE and GO based ICA algorithms can separate signals with arbitrary continuous distributions [6].

4   Chaotic Time Series Analysis

Delay coordinate reconstruction is a fundamental approach in chaotic time series analysis. It takes the current observed value $e(t)$ and the values after equally spaced time lags $e(t+\tau)$, $e(t+2\tau)$, $e(t+3\tau)$ ... of the one dimensional time series as the "observations" of several reconstructed variables of a multidimensional dynamical system, where τ is the delay. The first minimum of mutual information (MI) is the most widely used criterion for choosing τ [13]. Since the MI here is actually used as a measure of independence, we proposed to use the QE for measuring independence developed in the above ICA scenario as alternative criteria for choosing τ [8].

As an example, in Fig. 2, we reproduce four of the plots in [8]. They are the QE plots of the two reconstructed variables of Rossler chaotic system versus delay using four different convex functions, among which Fig. 2(b) is the minus entropy. We proposed a quality factor to characterize the sharpness of the minimum of QE [6]. In

(a) $f(u) = -\sqrt{u}$, $\mathcal{O}(l^2)$

(b) $f(u) = u\log_2 u$, $\mathcal{O}(l^2/\ln l)$

(c) $f(u) = u^{1.001}$, $\mathcal{O}(l^{1.999})$

(d) $f(u) = -\sin \pi u$, $\mathcal{O}(1)$
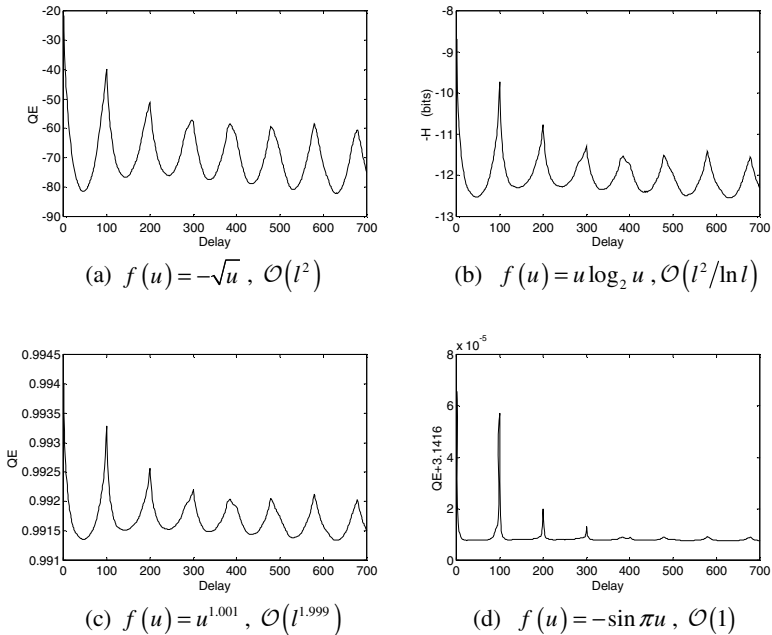
Fig. 2. QE of delay-coordinate variables of Rossler chaotic system versus delay. The convex function used and the order of quality factor are labeled under each subplot.

Fig. 2, we can see that the order of quality factor with respect to l, the number of quantization levels, correctly predicts the sharpness of the minima of QE, i.e., the higher the order, the sharper the minima. The abscissas of the first (leftmost) minima in Fig. 2(a)–2(c) are good choices of τ. In addition, the QE in Fig. 2(a) is better than the minus entropy in Fig. 2(b) since it can provide more accurate position of τ with its sharper minimum.

5   Image Processing

In this section, we shall report a new progress of the QE approach to image processing. As we mentioned at the end of Section 2, both the entropy and the QE can be viewed as measures of uniformity or nonuniformity. If there is some textural structure in an image, some kind of nonuniformity must exist. Therefore, we attempt to apply QE to textural features extraction.

5.1   Gray Level Co-occurrence Matrix (GLCM)

In 1973, Haralick et al. proposed the gray level co-occurrence matrix (GLCM), and 14 features based on GLCM [14]. In 1992, Ohanian and Dubes compared four types of textural features, namely Markov Random Field parameters, multi-channel filtering features, fractal based features, and co-occurrence features, and found that co-occurrence features perform best [15].

Suppose that $g(k,m) \in \{1,\dots,N_g\}$ is the gray level value of the (k,m) th pixel in a two dimensional image of size $N_y \times N_x$, where N_g is the number of gray levels. Let $L_x = \{1,\dots,N_x\}$, $L_y = \{1,\dots,N_y\}$. Then, the elements in a GLCM for direction angle θ are defined as [14]

$$P_\theta(i,j) = \text{Num}\left\{\left((k,m),(k+\Delta y, m+\Delta x)\right) \in \left(L_y \times L_x\right) \times \left(L_y \times L_x\right)\right.$$
$$\left. \Big| g(k,m) = i, g(k+\Delta y, m+\Delta x) = j\right\}. \tag{4}$$

Namely, $P_\theta(i,j)$ is the number of occurrence that two pixels spaced at a certain distance and direction have gray level values i and j, respectively. Therefore, P_θ is an $N_g \times N_g$ matrix. Especially, we take four directions $0°$, $45°$, $90°$, and $135°$, and let

$$\begin{cases} \Delta x = \pm 1, \Delta y = 0 & \text{if } \theta = 0°, \\ \Delta x = 1, \Delta y = 1 \text{ or } \Delta x = -1, \Delta y = -1 & \text{if } \theta = 45°, \\ \Delta x = 0, \Delta y = \pm 1 & \text{if } \theta = 90°, \\ \Delta x = -1, \Delta y = 1 \text{ or } \Delta x = 1, \Delta y = -1 & \text{if } \theta = 135°. \end{cases} \tag{5}$$

5.2   QE of GLCM

One of the features that were proposed in [14] is the entropy of GLCM. Before calculating the entropy, we should normalize the elements of P_θ so that they sum up to 1 as follows.

$$p_\theta(i,j) = \frac{P_\theta(i,j)}{\displaystyle\sum_{i=1}^{N_g}\sum_{j=1}^{N_g} P_\theta(i,j)}. \tag{6}$$

To be consistent with the sign of QE, we define the minus entropy of GLCM as follows.

$$-H = \sum_{i=1}^{N_g}\sum_{j=1}^{N_g} p_\theta(i,j)\log p_\theta(i,j). \tag{7}$$

Now, we define the QE of GLCM as follows.

$$\beta = \sum_{i=1}^{N_g}\sum_{j=1}^{N_g} f\big(p_\theta(i,j)\big), \tag{8}$$

where $f(\cdot)$ is a strictly convex function on $[0,1]$. Note that the angular second moment (ASM), another feature proposed in [14], is actually the QE of GLCM when $f(u) = u^2$. When the gray level values are smooth along direction θ, $p_\theta(i,j)$ will concentrate along the main diagonal $i = j$, and β will therefore be big according to Property 2 of QE. When the gray level values are rough along direction θ, $p_\theta(i,j)$ will become scattered, and β will therefore be small according to Property 1 of QE.

5.3   Experimental Results

Next, we show some experimental results. Fig. 3(a) is the original 256×256 Lena image. We reduce the number of gray levels from 256 to 8 and obtain the image in Fig. 3(b). We slide a 5×5 window pixel by pixel all over the image in Fig. 3(b). We calculate the QE of the GLCM of each 5×5 block and assign the value to the center point of that block. The first (last) row (column) of the image is replicated twice so that the values at the edges can be calculated. The QE of four directions, $0°$, $45°$, $90°$, and $135°$, are averaged to nullify direction information.

Fig. 4(a)–4(d) show the averaged QE image where the convex functions are identical with those in Fig. 2(a)–2(d), respectively, among which Fig. 4(b) is the minus entropy. Comparing Fig. 2(a)–2(d) with Fig. 4(a)–4(d), we can see clear correspondence between the shape of the QE curve and the appearance of the QE image. When the QE has sharp minima, it accentuates contrasty parts of the image, which are mainly boundaries between persons or objects. When the QE has blunt minima, it begins to emphasize less contrasty parts of the image, such as the skins of persons or the surfaces of objects. The QE curves Fig. 2(b) and 2(c) look similar in shape. Thus, the QE images Fig. 4(b) and 4(c) also look similar in appearance. By using different convex functions, we can adjust the degree to which the gentle textures on surfaces are to be emphasized, i.e., less than, equal to, or more than the degree of using the minus entropy. The QE with convex function $f(u) = -\sqrt{u}$ has sharper minimum than the minus entropy. Thus, it is more selective in detecting edges and gives a smoother look to Lena's face and shoulder than the minus entropy does, as can be seen by comparing Fig. 4(a) with 4(b).

(a) Original 256×256 image   (b) 256×256 image with 8 gray levels

Fig. 3. Lena image

(a) $f(u) = -\sqrt{u}$ (b) $f(u) = u \log_2 u$

(c) $f(u) = u^{1.001}$ (d) $f(u) = -\sin \pi u$

Fig. 4. Image of QE of GLCM averaged over four directions using 5×5 sliding window. The convex function used is labeled under each subplot.

6   Conclusions

We have reviewed the QE approaches to ICA and chaotic time series analysis, and presented the new progress of QE approach to textural features extraction. All these applications utilize the property that QE is a measure of uniformity or nonuniformity. Since the additivity of entropy is not necessary in solving many problems, QE can be applied to many areas that the entropy has been applied to. By selecting different convex functions, QE can achieve more abundant or even better results than the entropy.

Acknowledgements

References

1. Zemansky, M.W.: Heat and Thermodynamics. McGraw-Hill, New York (1968)
2. Shannon, C.E.: A Mathematical Theory of Communication. The Bell System Technical Journal 27, 379–423, 623–656 (1948)
3. Renyi, A.: On Measures of Entropy and Information. In: Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability, vol. 1, pp. 547–561 (1961)
4. Csiszar, I.: Information-Type Measures of Difference of Probability Distributions and Indirect Observations. Studia Scientiarum Mathematicarum Hungarica 2, 299–318 (1967)
5. Kapur, J.N.: Measures of Information and Their Applications. John Wiley & Sons, New York (1994)
6. Chen, Y.: Blind Separation Using Convex Functions. IEEE Trans. Signal Processing 53(6), 2027–2035 (2005)
7. Chen, Y.: A Novel Grid Occupancy Criterion for Independent Component Analysis. IEICE Trans. Fundamentals E92-A(8) (2009)
8. Chen, Y., Aihara, K.: New Results on Criteria for Choosing Delay in Strange Attractor Reconstruction. In: Huang, D.-S., Wunsch II, D.C., Levine, D.S., Jo, K.-H. (eds.) ICIC 2008. LNCS, vol. 5226, pp. 946–953. Springer, Heidelberg (2008)
9. Comon, P.: Independent Component Analysis, A New Concept?. Signal Processing 36, 287–314 (1994)
10. Jutten, C., Herault, J.: Blind Separation of Sources, Part I: An Adaptive Algorithm Based on Neuromimetic Architecture. Signal Processing 24, 1–10 (1991)
11. Yang, H.H., Amari, S.: Adaptive Online Learning Algorithms for Blind Separation: Maximum Entropy and Minimum Mutual Information. Neural Computation 9, 1457–1482 (1997)
12. Bell, A.J., Sejnowski, T.J.: An Information-Maximization Approach to Blind Separation and Blind Deconvolution. Neural Computation 7, 1129–1159 (1995)
13. Fraser, A.M., Swinney, H.L.: Independent Coordinates for Strange Attractors from Mutual Information. Physical Review A 33(2), 1134–1140 (1986)
14. Haralick, R.M., Shanmugam, K., Dinstein, I.: Textural Features for Image Classification. IEEE Trans. Systems, Man, and Cybernetics SMC-3(6), 610–621 (1973)
15. Ohanian, P.P., Dubes, R.C.: Performance Evaluation for Four Classes of Textural Features. Pattern Recognition 25(8), 819–833 (1992)

Gait Recognition Using Hough Transform and Principal Component Analysis

Ling-Feng Liu[1,2], Wei Jia[1], and Yi-Hai Zhu[1,2]

[1] Hefei Institute of Intelligent Machines, CAS, Hefei, China
[2] Department of Automation, University of Science and Technology of China
fengling963@gmail.com

Abstract. In this paper, we propose a new spatio-temporal representation for gait recognition. Firstly, the new representation of gait is constructed, which is the average of the Hough transformed images in one complete cycle of a silhouette sequence. Secondly, we project the new representation to low dimension by applying Principal Component Analysis. Finally, the nearest neighbor rule is adopted for recognition. The results of experiments conducted on CASIA-A Gait Database show that the proposed gait recognition approach can obtain encouraging accurate recognition rates.

Keywords: Gait Recognition, Hough Transform, PCA.

1 Introduction

Nowadays, the demand of automatic human identification technology is increasing in many important applications, such as visual surveillance, access control, e-business, and human-machine interface etc. There are many kinds of human physical or behavioral features which are often used for human identification such as fingerprint, voice, face, iris, hand signature, and gait etc. For the advantages of being non-contact, non-invasive, easily gained at a distance and difficult to conceal, gait recognition, which identifies people based on the way they walk or run, can be effectively used in many security-sensitive environments, such as airports, banks, car parks and so on. As a result, it has recently gained growing interests from the researchers.

There are a lot of interesting approaches proposed for gait recognition, which can be roughly divided into two categories: model-based approaches and model-free approaches [1, 2]. Model-based approaches generally aim to recover gait features, such as stride dimensions, limb lengths, and kinematics of joint angles. Meyer et al. [3] extracted and tracked the contours of different parts of the human body. Lee et al. [4] fitted seven ellipses in the human body area, and used their locations, orientations, and aspect ratios as features to represent the gait. In [1], Zhang et al. proposed a novel two-step, model-based approach to recognize gait by employing a five-link biped locomotion human model. Generally speaking, all features used in these model-based approaches are insensitive to noise and background cluttering. However, Mode-based approaches are not practical in real environment since the processing on feature extraction and matching is very complicated.

Model-free approaches focus on the spatio-temporal information contained in the silhouette images. In [5], a statistical approach was proposed to extract features from spatial and temporal templates. In [6], key frames from a gait sequence were compared to training frames using normalized correlation, and then classification was performed by nearest neighbor rule. Ju et al. [7] adopted Hough Transform to extract the lines representing legs, and calculated inclination angles between upper legs and lower legs, which is used for final recognition. Han et al. [8] proposed the Gait Energy Image (GEI) to characterize human walking properties for individual recognition by gait. Generally speaking, it is easy to extract the gait features in those Model-free approaches. On the other hand, Model-free approaches have lower computational complexity. Thus, they are more suitable for being used in the real-time system.

It is well-known that the information of straight lines in gait silhouettes is very important for gait recognition. In this paper, motivated by the work of [12], we present a new spatio-temporal gait representation based on the Hough transform. The new representation contains more straight lines information, and is more insensitive to image noise. Finally, we use PCA for feature extraction and NN for classification. An encouraging result of experiment conducted on CASIA-A Gait database is obtained.

2   Preprocessing of Gait Sequences

2.1   Aligning Gait Sequences

As we know, when a person is walking, the relative position between him and the camera will be changed. That is, the nearer the walking person gets to the camera, the bigger the gait silhouette image will be. Thus, the preprocessing is needed to align the binary gait silhouettes before constructing templates and feature extraction. Here, we assume that silhouettes have been extracted from original human walking sequences [9]. And we denote the binary gait silhouette by:

$$s(i, j) = \begin{cases} 1 & \text{if } (i, j) \text{ belongs to the foreground} \\ 0 & \text{otherswise} \end{cases} \tag{1}$$

The area center (i_c, j_c) of each silhouette can be computed by:

$$i_c = \frac{1}{N} \sum_{i,j} i \cdot s(i, j) \quad j_c = \frac{1}{N} \sum_{i,j} j \cdot s(i,j) \tag{2}$$

where N is the number of foreground pixels, given by $N=\sum S(i,j)$. We resize gait silhouettes so that all silhouettes have the same height (the images are 120×120 in our approach), and then center each silhouette image according to the horizontal center i_c [10]. Fig.1 shows several gait silhouette images after aligning.

Fig. 1. The aligned gait silhouette images

2.2 Gait Cycle Detection

Before constructing the Hough Template, another important work is to compute the periodicity of walking in a gait sequence. We know that when a person is walking, the maximum width of his/her lower half silhouette presents obviously periodic change. We define the maximum width in a gait sequence as $V=[v_1,v2,...,vn]$, as shown in Fig. 2a. Since the maximum value in Fig. 2a is noisy, we identify the cycle length T_G by the approach presented in [11] which computed the autocorrelation coefficient $R=[r_1,r_2,...,r_n]$ (see Fig.2b) of the signal V. The autocorrelation is defined as:

$$R_m = \frac{1}{n}\sum_{k=1}^{n-m}(v_k - \mu)(v_{m+k} - \mu)$$

(3)

(a)   (b)
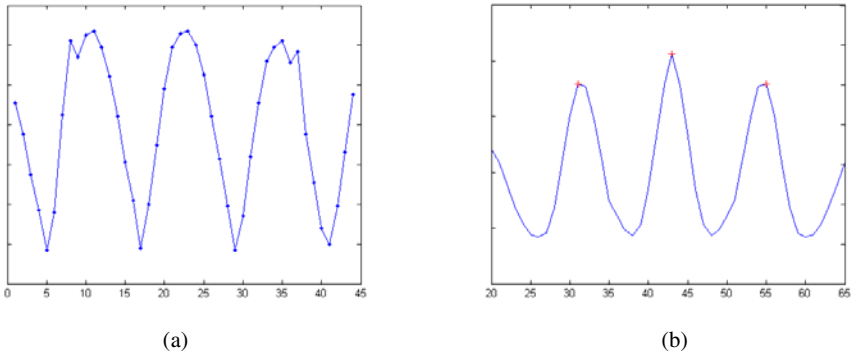
Fig. 2. (a) The maximum width V in a gait sequence (b) the autocorrelation coefficient

We can see that the curve of autocorrelation in Fig. 2b is more smoothing than Fig. 2a. Thus, we will determine the T_G from the autocorrelation curve. In Fig. 2b, the local maximum points of the curve are labeled with '+', and their locations are 31, 43, and 55, respectively. Since the gait sequence is highly relative to the translation sequence after half gait cycle, the cycle length T_G here is 24 frames.

3   Constructing Hough Transformed Template

3.1 Hough Transform

Hough Transform, named after Paul Hough who patented the method in 1962, is a feature extraction technique which can be used to isolate features of a particular shape within an image [7]. For finding imperfect instances of objects, it requires that the particular shape can be described by the mathematical functions that describe the boundary curves. Because of the computational complexity, Hough Transform is normally restricted to first and second order equation. Generally, the classical Hough Transform is most commonly used for the detection of regular curves such as lines, circles, and ellipses, etc.

In the image space, a straight line can be described as $y=ax+b$. We treat the two constants a and b as parameters representing the lines. Obviously, the characteristics of one straight line are not x and y, but its slope a and intercept b. As a result, one point in the image space corresponds to a line in the slope-intercept space. As shown in Fig. 3b, the intersection points in slope-intercept space estimate the line in image space which is shown in Fig. 3a.
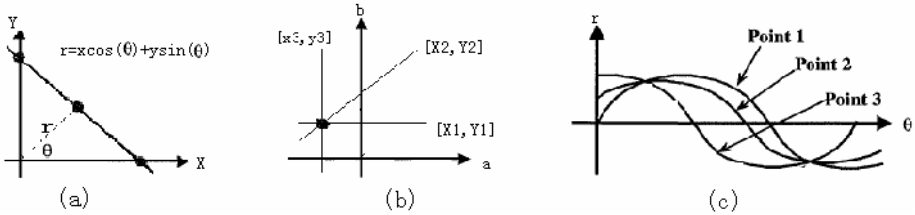
Fig. 3. Hough Transform

Since both a and b parameters are unbounded, a straight line can be represented by its shortest distance from the origin (r) and its orientation (θ) (See Fig. 3a). Thus, the straight line can be described as $r=x\cos(\theta)+y\sin(\theta)$. In this time, a point in the image space corresponds to a sine wave in r-θ space. As shown in Fig. 3c, the intersection points in r-θ space can estimate the line in the image space. A voting procedure is carried out in the parameter space, from which object candidates are obtained as local maxima. So the parameter space is often called the accumulator space.

3.2   Gait Template Construction

We use the Laplacian of Gaussian method to detect edges in gait intensity images, as shown in Fig. 4a. Then we establish a mapping between the image space and the accumulator space when the Hough Transform is applied to edges in gait intensity image. Let H_t denote the results in the accumulator space which are transformed from edges in gait intensity image B_t at time t, as shown in Fig.4.

(a)   (b)
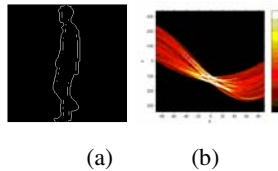
Fig. 4. Hough transform of edges in a gait intensity image. (a) edges in gait intensity image (b) transformed edges in accumulator space.

Jing et al. used Hough Transform to extract the lines which represent legs, and obtained inclination angles between upper legs and lower legs [7]. In our paper, we use the results H_t in accumulator space to construct the gait template. Given a person's gait transformed image H_t, the Hough template is defined as follows:

$$G \; = \; \frac{1}{N} \sum_{t=1}^{N} H_t \tag{4}$$

where N is the number of frames in the complete cycle(s) of a silhouette sequence, t is the frame number in the sequence (moment of time) [8]. Fig.5 is the sketch map for computing a gait template. Fig. 6 shows gait Hough templates coming from 4 persons.
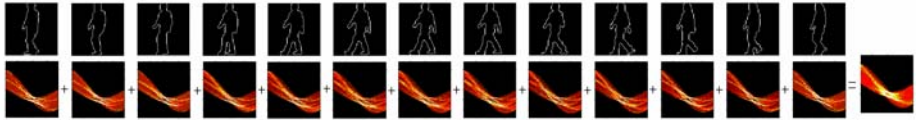
Fig. 5. An example of a gait squence. The rightmost image is the corresponding gait template
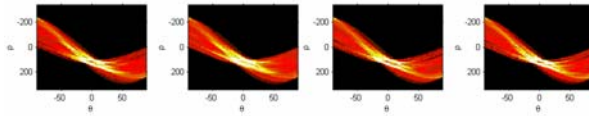
Fig. 6. Gait templates based on Hough Transform

Compared to [7], our approach, constructing the gait Hough template with the transformed images in accumulator space, takes account of more straight lines information in the gait intensity image. Another advantage of the Hough template is that it is insensitive to image noise since it computes a global description of features. The angles which the leg and arm axes form with the horizontal axis vary widely. Same as the Radon template proposed in [12], the Hough template can guarantee that much of the energy of the original silhouettes appears in G.

4   PCA for Feature Generation

Performing recognition in a high dimensional space can significantly diminish the efficiency of system performance. Principal Component Analysis (PCA) is widely utilized to reduce the dimensionality of the data. The goal of PCA is to reduce the dimensionality of the data while retaining as much as possible of the variation present in the original dataset. PCA allows us to compute a linear transformation that maps data from a high dimensional space to a lower dimensional space.

Suppose we have $x_1, x_2, \ldots, x_M$ templates for training which are $N \times 1$ vectors. Similar to the illustration in [13], the mean $\bar{x}$ and the global covariance matrix C of the training samples are computed:

$$\bar{x} = \frac{1}{M} \sum_{i=1}^{M} x_i \tag{5}$$

$$C = \frac{1}{M} \sum_{n=1}^{M} (x_n - \bar{x})(x_n - \bar{x})^T \tag{6}$$

If the rank of the matrix C is N, then we can compute N eigenvalues $\lambda_1 > \lambda_2, \ldots, > \lambda_N$ and the associated eigenvectors $e_1, e_2, \ldots, e_N$ based on Singular Value Decomposition (SVD). We keep only the terms corresponding to the k largest eigenvalues and their corresponding eigenvectors using a threshold value

$$W_k = \sum_{i=1}^{k} \lambda_i \Bigg/ \sum_{i=1}^{N} \lambda_i > T_s \tag{7}$$

where W_k is the accumulated variance. In our experiments, T_s is chosen as 0.99. Accordingly, a Hough gait template G can be mapped into a point G_p in the k-dimensional eigenspace.

$$G_p = [e_1, e_2, \cdots, e_k]^T G \tag{8}$$

5   Experimental Results

Our experiments are carried out on the CASIA-A database [16]. All persons walk on a straight line under normal conditions. And every subject is captured in three different view angles i.e., frontally (90º), laterally (0º), and obliquely (45º), respectively [14]. The database consists of 20 different persons. Each person has 4 sequences per view. The database thus includes a total of 240 (20×4×3) sequences. The length of each collected sequence varies with the pace of the walker, but the average is about 90 frames [14].

In our experiments, we apply the nearest neighbor rules for classification and the leave-one-out cross-validation rule to compute an unbiased estimate of the true classification rate. We compared our approach to GEI [8] and GRRT [2] on the CASIA-A gait database. The correct classification rates (CCR) has been calculated and the results are summarized in Table 1, in which our methodology is denoted as GRHT.

Table 1. CCRs under different viewing angles

Methodology	CCRs (%)		
	0°	45°	90°
GEI [8]	97.50	96.25	96.25
GRRT [2]	92.50	90.00	95.00
GRHT	97.50	98.75	96.25

From Table 1, it can be seen that the whole results of the methodology proposed in our paper are encouraging, with mean CCR is 97.5%.

Another useful classification performance measure that is probably more general than CCR is the Rank Order Statistic [15]. It is defined as the cumulative probability P that the real class of a test measurement is among its top k matches. The performance statistics are reported as the cumulative match scores (CMS). Under the closed-universe assumption, we report in Table 2, CMS at rank 1, rank 5 and rank 10 and plot on graphs under different views where the horizontal axis is rank and the vertical axis is the probability of identification, as see in Fig.7. In Table 2, rank1 performance means the percentage of the correct subjects appearing in the first place of the retrieved rank list. In Fig.7, we plot respectively CMS in three different views.

Table 2. CMS (in percentage) at rank 1 rank 5 and rank 10 under laterally

Methodology	Rank 1	Rank 5	Rank 10
GEI	97.50	98.75	98.75
GRRT	92.50	97.50	100
GRHT	97.50	100	100
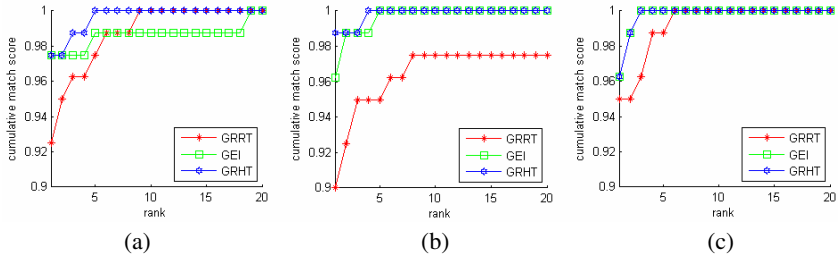
Fig. 7. Cumulative match scores. (a) 0º, (b) 45 º, (c) 90º.

It can be seen that the performance of the proposed approach in our paper is better than GEI and GRRT. The gait Hough template takes more straight lines information which is useful for recognition. The Hough template includes contributions from several gait transformed intensity images which consider the global features in the image space and it is insensitive to image noise.

6   Conclusion

In this paper, we proposed an efficient algorithm for gait recognition based on Hough templates. Experimental results have demonstrated the effectiveness of the proposed algorithm on CASIA-A Gait database. There are two contributions should be highlighted. First, it is the first try to construct the gait templates in the accumulator space of Hough Transform. Second, the algorithm in our experiments can extract the useful straight lines information in the gait sequences. Although we have gain an encouraging results on CASIA-A Gait database, more research on a more realistic gait database is still required before gait recognition can be used for human identification in actual environment.

References

1. Zhang, R., Vogler, C., Metaxas, D.: Human Gait Recognition at Sagittal Plane. Image Vision Computing 25(3), 321–330 (2007)
2. Gu, L.: Research on Human Gait Recognition Methods Based on Image Sequences. Ph.D. thesis, Nanjing University of Science and Technology (2007)

3. Meyer, D., Denzler, J., Niemann, H.: Model Based Extraction of Articulated Objects in Image Sequences for Gait Analysis. In: Proceedings of International Conference on Image Processing, October 1997, vol. (3), pp. 78–81 (1997)
4. Lee, L., Grimson, W.E.L.: Gait Analysis for Recognition and Classification. In: Proc. IEEE Int. Conf. Automatic Face and Gesture Recognition, Washington, DC, pp. 148–155 (2002)
5. Huang, P.S.: Automatic Gait Recognition via Statistical Approaches for Extended Template Features. IEEE Transactions on Systems, Man, and Cybernetics-Part B: Cybernetids 31(5), 818–823 (2001)
6. Collins, R.T., Gross, R., Shi, J.: Silhouette-based Human Identification from Body Shape and Gait. In: Fifth IEEE International Conference on May 21-21, pp. 366–371 (2002)
7. Yu, J., Duan, J., Su, K.: A Hough Transform Based Method for Gait Feature Extraction. Journal of Image and Graphics 10(10), 1304–1309 (2005)
8. Han, J., Bir, B.: Statistical Feature Fusion for Gait-based Human Recognition. In: Proc. of the 2004 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, pp. II842–II847 (2004)
9. Boulgouris, N.V., Plataniotis, K.N., Hatzinakos, D.: An Angular Transform of Gait Sequences for Gait Assisted Recognition. In: Proc. IEEE Int. Conf. Image Processing, Singapore, pp. 857–860 (2004)
10. Sarkar, S., Phillips, P.J., Liu, Z., Vega, I.R., Grother, P., Bowyer, K.W.: The HumanID Gait Challenge Problem: Data Sets, Performance, and Analysis. IEEE Trans. Pattern Analysis and Machine Intelligence 27(2), 162–177 (2005)
11. Boulgouris, N.V., Plataniotis, K.N., Hatzinakos, D.: Gait Recognition Using Dynamic Time Warping. In: Proc. IEEE Int. Symp. Multimedia Signal Processing, Siena, Italy, pp. 263–266 (2004)
12. Boulgouris, N.V., Chi, Z.X.: Gait Recognition Using Radon Transform and Linear Discriminate Analysis. IEEE Transactions on Image Processing 16(3), 731–740 (2007)
13. Wang, L., Tan, T.N., Ning, H.Z., Hu, W.M.: Silhouette Analysis-Based Gait Recognition for Human Identification. IEEE Transaction on Pattern Analysis and Machine Intelligence 25(12), 1505–1517 (2003)
14. Wang, L., Tan, T.N., Hu, W.M., Ning, H.Z.: Automatic Gait Recognition Based on Statistical Shape Analysis. IEEE Transactions on Image Processing 12(9), 1120–1129 (2003)
15. Phillips, P.J., Moon, H., Rizvi, S., Raus, P.: The feret evaluation methodology for face recognition algorithms. IEEE Trans. Pattern Anal. Mach. Intell. 22(10), 1090–1104 (2000)
16. CASIA-A Gait Database [OL/DB], http://www.cbsr.ia.ac.cn/

Improved Competitive Code for Palmprint Recognition Using Simplified Gabor Filter

Jing Wei[1,2], Wei Jia[1], Hong Wang[2], and Dan-Feng Zhu[1,2]

[1] Intelligent Computing Lab, Hefei Institute of Intelligent Machines,
Chinese Academy of Sciences, P.O. Box 1130, Hefei, Anhui, 230031, China
[2] College of Mathematics and Computer Science,
Shanxi Normal University, Linfen, Shanxi, 041004, P.R. China
wjkathy@gmail.com

Abstract. This paper presents a fast algorithm for extracting features using the Simplified Gabor (SG) for Competitive Coding-based palmprint recognition. The competitive code convolves the palmprint image with a bank of Gabor filters with different orientations. We use a simplified version of Gabor filters and an efficient algorithm for extracting features to modify the competitive code. Experimental results indicate that, using SG can achieve the verification accuracy similar to using common Gabor filters, while the runtime for feature extraction using SG is very fast compare to the original algorithm.

Keywords: Competitive code, Simplified Gabor (SG), Feature extraction.

1 Introduction

Biometric is an emerging technology that can effectively solve security problems in our networked society. Generally, the aim of biometrics is to identify or verify individuals using physiological or behavioral characteristics, such as iris, retina, palmprint, fingerprint, hand geometry, face, voice, signature and gesture etc [1]. Particularly, biometrics can provide advantages over non-biometric methods such as password, PIN, and ID cards.

Among aforementioned biometric techniques, palmprint recognition system has several unique advantages. A palmprint image contains various features, including principal lines [2], wrinkles, ridges, minutiae points, singular points and texture. Lines and texture [3] are the most clearly observable features in low-resolution palmprint images (such as 100 dpi). Compared with other biometric traits, the advantages of palmprint are the availability of large palm area for feature extraction, the simplicity of data collection and high user acceptability. Palmprint recognition [4] has received considerable recent research interest due to its low-price capture device, fast execution speed, and high verification accuracy etc.

In palmprint recognition, methods which based on the direction feature [4] are very effective, such as Competitive Code [5], Ordinal Code [6], RLOC [3] etc, since the direction feature of Palmprint lines carry more information to identify, and they are

robust to illumination changes. Competitive Code is the algorithm exploiting Gabor filter bank to extract direction feature. However, using convolution between images and original Gabor filter bank might be time-consuming. In this paper, we use simplified Gabor filter instead of original Gabor filter to speed up the processing of feature extraction.

2 The Original Competitive Code Algorithm

2.1 Competitive Code

The basic idea of Competitive Code [5] is to extract the orientation field as features and use angular distance as a matching function. The original competitive coding scheme simply selects $\theta_p = p\pi/6$, $p = \{0, 1... 5\}$ as orientations of Gabor filters. And we will descript Gabor filter in detail in the next section. A brief summary of Competitive Code is given below:

Six real parts of Gabor filters $\Psi_R(x,y,\theta_j)$ with different directions θ_j are applied to a preprocessed palmprint image $I(x, y)$ [7].

The orientation of a local region is obtained by the competitive rule,

$$k = \arg \quad \min_j (I(x, y) * \psi_R(x, y, \theta_j)) \tag{1}$$

where $j = 0, 1, 2, 3, 4, 5$ and k is called the winning index.

Two Competitive Codes are compared by their angular distance. The bit representation of angular distance is defined as,

$$D(P,Q) = \frac{\sum_{y=1}^{N} \sum_{x=1}^{N} \sum_{i=1}^{3} (P_M(x, y) \bigcap Q_M(x, y)) \bigcap (P_i^b(x, y) \otimes Q_i^b(x, y))}{3 \sum_{y=1}^{N} \sum_{x=1}^{N} P_M(x, y) \bigcap Q_M(x, y)} \tag{2}$$

where $P_i^b(Q_i^b)$ is the ith bit plane of Competitive Code $P(Q)$; $\otimes$ is bitwise exclusive OR; $\bigcap$ is bitwise AND; $P_M(Q_M)$ is the mask of $P(Q)$ to denote the non-plamprint pixels and N^2 is the size of Competitive Code. Fig.1 shows a preprocessed image and the corresponding Competitive Code where different color represents different orientations. We can observe the palm lines in the Competitive Code.

| (a) | (b) | (c) |
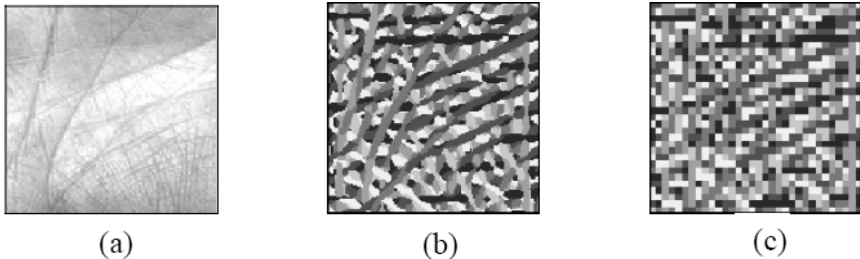
Fig. 1. (a) A preprocessed palmprint image, (b) corresponding Competitive Code (128×128) and (c) corresponding Competitive Code(32×32)

More computational detail, comparison and discussion about this method can be found in [5].

2.2 Gabor Function

The Gabor wavelet [8] is well known for its effectiveness as a feature for image processing and pattern recognition. Its kernels are similar to the response of the two-dimensional receptive field profiles of the mammalian simple cortical cell, and exhibit the desirable characteristics of capturing salient visual properties such as spatial localization, orientation selectivity, and spatial frequency selectivity. In the spatial domain, a Gabor Function is a complex exponential modulated by a Gaussian function, which is defined as follows [9]:

$$G(x, y, \theta, \mu, \sigma) = \frac{1}{2\pi\sigma^2} \exp\left\{-\frac{x^2 + y^2}{2\sigma^2}\right\} \exp\left\{2\pi i(\mu x \cos\theta + \mu y \sin\theta)\right\} \tag{3}$$

where $i = \sqrt{-1}$, μ is the frequency of the sinusoidal wave, θ controls the orientation of the function, and σ is the standard deviation of the Gaussian envelop.

By selecting different center frequencies (also called scales) and orientations, we can obtain a family of Gabor filters. Fig.2 illustrates the real part and the imaginary part of a Gabor filter.

(a)   (b)
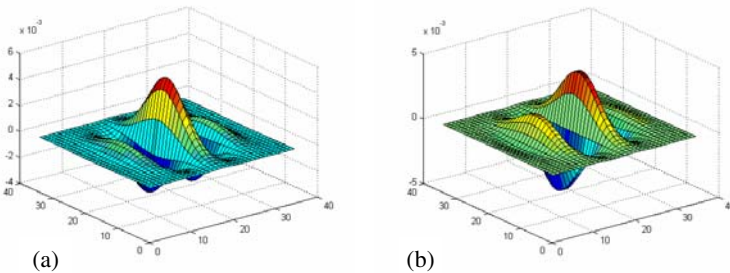
Fig. 2. (a) The real part and (b) the imaginary part of a Gabor filter

3 Simplified Gabor Based the Competitive Code

Although fast algorithms such as the Fast Fourier Transform (FFT) can be employed for convolution, the computational complexity is still very high because a large number of Gabor filters are involved in computing the Competitive Code. In [10], simplified Gabor wavelets (SGW) have been proposed for face recognition. This simplified Gabor filter (SG) can be viewed as an approximation of the original Gabor wavelets (GW). An SG is generated by quantizing a corresponding GW into a certain number of levels.

With SG, features can be extracted efficiently and fast. The SG can replace the GW for real-time processing and applications.

In [10], the positive and negative parts of a GW are quantized uniformly according to the corresponding number of levels. We assume the most positive and negative values of a GW are $A+$ and $A-$, respectively. The corresponding quantization levels for positive levels $q+(k)$ and negative levels $q-(k)$ are as follows:

$$q+(k) = \frac{A+}{2n_p+1} \cdot 2k \quad where \quad k=1,...,n_p$$

$$(4)$$

$$and \quad q-(k) = \frac{A-}{2n_n+1} \cdot 2k \quad where \quad k=1,...,n_n$$

The real part and the imaginary part of the corresponding simplified Gabor wavelet are shown in Fig. 3.

In order to further simplify the Gabor filter, we assume the positive and negative values of a SG are only $a+$ and $a-$, and $a+=1$, $a-=-1$, where the SG is must zero mean. In Competitive Code, we only use the real part of the SG as shown in Fig.4a.

(a)   (b)
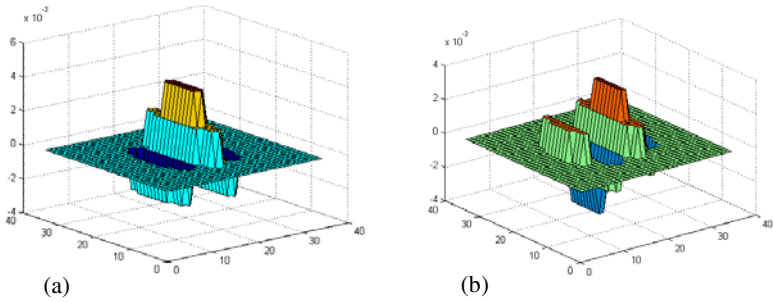
Fig. 3. (a) The real part and (b) the imaginary part of the corresponding SGW

(a)   (b)   (c)
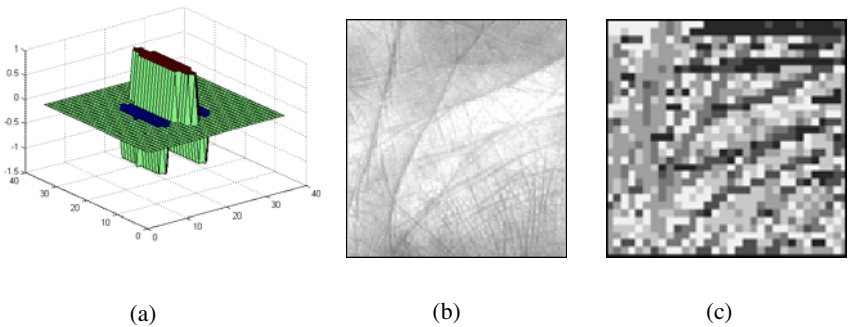
Fig. 4. (a) The real part of the SG, (b) a preprocessed palmprint image and (c) corresponding Competitive Code (32 x 32) using the real part of the SG

In order to save time and improve speed, the features are extracted not by convolving $I(x,y)$ with $\Psi_R(x,y,\theta_j)$. Convolve is defined as the integral of the product of the two functions after one is reversed and shifted. Because the values of the real part of the SG have three levers, 1,-1 and 0, we only need to add and subtract the value of the corresponding position in the original image. In the same words, firstly, we add the positive part. Then we add the negative part. Finally we let the positive part add the negative part to get the final result. The result of this method is equal to the result of convolution. Fig.4 (b),(c) show a preprocessed image and the corresponding Competitive Code using SG. In this way, the runtime for feature extraction will be improved greatly.

4   Experiment

4.1   Palmprint Database

PolyU Palmprint Database [11] is used for the evaluation of the recognition perform-ance. In the database there are totally 600 palmprint images of 100 different palms (100 classes). Six samples from each of these palms were collected in two sessions. The average interval between the first and the second collection was two months. In our paper, the 3 samples were captured in the first session are selected as training samples and the 3 samples were captured in the second session are selected as test samples. The experiments were conducted on a personal computer with an Intel Duo E4500 processor (2.20GHz) and 2.0G RAM configured with Microsoft Xp and Matlab 7.4.

4.2   Experimental Results

All images are normalized to 128×128 subimage after preprocessing. Then four meth-ods, i.e., palm code [7], fusion code [12], competitive code [5], and competitive code using SG are compared. After all possible intra-class and inter-class comparisons are implemented in PolyU Palmprint Database, the comparison of the four methods using Gabor filters are illustrated in Table 1. In our experiments, we also measure the time to extract feature by above algorithms and the results are displayed in Table 1. Fig.5 shows FAR and FRR curves by original Competitive Code and Competitive Code using SG.
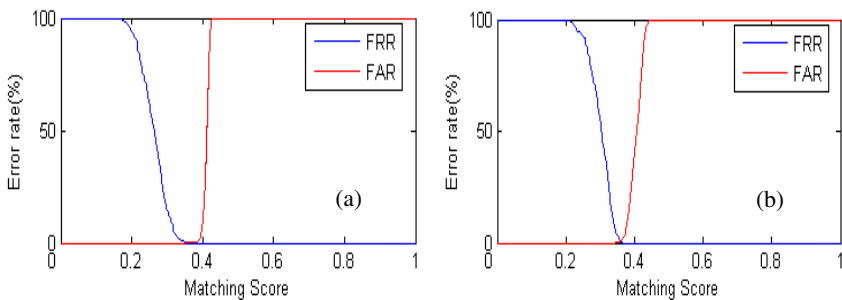
Fig. 5. FAR and FRR curves (a) by Competitive Code and (b) by Competitive Code using SG

Table 1. Comparison of some algorithms based-Gabor on PolyU database

Algorithm	Feature size	Feature extraction time	EER	Recognition rate
Palm Code ($\theta=0^o$) [7]	256 Bytes	51ms	1.13%	98.67%
Palm Code ($\theta=45^o$) [7]	256 Bytes	51ms	0.12%	100%
Palm Code ($\theta=90^o$) [7]	256 Bytes	51ms	0.52%	100%
Palm Code ($\theta=135^o$) [7]	256 Bytes	51ms	0.13%	99.67%
Fusion Code [12]	256 Bytes	279ms	0.11%	100%
Competitive Code [9]	384 Bytes	226ms	0	100%
Comp using SG by convolving	384 Bytes	187ms	0.33%	100%
Comp using SG not by convolving (our method)	384 Bytes	145ms	0.33%	100%

From the experimental results, we can see that using SG can achieve the verification accuracy similar to original competitive code, while the runtime for feature extraction using SG is very fast compared to the original algorithm.

5   Conclusions

In this paper, we used simplified Gabor for competitive code, which can achieve a performance similar to the original competitive code for palmprint recognition. This method is robust to illumination, contrast and misalignment variations [3]. In addition, the speed of feature extraction using the algorithm by simplified Gabor is faster than the Competitive Code algorithm. Although the EER of the improved method is slightly higher, the recognition rate is similar to the Competitive Code while the time of extraction is greatly reduced. This method of the simplified filter can also be used in ordinal code [6] in the same way. The time of extraction feature of the improved ordinal code can be decreased greatly too. We will try our best to reduce the ERR of the improved method while keeping the time of extraction unchanged or reduced in the future.

Acknowledgement

References

1. Jain, A., Bolle, R., Pankanti, S.: Personal Identification in Networked Society, Boston, Mass. Kluwer Academic Publishers, Dordrecht (1999)
2. Huang, D., Jia, W., Zhang, D.: Palmprint Verification Based on Principal Lines. Pattern Recognition 41(4), 1316–1328 (2008)

3. Jia, W., Huang, D., Zhang, D.: Palmprint Verification Based on Robust Line Orientation Code. Pattern Recognition 41(5), 1504–1513 (2008)
4. Jia, W., Huang, D., et al.: Palmprint Identification Based on Directional Representation. In: 2008 IEEE International Conference on SMC, Singapore, pp. 1562–1567 (2008)
5. Kong, A.W.K., Zhang, D.: Competitive Coding Scheme for Palmprint Verification. J. Proc. ICPR (2004)
6. Sun, Z.N., Tan, T.N., Wang, Y.H., Li, S.Z.: Ordinal Palmprint Represention for Personal Identification. In: Proc. of IEEE Computer Sodety Conference on Computer Vision and Pattern Recognition, vol. 1, pp. 279–284 (2005)
7. Zhang, D., Kong, W.K., You, J., Wong, M.: On-line palmprint identification. IEEE Trans. on PAMI 25(9), 1041–1050 (2003)
8. Lee, T.S.: Image Representation Using 2D Gabor Wavelets. IEEE Trans. Pattern Anal. Mach. Intell. 18(10), 959–971 (1996)
9. Liu, D., Lam, K.M., Shen, L.S.: Optimal sampling of Gabor features for face recognition. J. Pattern Recognition Lett. 25(2), 267–276 (2004)
10. Choi, W.P., Tse, S.H., Wong, K.W., Lam, K.M.: Simplified Gabor wavelets for human face recognition. J. Pattern Recognition 41, 1186–1199 (2008)
11. PolyU Palmprint Database, http://www.comp.polyu.edu.hk/~biometrics/
12. Kong, W.K., Zhang, D.: Feature-level fusion for effective palmprint authentication. In: Zhang, D., Jain, A.K. (eds.) ICBA 2004. LNCS, vol. 3072, pp. 761–767. Springer, Heidelberg (2004)

Palmprint Recognition Combining LDA and the Center Band of Fourier Magnitude

Dan-Feng Zhu[1,2], Hong Wang[2], Yi-Hai Zhu[1], and Jing Wei[1,2]

[1] Intelligent Computing Lab, Hefei Institute of Intelligent Machines,
Chinese Academy of Sciences, P.O. Box 1130, Hefei, Anhui, 230031, China
[2] College of Mathematics and Computer Science,
Shanxi Normal University, Linfen, Shanxi 041004, P.R. China
danfeng0710@gmail.com

Abstract. In this paper, an effective algorithm has been proposed for palmprint recognition combining Fourier Transform and Linear Discriminant Analysis (LDA). For Fourier representation, we only exploit the center band of Fourier magnitude for recognition. The results of experiments conducted on PolyU palmprint database demonstrate the effectiveness of proposed method. All in all, the proposed method that is robust against illumination in this paper is a suitable even wonderful method for palmprint recognition.

Keywords: Biometrics, Palmprint recognition, LDA, Fourier Transform, The center band.

1   Introduction

In the information society, there are occasions on which the personal authentication is required. There is no doubt that biometrics is one of the most important and effective solutions to this task. Generally speaking, biometrics is a field of technology that uses automated methods for identifying or verifying a person based on physiological or behavioral traits such as face, fingerprints, iris and palmprint etc. Due to its importance and promising market prospect, biometrics has drawn wide attention from researchers [1].

Palmprint recognition, developed in recent years, is a new biometrics technology. Compared with fingerprint or iris based personal biometrics system, palmprint based biometrics system has several unique advantages such as stable line features, low-resolution imaging, low cost capture devices, easy self positioning etc. [2]. What is more important is that it can reach accurate recognition rate with processing speed [3]. However, palmprint recognition is still staying at the developing stage, therefore the investigation of its theory and application should be enhanced and improved further.

It is well-known that subspace methods are often used for palmprint recognition. For example, Lu et al. adopted Principal Components Analysis (PCA) [4], and Wu et al. used LDA for palmprint recognition [5]. On the other hand, some researches show that by using Fourier representation, the recognition performance of subspace methods can be improved. For instance, Savvides et al. exploited the phase components based on PCA for face recognition, and better recognition rates were reported [6]. Furthermore,

some recent researches show that using the center band of Fourier spectrum, we can also improve the recognition performance since the high frequency of Fourier Transform might be meaningless for recognition. In [7], ITO et al. exploited Phase-Only Correlation (POC) using the center band of Fourier phase, and the improvement of recognition is obvious.

In this paper, an effective algorithm has been proposed for palmprint recognition combining Fourier Transform and Linear LDA, and we also investigate the recognition performance of using the center band of Fourier magnitude.

2 Palmprint Recognition Combining LDA and the Center Band of Fourier Magnitude

2.1 Fourier Transform

2D Discrete Fourier Transform (2DDFT) has been widely discussed and applied in image processing [8]. Assume f $(m,$ $n)$ $(m=0,1,\cdots, M-1; n=0,1,\cdots, N-1)$ is a $M \times N$ palmprint image. Let $F(k, l)$ denote the 2DDFT of the images $f(m, n)$, which is given by:

$$F(k,l) = \sum_{m=0}^{M-1} \sum_{n=0}^{N-1} f(m,n)e^{-j2\pi(\frac{mk}{M}+\frac{nl}{N})}$$
$$= A_F(k,l)e^{j\theta_F(k,l)},$$

(1)

where $A_F(k, l)$ is magnitude components, and $\theta_F(k, l)$ is phase components. $F(k, l)$ is returned a matrix which size is $M \times N$, and the size of $A_F(k, l)$ is also $M \times N$. Fig. 1 shows respectively the magnitude and phase components of two palmprint images.
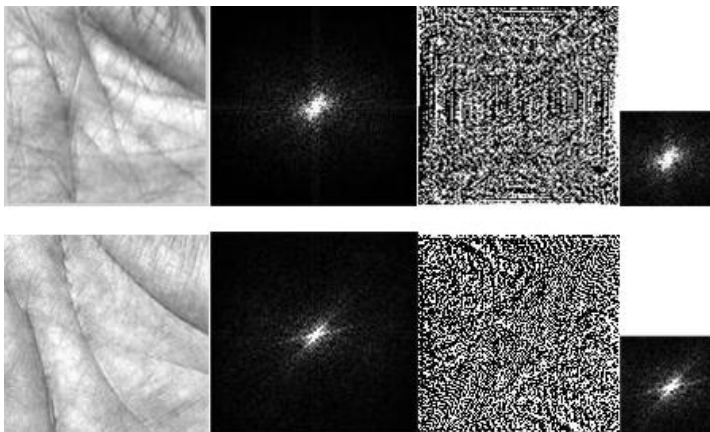
Fig. 1. Examples of the palmprint images used in the experiment. The original palmprint image(the first column), Fourier magnitude components(the second column), Fourier phase components(the third column) and the center band of Fourier magnitude(the fourth column).

2.2   The Center Band of Fourier Magnitude

ITO et al. found that using all the phase information cannot get the best recognition rate, because the phase information in the high frequency region might be meaningless for recognition [7]. In fact, the magnitude information is also in line with the law.

Our approach here is that the center band of Fourier magnitude is extracted after centering the Fourier magnitude, as shown in Fig. 1. Generally, using the center band of Fourier magnitude, we can eliminate meaningless high frequency magnitude components even noises of illumination, because the frequency spectrum of noise often lies in high frequency components. Therefore, using the center band of Fourier magnitude, it has been found that the proposed method can be tolerant of illumination variations and reduce required space. As a result, the recognition performance can be improved, and the response time can be reduced. Now we give an example to demonstrate its role. The size of a palmprint image is 128×128, so the size of Fourier magnitude is also 128×128. If we set the center bandwidth t to 64, the size of the center band of Fourier magnitude is 64×64 and a quarter of the original Fourier magnitude. It is illustrated in Fig.1, in which the second column show the original Fourier magnitude and the fourth column show the center band of Fourier magnitude. As can be seen from the above analysis, it is evident that performing the center band of Fourier magnitude plays a key role.

2.3   LDA

LDA is a well-known supervised method for dimensionality reduction, but before that, the center band of Fourier magnitude needs to be adjusted.

A $k×k$ the center band of Fourier magnitude can be reshaped to a k^2 dimension array with the same elements as the center band of Fourier magnitude. The space spanned by the array is called the Fourier magnitude space (FMS). Generally, the dimensionality of FMS is too high to be used directly. Thus, we perform LDA to reduce the dimensionality of the Fourier magnitude array of FMS. When dealing with high-dimensional and small sample size data, classical LDA suffers from the singularity problem. A common way to deal with the singularity problem is to apply an intermediate dimensionality reduction stage, such as PCA [9]. The algorithm is known as PCA+LDA. So we apply PCA+LDA to the Fourier magnitude array here. The Fourier magnitude array is firstly projected to a lower intermediate dimensional space by using PCA, which stems from the fact that the intermediate dimensionality is at most n-c, then the standard LDA is employed to process the projected samples and reduce the dimension to at most c-1.

Given a data matrix $X=\{x_1, x_2, \ldots, x_n\} \in R^{d×n}$ consisting of n samples $\{x_i\}_{i=1}^{n}$ in R^d, compute the transformation matrix W_{PCA} :

$$W_{PCA} = \arg \max_{W} \left| W^T S_T W \right| = [u_1, u_2, \cdots u_{m_1}] \tag{2}$$

Where $\{u_i | i=1,2,\ldots,m_1\}(m_1 \leq n\text{-}c)$ is the set of eigenvector of S_T corresponding to the m_1 nonzero largest eigenvalues, and m_1 is the intermediate dimensionality reduction by using PCA. S_T is the total scatter matrix defined as

$$S_T = \sum_{k=1}^{n} (x_k - \mu)(x_k - \mu)^T \tag{3}$$

Compute the transformed within-class scatter matrix S'_W, which is a full rank matrix and the transformed between-class scatter matrix S'_B:

$$S'_w = W_{PCA}{}^T S_w W_{PCA} \tag{4}$$

$$S'_B = W_{PCA}{}^T S_B W_{PCA} \tag{5}$$

W_{LDA} can be obtained as follows:

$$W_{LDA} = \arg\max_w \frac{\left|W^T S'_B W\right|}{\left|W^T S'_w W\right|} \tag{6}$$

Thus, the W_{opt} can be computed by the following formula:

$$W_{opt}^T = W_{LDA}^T W_{PCA}^T = [w_1, w_2, \cdots, w_m]^T \tag{7}$$

where the columns of $W_{opt}\{w_1, w_2, ..., w_m\}(m \le c\text{-}1)$ are orthonormal vectors, c is the number of palmprint classes. The space spanned by these vectors is called the Fisher-palm space(FPS). Note that there are at most c-1 nonzero generalized eigenvalues, and an upper bound on m is c-1, where c is the number of palmprint classes.

There are two stages in our system: the enrollment stage and the recognition stage. In the enrollment stage, the W_{opt} is computed by using the training samples (Eqs. (2)–(7)) and stored as a FPS at first, and then the mean of each palmprint class is projected onto this FPS:

$$Y = W_{opt}^T X, \tag{8}$$

where $X = \{x_1, x_2, \cdots, x_c\}$, c is the number of palmprint classes and each column of X, $x_i(i=1, 2, \cdots, c)$ is the mean of the ith class palmprints. Y is stored as the template for each palmprint class. In the recognition stage, the input the Fourier magnitude array is projected onto the stored FPS to get its feature vector V, and then V is compared with the stored templates to obtain the recognition result [5]. After using LDA perform feature extraction, a nearest neighbor classifier based on Euclidean distance is employed for classification.

3   Experiments and Discussions

Evaluating the performance of the proposed method, we conduct the experiments on the PolyU palmprint database including 600 gray-scale palmprint images from 100 palms [14]. The database was collected in two sessions, and the average interval between the first and the second collection was two months. After preprocessing, the palmprint image is normalized to 128×128 ROI image [5]. That is, the size of all the original palmprint images is 128×128 pixels. Some samples in our database are shown in Fig.2.

In these experiments, the first three samples captured in the enrollment stage are selected as training samples and the three samples captured in the recognition stage are selected as test samples. All of the experiments are conducted on a PC with Microsoft XP and MATLAB 7.5.0 platform. After using LDA perform feature extraction, the nearest neighbor classifier is employed for classification. Then, the output for a query

may be correct or not and we count the number of correct answers to evaluate the accuracy of this method. Also the time for response is recorded to evaluate the efficiency of the method. To sum up, all experiments are carried out in order to measure the performance of the proposed method.

Fig.3 shows the situation of the recognition rate versus the variation of dimension. In these experiments, there are five kinds of methods to compare altogether. They are Eigenpalms(PCA), Eigenphases(Phase+PCA), Fisherpalms(LDA), CBFP+LDA and CBFM+LDA(the proposed method), and the algorithms in brackets are main algorithms of these methods. The CBFP+LDA represents temporarily the algorithm combining the center band of Fourier phase and LDA, and the CBFM+LDA represents the proposed method for palmprint recognition combining the center band of Fourier magnitude and LDA.
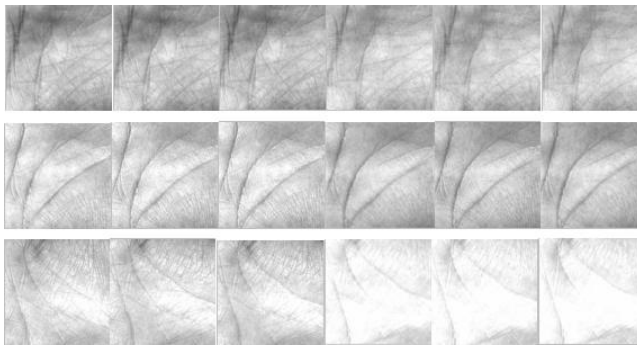
Fig. 2. Some typical samples of the cropped images found in the PolyU Palmprint Database
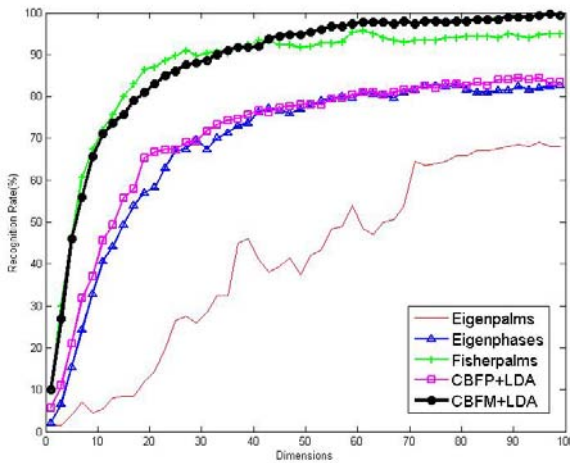
Fig. 3. The recognition rates of Eigenpalms, Eigenphases, Fisherpalms, CBFP+LDA and CBFM+LDA (the proposed method) versus the dimensions

As can be seen from the Fig.3, the recognition rates of the five methods all revealed a trend of fast increase though the margins of the rises varied. It is evident that the feature length plays a key role in the matching process, and long feature lengths lead to a high recognition rate. Specifically, in this experiment, the Eigenphases, the CBFP+LDA and the CBFM+LDA have a common parameter, the center bandwidth t. We set the center bandwidth t of the Eigenphases and the CBFP+LDA to 20 (about 16% of the entire Fourier phase), and the center bandwidth t of the CBFM+LDA to 38 (roughly equivalent to 30% of the entire Fourier magnitude). The black curve with circle represents the CBFM+LDA that is the proposed method for palmprint recognition, and its recognition performance increases rapidly with the dimension and achieved the maximum 99.7 percent while the dimension is 98.
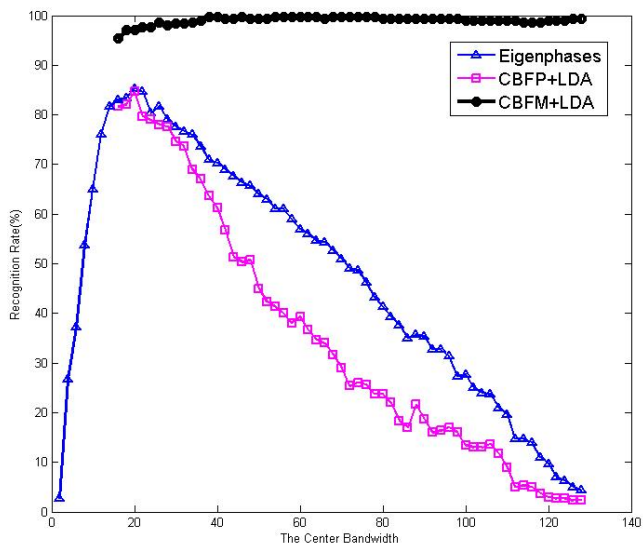
Fig. 4. The recognition rates of Eigenphases, CBFP+LDA and CBFM+LDA (the proposed method) versus the variation of the center bandwidth t

With the variation of the center bandwidth t, the recognition performance situation of Eigenphases, CBFP+LDA and CBFM+LDA (the proposed method) are shown in Fig.4. This experiment show the recognition performance of above three methods versus the center bandwidth when a certain fixed dimension. Generally, the CBFM+LDA has two parameters in this experiment, the intermediate dimension m_1 (using PCA) is set to 158 and the ultimate dimension m (using LDA) is set to 98. For Eigenphases and CBFP+LDA, the recognition rates gradual increase and reach the maximum, while the center bandwidth t is set to 20 (about 16% of the entire Fourier phase), and then decline rapidly. As is indicated in the figure, however, the CBFM+LDA is still represented by the black curve with circle. Through experiments, we can find that the recognition rate of the CBFM+LDA cannot be obtained when the center bandwidth is less than 15. Then the larger is the center bandwidth, the higher is

the recognition rate of the CBFM+LDA and reached a stable situation since 38 (roughly equivalent to 30% of the entire Fourier magnitude). In short, it is evident that the center bandwidth t has a great impact on the recognition performance.

Table 1 lists the maximal recognition performance of each method and corresponding response time. From Table 1, we can see comparative results which the CBFM+LDA (the proposed method) outperforms Eigenpalms, Eigenphases, Fisherpalms and the CBFP+LDA. Specifically, when the center bandwidth t is set to 38 (roughly equivalent to 30% of the entire Fourier magnitude), the intermediate dimensionality m_1 (using PCA) is set to 158 and the ultimate dimensionality m (using LDA) is set to 98, the performance of the CBFM+LDA were shown in the Table 1. The identification accuracy of the CBFM+LDA is up to 99.7 percent, i.e., only one sample is missed. In addition, the average response time is 9.60 seconds. Using the center band of Fourier magnitude, we found a great reduction in response time. Therefore, it shows that the proposed palmprint identification method is with high performance in terms of accuracy and efficiency on our palmprint database. All these results demonstrate intuitively that the CBFM+LDA (the proposed algorithm) is also a good tool for palmprint recognition.

Table 1. The Table Show the Maximal Recognition Rates, the Corresponding Response Time on the PolyU Palmprint Database and the Corresponding Center Bandwidth

Methods	Recognition Rate	Response Time
Eigenpalms	74.7%	9.95s
Eigenphases (t=20)	85.3%	6.65s
Fisherpalms	95.7%	9.70s
CBFP+LDA (t=20)	84.7%	6.75s
CBFM+LDA (the Proposed Method) (t=128)	99.3%	34.50s
CBFM+LDA (the Proposed Method) (t=38)	99.7%	9.60s

4   Conclusion

In this paper, we proposed a method for palmprint recognition combining the center band of Fourier Transform and Linear Discriminant Analysis. The experimental results show that the proposed method consistently outperforms Eigenpalms, Eigenphases and Fisherpalms. It is clear that the advantages of our proposed method are that it does not require any special preprocessing or any a-priori knowledge to handle illumination variations during the testing phase. For Fourier representation, we only exploit the center band of Fourier magnitude for recognition, which can improve the recognition

accuracy and response time considerably. All in all, the new method that we mentioned in this paper is a suitable even wonderful method for palmprint recognition.

Acknowledgment

References

1. Huang, D., Jia, W., Zhang, D.: Palmprint Verification Based on Principal Lines. Pattern Recognition 41(4), 1316–1328 (2008)
2. Jia, W., Huang, D., Zhang, D.: Palmprint Verification Based on Robust Line Orientation Code. Pattern Recognition 41(5), 1504–1513 (2008)
3. Jia, W., Huang, D., et al.: Palmprint Identification Based on Directional Representation. In: 2008 IEEE International Conference on SMC, Singapore, pp. 1562–1567 (2008)
4. Lu, G.M., Zhang, D., Wang, K.Q.: Palmprint Recognition Using Eigenpalms Features. J. Pattern Recognition Letter. 24, 1463–1467 (2003)
5. Wu, X.Q., Zhang, D., Wang, K.Q.: Fisherpalms Based Palmprint Recognition. J. Pattern Recognition Letter. 24, 2829–2838 (2003)
6. Savvides, M., Vijaya Kumar, B.V.K., Khosla, P.K.: Eigenphases vs. Eigenfaces. In: IEEE International Conference on Pattern Recognition, pp. 1051–4651 (2004)
7. Ito, K., Aoki, T., Nakajima, H.: A Palmprint Recognition Algorithm Using Phase-Only Correlation. IEICE Trans. Fundamentals E91–A, 4 (2008)
8. Li, W.X., Zhang, D., Xu, Z.Q.: Palmprint Recognition Based on Fourier Transform. J. Journal of Software, 0879–0886 (2002)
9. Ji, S.W., Ye, J.P.: Generalized Linear Discriminant Analysis: A Unified Framework and Efficient Model Selection. IEEE Trans. on Neural Networks 19(10), 1768–1782 (2008)
10. Belhumeur, P., Hespanha, J., Kriegman, D.: Eigenfaces vs Fisherfaces: Recognition Using Class Specific Linear Projection. IEEE Trans. 19(7), 711–720 (1997)
11. Chen, L.F., Liao, H.Y.M., Lin, J.C., Kao, M.D., Yu, G.J.: A New LDA-Based Face Recognition System which Can Solve the Small Sample Size Problem. Pattern Recognition 33(10), 1713–1726 (2000)
12. Oppenheim, A.V., Lim, J.S.: The Importance of Phase in Signals. Proc. IEEE 69(5), 529–541 (1981)
13. Hayes, M.H., Lim, J.S., Oppenheim, A.V.: Signal Reconstruction from Phase or Magnitude. IEEE Trans. Acoust., Signal Processing ASSP-28, 672–680 (1980)
14. PolyU Palmprint Database, http://www.comp.polyu.edu.hk/~biometrics/

A Framework for Recognition Books on Bookshelves

Nguyen-Huu Quoc and Won-Ho Choi

Graduate School of Electrical Engineering, University of Ulsan, Korea
Tel.: +82-52-259-1627; Tel.: +82-52-259-2203
robetaynguyen@yahoo.com, whchoi@mail.ulsan.ac.kr

Abstract. In this paper, we present a framework to recognize books on bookshelves by reading its title on book spines. The framework consists of control and recognition module. Control module moves camera to suitable positions for image capturing while recognition one processes taken images to know which books are shelved on the shelves. Firstly, images are captured from random position. Secondly, we separate it into book and non-book regions. Then, books in book region are segmented by using line segment and MSAC based dominant vanishing point (DVP). After book verification stage, adaptive Otsu's thresholding is employed to extract book titles and ready for recognition of next stage. In case recognizing unsuccessfully, we feedback control information to control module to adjust camera location and repeat the above procedure.

Keywords: Book recognition, line segment, dominant vanishing point, MSAC, Singular Value Decomposition, OCR, RANSAC.

1 Introduction

Nowadays, regardless to extremely develop of many kinds of media, books are indispensable information source to life. With the huge number of books keep growing everyday, some book related works where formerly done by human beings such as arranging, finding, storing books, etc., are getting more and more exhausting. Intelligent robots can be a solution of above issues, in which, book recognition is the core module of the robots. In order to successfully implement this kind of module onto robots, we need to layout a framework for long term research.

This paper considers a comprehensive view of the system, addresses every aspects and challenges of the problems and provides a clear and effective framework for recognition of books on bookshelves. The framework should be inheritable and consistent for future development. Consequently, we design a pattern of framework which covers all essential steps for recognition of books on bookshelves including: extracting regions of book on bookshelves, segment books from book regions, extract book titles, recognize characters of title and syntactic verification. With feedback information from the verification block, a control

module can adjust camera position to take photos from better view in case the recognition module cannot find a reasonable solution. Therefore, our approach can recognize the books when boundaries among books are unclear or characters are almost invisible. Currently, we concentrate on solving the core problems of segmentation including books and characters segmentation as well as propose the algorithm for extracting characters in book titles.

In most of earlier related works, authors focused on a specific part of the system and many assumptions were made to boost the final outcomes. In [1], the authors proposed a method for characters recognition of books on bookshelves using multiple dictionaries selection. Characters of book titles were easily extracted from books by close shot, clear image and uniform background on book spines. Meanwhile, [2], [3] and [10] concentrated on detecting boundary of books. Bookshelves images were taken in the fixed distance and intentionally captured the region of books with single cell per image. Their approach could be fail in case the non-book regions or multiple cells exist in one image. The implementation of book recognition was presented in [5], UJI Librarian robot. The control and vision strategy were both mentioned in the papers. However, the implementation was limited in range on application because of recognizing books based on book tags. Robot cannot handle the works where book tags don't lie in the right position caused by collapsing books or confused arrangement. In addition, distance between camera and robots must be kept unchanged during operation time. This can reduce robot's visibility, flexibility and increase searching time.

2 The Framework for Recognition Books on Bookshelves

2.1 Summarization of Problems of Recognition Books on Bookshelves

In general, we can classify the problems need to be solved of the system into 3 categories: bookshelf structures, book features and image quality. In the first group, we have to deal with various kinds of bookshelves corresponding to their location such as libraries, bookstores, offices, etc. The differences among their shapes, background and materials can take a large effect on recognition process.
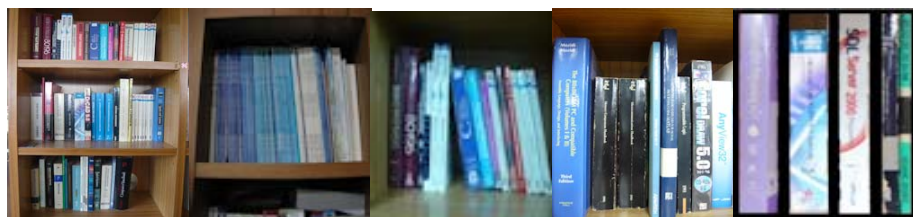
(a) Multiple cells (b) Thin books (c) Non-parallel (d) Refection (e) Non-uniform

Fig. 1. Typical problems of the system

Moreover, arranging method also raises difficulties when books are not parallel, collapsed or mixing non-book objects into book regions.

The second class relates to the patterns and size of books. Many algorithms could fail in book segmentation stage in case books are thin, dissimilar height while non-uniform spine background, uncommon fonts, color and character's size can cause false alarms when recognizing characters.

The last category addresses some problems of image quality due to light source, camera types or vibration when taking pictures. As the result, the images can be blurring, glaring colors. It leads to missing information or wrong detection. Fig.1 shows the typical challenges of designing system.

2.2   Books on Bookshelves Recognition Framework

The main idea of this paper is to design a system that can combine various information including book profiles, recognized characters of title strings and syntactic verification in order to minimize the risk of false detection. Our system also emphasizes the role of control module in case recognition module can not find a reasonable solution. The general block diagram for the system is depicted in fig. 2a and fig. 2b is our approaching methods for each steps.

In recognition module, book regions extraction is done by frequency difference between book and non-book regions. Line segmentation, MSAC algorithm,
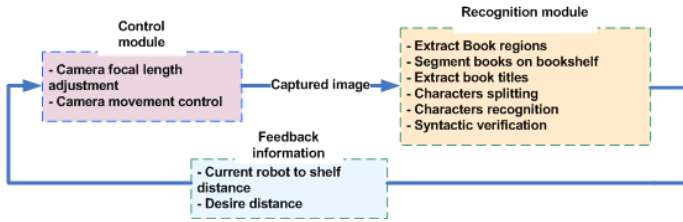
(a) General view
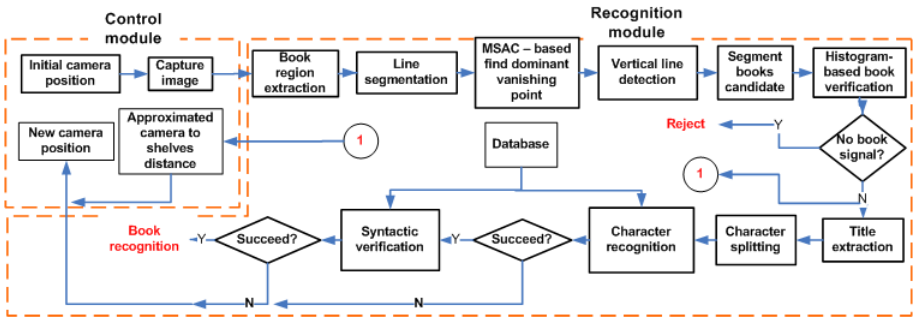
(b) Approaching methods

Fig. 2. The framework design

vertical line detection and histogram based verification are implemented methods for book segmentation. Every single obtained character in title extracting and characters splitting process is recognized by recognition stage. Finally, we use syntactic verification to determine the book title and feedback control information.

3   Technical Approach

3.1   Book Region Extraction Using High Frequency Filtering

The goals of this step are to extract regions of book out of the other ones and divide multiple cells image into sub-images where only one cell exists in each sub-image (shown in fig. 3). The method is based on the fact that regions of books take higher frequency than remaining regions. For the reason, high pass filter is applied to filter out low frequency areas correspond to regions of non-book. Firstly, converted gray image is preprocessed by rescaling and Gaussian filtering for noise suppression. then,3^{rd}order high pass Butterworth filter is employed with details describe in (1, 2) Where $g(x,y)$ is output image , $I(u,v)$ is Fourier transform of original image with zero-frequency component is shifted to center of spectrum. $F^{-1}(.)$ is inverse Fourier transform operator, (cx,cy) is center of image and fc, n are cut frequency and filter order respectively.

$$g(x,y) = F^{-1}\left(H(u,v)I(u,v)\right) \tag{1}$$

$$H(u,v) = \frac{1}{1+\left(\frac{fc^2}{(u-cx)^2+(v-cy)^2}\right)^{2n}} \tag{2}$$

Since the filter is to reject low frequency areas, the remaining high frequency regions mostly cover regions of books and small noise areas. These areas appear due to the uneven illumination or color difference of bookshelf's material. They can be eliminated using opening by reconstruction [7] in combination with connected component analysis (CCA) to get the book regions.

(a) Original image (b) filtered image   (c) result
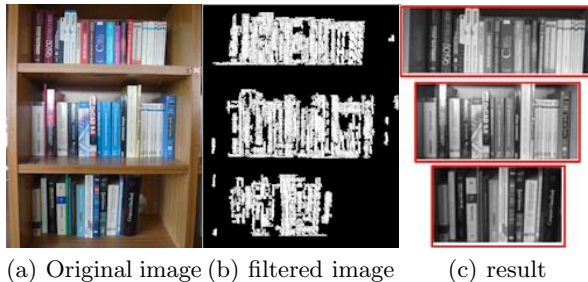
Fig. 3. Frequency difference in bookshelf image

3.2   Books Segmentation and Verification

This step separates every single book from extracted book regions and checking the valid books after segmented:

Line segment: Vertical lines are the most important feature to distinguish between 2 adjacent books. The first stage is computation of edges using Canny edge detector. Then we link edge's pixels together into lists of sequential edge points, one list for one edge contour. The contour (edge list) starts and stops at an ending or a junction of another contour (edge list). Simultaneously, we discard contours less than 50 pixels. Line segments are fitted with the maximum deviation from the original edge of 2 pixels and coarsely separated into two groups. The vertical one contains line segments which create an acute angle 20° in maximum with vertical axis.

MSAC based calculation of vertical dominant vanishing point (DVP): We search in bookshelves image to find the vertical DVP corresponding to vertical lines pass through. To estimate DVP, we use MSAC (m-estimator sample consensus, [8]) algorithm. Suppose a line segment in image plane is described as $l = (a, b, c)^T$ where $a^2 + b^2 = 1$ (normalization); With 2 segments l_i, l_j, intersection of l_i, l_j is computed as: $v = l_i \times l_j$, in which $v = (v1, v2, v3)^T$. So v is satisfied $l_k^T v = 0; k = i, j$ .In reality, an error always occurs when detecting line segments. Thus, the intersection v of n lines through v is described,

$$e = \sum_{k=1}^{n} e_k = \sum_{k=1}^{n} \left(l_k^T v \right)^2 \simeq 0 \tag{3}$$

We run RANSAC [4] for K trials to find the largest consensus set N_i contains N_i segments. The best choice of intersection v among those segments is re-estimated using all the points of the subset N_i. The number K is computed:

$$K = \left(\frac{\log(1-p)}{\log(1-(1-\varepsilon)^n)} \right) \tag{4}$$

Where, p is probability that after K trials, at least one of the random sample of n segments is free from outliers; ε is probability that a selected data point is an outlier. In our implementation, the cost function is calculated as $C = \sum_k \rho \left(e_k^2 \right)$.

Where, $\rho(.)$ is the robust error term is defined in MSAC algorithm:

$$\rho \left(e^2 \right) = \begin{cases} e^2 (inliers) \\ t^2 (outliers) \end{cases} \tag{5}$$

In Equ.6, $t = 1.96\sigma$; and σ is computed by the suggestion of Rousseeuw et al [9]

$$\sigma = 1.4826 \left[1 + \frac{5}{(N-n)} \right] \sqrt{med_i \left(e_i^2 \right)}; n = 2 \tag{6}$$

Outputs of the MSAC algorithm are a set of inliers V^{in} and vertical vanishing point v. Fig. 4c show the vertical group V^{in} correspond to the vertical vanishing point v marked by yellow color.
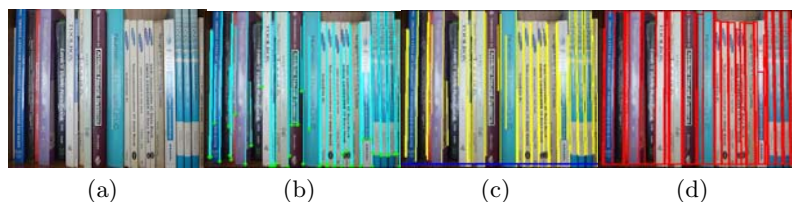
Fig. 4. Book segmentation process: (a) original image; (b) line segment; (c) MSAC find DVP; (d) book candidates

Segment book candidates: After a set of vertical lines was selected, lines with length less than L_T are rejected. Each end point of the remains is grouped together to find the best fit bottom boundary of books by using least square error estimation. To find the upper boundary of book, we consider pair of starting points of two adjacent lines; the upper boundary of a book is the horizontal line through the higher starting point. With 2 vertical lines, an upper and a bottom line, we now found out 4 vertexes of a book and easily segment book candidates and skew to get their straight view.

Histogram based books verification: In previous stage, we assume 2 consecutive vertical lines form a book. This assumption is not always true. Books are distinguished to non-books by non-uniform of their histogram. As the result, variance of histogram of a non-book candidate is extremely lower than that of book candidate. We exploit such a feature to decide whether a candidate is a book.

3.3 Title Extraction and Characters Splitting

The objects of this step are to localize titles on the segmented books and split strings contain the titles into single character.

Title extraction. Book titles are located through 4 sub-steps:

Heuristic Filtering: Canny edge map obtained previously is normalized to standard size and named as ME. ME is then clustered into number of blobs .i.e. connected regions. We specify some heuristic criteria including centroid position, aspect ratio, width, height and area of bounding box to discard some regions not likely belong to region of text.

Text line location: vertical projection which pre-processed by Gaussian filter used to detect the vertical location of title. The reasonable position of title must appear a very steep peak in the vertical direction. The initial position of peak is corresponding with the initial column position of title; the terminative column position is corresponding too. According to the position, we do vertical cutting. The distance between 2 consecutive peaks and their magnitude are also taken into consideration to reduce the risk of false cutting.
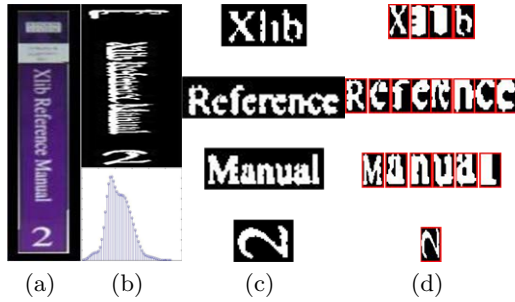
Fig. 5. Title extraction and character splitting process: (a) original image; (b) Vertical projection; (c) Words extraction; (d) Characters splitting

Word location: Morphology dilation operator of size 3x2 is performed on every line of text. The operator links characters of words in to various connected regions which are treated as a word. The word is then rotated 90° to get its horizontal alignment.

Object enhancement and binarization process: the quality of text regions vary so much in different capture conditions. As the result, some image enhancement methods should be adopted to obtain better text segmentation. The accelerated, damped Lucy-Richardson algorithm [11] is used in order to deblur regions of text. The algorithm maximizes the likelihood that the resulting image, when convolved with the PSF (Point Spread Function describes the shape of the image), is an instance of the blurred image, assuming Poisson noise statistics. In our application, we assign PFS with Gaussian function of size 5x5 and sigma equal to 5. Due to non - uniform in background of text regions, using global thresholding for binarization can cause low quality segmented characters. Therefore, Otsu's threshold value is computed on each word's region to get more precise segmentation result.

Characters splitting with multi-threshold method. Due to the book's decoration or image quality, characters are frequently merged together. Finding the best cutting position to split a word into single character is the goal of this part. Normally, characters within a line have similar width. Thus, the average width of characters in non-merged group is a reference value for verify result of splitting process on merged group. For each merged character in the merged group, the multi-threshold is performed depend on the idea: the area where 2 characters are merged must have the smaller gray value than others. Therefore, we gradually increase the threshold value in the merged characters regions until number of connected components more than 2. A connected components is accepted as a single character if it satisfies condition of width equality (compare with reference width), height and aspect ratio. The process is repeated until the merged group is empty. All segmented characters are normalized to standard size of 24x42 pixels and ready for recognition stage. Fig. 5 depicts the whose procedure of extracting characters in book titles.

4 Experimental Results

We conducted number of experiments on a collected dataset to test the effectiveness of proposed method of each steps in the framework. The dataset contains more than 1000 bookshelves images taken in different distances, places and conditions. We simulate movement of camera by classifying bookshelves images into 4 sequences of different distance as shown in fig. 6a (the closest image is distance 0; the further ones are distance 1, 2, 3 respectively).

Fig. 6b to fig. 6e show the typical results of proposed methods. In fig. 6c is result of book regions extraction depend on frequency difference between book and non-book areas. The remains in fig. 6c only cover regions of books. The segmented books of the bottom cell of fig. 6c are shown in fig. 6d. We can see that segmentation process is quite effective. However, some books are still stick together due to their ambiguous boundaries. This problem can be solved by moving camera closer to the shelf. Table 1 and 2 show more results of book extraction and book segmentation process. Fig. 6e is several results of title segmentation and splitting process. In the far distance, characters are almost unreadable or extremely overlapped. It causes difficulties to find the optimal splitting

(a)

(b)   (c)   (d)   (e)
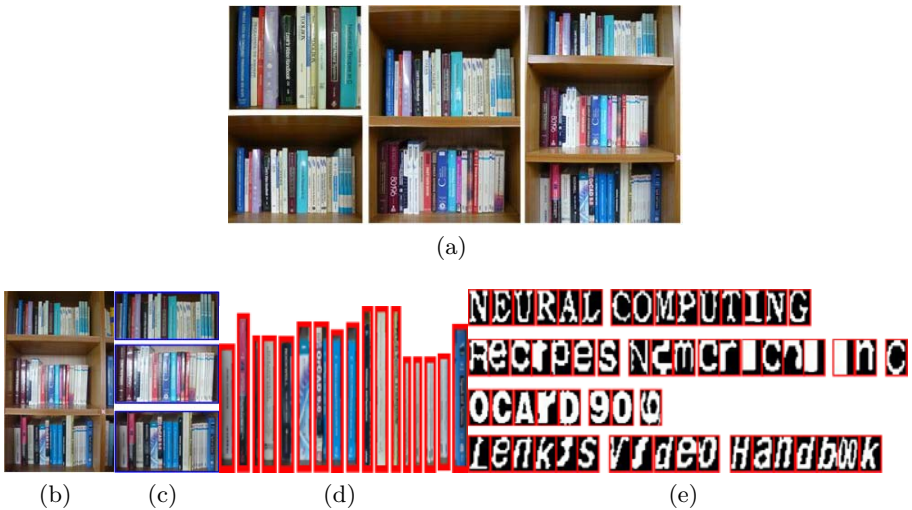
Fig. 6. Experimental results: (a) Typical images in testing dataset; (b) original image; (c) book regions extraction; (d) book segmentation; (e) title segmentation and characters splitting.

Table 1. Book regions extraction

	Number of images	Extraction rate
Distance 1	66	100%
Distance 2	52	100%
Distance 3	37	97.3%

Table 2. Book segmetation

Number of segmented books	Number of good segmentations	Book segmentation rate
2759	2575	93.33%

algorithm. We can overcome such a challenge by the combination among recognition algorithm; syntactic verification and camera movement control.

5   Future Works

The research is getting underway with the remained stages to make the complete assessment of the framework's effectiveness. In order to recognize characters with the wide range of fonts, sizes as well as geometric transformation, SVD is exploiting with adaptive database directly built on segmented characters. Our

Fig. 7. Creating standard "A" from various segmented "A"

database can extremely reduce natural noise and adapt to the standard fonts which regular appear in book titles. Fig. 7 is an example of creating a standard "A" of 50 different "A" segmented formerly. The syntactic verification is being implemented using semantic network.

6   Conclusion

This paper has attempted to provide a comprehensive framework for recognition books on the bookshelves. The main contributions of the paper are as follow:

Designed the system framework that promise to handle almost every addressed challenges of recognition books on bookshelves.

Investigated the main feature difference between book and non-books regions for filter out regions of non-books.

Exploited MSAC based Dominant Vanishing Points in separating books with a good result.

Multi-threshold method for splitting merged characters in book title.

References

1. Minako, S., Hiroshi, M., Norihiro, H.: Character Recognition on Bookshelf Images by Automatic Template Telection. Pattern recognition. In: Pro. 14th International conference, vol. 2, pp. 1117–1120 (1998)
2. Yukari, A., Minoru, I.: Book Recognition from Color Images of Book Shelves. In: MVA 1998 IAPR Workshop on Machine Vision Applications, Makuhari, Chiba Japan (1998)

3. Eiji, T., Seiichi, U., Hiroaki, S.: Block Boundary Detection and Title Extraction for Automatic Bookshelf Inspection. In: 10th Korea-Japan Joint Workshop on Frontiers of Computer Vision, FCV 2005, Fukuoka, Japan (2004)
4. Fischler, M.A., Bolles, R.: Random Sample Consensus. A Paradigm for Model Fitting With Apphcahons to Image Analysm and Automated Cartography. In: Baurnann, L.S. (ed.) Proc. Image Understanding Workshop, College Park, Scmnce Apphcatlons, McLean, Va, pp. 71–88 (1980)
5. Prats, M., Sanz, P.J., del Pobil, A.P.: Model-based Tracking and Hybrid Force/Vision Control for the UJI Librarian Robot. In: Proc. IEEE/RSJ International Conference on Intelligent Robots and Systems, Edmonton, Canada, pp. 1090–1095 (2005)
6. Chitrakala, G., Manjula, D.: Contourlet Based Approach for Text Identification and Extraction from Heterogeneous Textual Images. International Journal of Computer Science and Engineering (2008)
7. Terol, I.R., Vargas, D.: A Study of Openings and Closings with Reconstruction Criteria. In: Proc. 6th International Symposium in Marhemadcal Morphology (2002)
8. Torr, P.H.S.: MLESAC: A New Robust Estimator with Application to Estimating Image Geometry. In: CVIU, vol. 78, pp. 138–156 (2000)
9. Rousseeuw, P.J., Leroy, A.M.: Robust Regression and Outlier Detection. Wiley Inter-science, Chichester (2003)
10. Taira, P.E., Uchida, S., Sakoe, H.: Book Boundary Detection from Bookshelf Image Based on Model Fitting. In: International Symposium on Information Science and Electrical Engineering (2003)
11. Shan, Q., Jiaya, J., Agarwala, A.: High-quality motion deblurring from a single image. ACM Transactions on Graphics 27, 1–73 (2008)

Performance Enhancement of Sum of Absolute Difference (SAD) Computation in H.264/AVC Using Saturation Arithmetic

Trung Hieu Tran, Hyo-Moon Cho, and Sang-Bock Cho

Graduate School of Electrical Engineering, University of Ulsan,
San 29, Mugeo-dong, Namgu, Ulsan 680-749, Korea
tthieu@mail.ulsan.ac.kr, {hmcho67,sbcho}@ulsan.ac.kr

Abstract. Sum of Absolute Difference (SAD) Computation is commonly used for motion estimation in video coding. It is usually the computationally intensive part in video processing. Therefore, a method to reduce the computational complexity is strictly required. In this paper, the effectiveness of saturation arithmetic on SAD computation is presented. Our goal is to use saturation arithmetic to reduce the complexity of SAD computation for the encoding process while the accuracy in finding the best matching block from the reference frame is still maintained. Experiment results show that the computational complexity of SAD computation is reduced efficiently by saving a number of bits for SAD values representation while the video quality is kept.

Keywords: H.264/AVC, SAD, saturation arithmetic, motion estimation.

1 Introduction

Sum of Absolute Difference (SAD) is a very commonly used technique for motion estimation in various video coding standards like H.26x and MPEG-1/2/4 [1]. Motion estimation technique is used to aid temporal prediction in video encoding. The basic need for motion estimation stems from the fact that the consecutive video frames have similarities in most cases that can be intelligently exploited to reduce the number of bits in the encoding process. Most of the consecutive video frames will be similar except for the changes that might be induced by objects moving within frames. The best case for motion estimation will be when the consecutive frames have no frame differences except for the differences caused by noise. In this case, it will be very easy for the encoder to efficiently predict the frame as a duplicate of the reference frame. Once the encoder finds that the frames are same, the only information that needs to be transmitted to the decoder will be the syntactic overhead necessary to reconstruct the picture from the original reference frame. This sounds simple since we are saying that the two consecutive frames are exactly similar. However, in most of the cases, the motion scenes differ at least by a few frames because of the objects moving across the frames in the video. This is when SAD calculation comes into play. It has been shown that this is one of the most effective techniques in motion estimation.

The commonly used arithmetic technique in many video coding standards is the use of modular arithmetic, or modulo arithmetic. In modular arithmetic, the numbers wrap around after they reach a certain value. Suppose if we are using 8 bits for the

representation of a number in its binary form, if we try to add 1 to the binary number 11111111, the actual value in decimal should have been 256. But in modular arithmetic, because the number 256 cannot be represented using 8 bits, it wraps around and this number will be represented as 0, the next number in a circular fashion after 255. Another technique that is gaining wide popularity nowadays is the use of saturation arithmetic. Saturation is the process for determining a maximum value or minimum value which is a correct sign as the result of the arithmetic operation when an overflow or an underflow occurs in an arithmetic system. Let us consider the same example that we discussed for modular arithmetic. Now, after addition, the value becomes 256 and hence cannot be represented using 8 bits. Instead of wrapping around, the value "saturates" at the highest value possible to represent using 8 bits, which will be 255 in this example. So, even if we add any number to any number, the maximum possible number that can be represented using 8 bits in saturation arithmetic will be 255. The representation of information in this case is much better than the former.

In this paper, we present the using of saturation arithmetic to enhance the performance on SAD calculation, an operation frequently used by a number of algorithms for motion estimation. Many techniques have been proposed to speed up the computation of SAD [2], [3], [4]. Saturation arithmetic has become popular in the recent days because the use of it can reduce the computation complexity. Therefore, we can save the resources and the encoder process also will be improved. The rest of this paper is organized as follows: Section II gives some background of motion estimation and the calculation of SAD. Section III presents the use of saturation arithmetic on SAD. The experimental results are shown in section IV, and section V states the conclusion of our paper.

2   Motion Estimation and SAD Calculation

2.1   Motion Vector Search

In video compression [5], [6], [7], video sequences are compressed by exploiting both spatial and temporal redundancy. Spatial redundancy can be seen as small differences between local pixels and temporal redundancy can be seen as small differences between two temporally close video frames. Motion estimation calculates motion vectors by finding matching blocks in the current frame corresponding to blocks in the reference frame. The best match is usually established with the use of the Mean of Absolute Difference (MAD) and the SAD.

The MAD is performed by:

$$MAD(x,y,r,s) = \frac{1}{NN} \sum_{i=0}^{N-1} \sum_{j=0}^{N-1} | A_{(x+i,y+j)} - B_{((x+r)+i,(y+s)+j)} | \qquad (1)$$

MAD can be also written as:

$$MAD(x,y,r,s) = \frac{SAD(x,y,r,s)}{NN} \qquad (2)$$

Where SAD is given by:

$$SAD(x,y,r,s) = \sum_{i=0}^{N-1} \sum_{j=0}^{N-1} | A_{(x+i,y+j)} - B_{((x+r)+i,(y+s)+j)} | \qquad (3)$$

In these equations *(x, y)* is the position of current block and *(r, s)* denotes the motion vector, *N x N* is the block size, usually 16x16, 8x8, or 4x4.

2.2   Computing the Sum of Absolute Difference

The general algorithm computing the SAD of two blocks is depicted in (3). A direct approach in computation the SAD consists of the following steps:

- Compute $(A_i - B_i)$ for all N x N pixels in the two block A and B
- Determine which $(A_i - B_i)$ are less than zero and produce in that case $(B_i - A_i)$ as the absolute value, else produce $(A_i - B_i)$
- Perform the accumulate operation to all N x N absolute values

3   Saturation Arithmetic on SAD

Saturation arithmetic is a version of arithmetic in which all operations such as addition and multiplication are limited to a fixed range between a minimum and maximum value. If the result of an operation is greater than the maximum, it is set to the maximum, while if it is below the minimum it is clamped to the minimum. The name comes from how the value becomes "saturated" once it reaches the extreme values. Nowadays, saturation arithmetic becomes to gain popularity because the use of it can reduce computation complexity; however, this reduction in computation complexity is achieved only at the expense of the resultant value being "approximates". Many assembly level languages like MMX started to support saturation arithmetic [8], [9], specifically for such signal processing applications [10].

The idea used in this paper is to use saturation arithmetic to reduce the complexity of the SAD computation and make sure that we are not losing the accuracy in finding the best matching block from the reference frame. Let us explain the concept by using an example: consider a frame with 100 blocks that can be searched to find the best match from a reference frame. If we don't use saturation arithmetic, the maximum value of SAD can reach up to a value that will use 16 bits. However, in most of the cases, the SAD value will not be as high as a 16 bits value. If there is no difference between two successive frames, the value will be 0 (ideally). Therefore, we would use any bit between 0 to 16 bits to represent the SAD values. Suppose if we have a set of SAD values {6367, 9573, 2184, 12353}, the value with the best matching block will be the one with the minimum SAD value and in this case it will be the block that gave a SAD value of 2184. If we use saturation arithmetic with 14 bits to represent, we can see that some of these SAD values will saturate resulting in a new set of SAD values: {4095, 4095, 2184, 4095}. From this example, it can be seen that since the minimum SAD value is less than the maximum possible SAD value with saturation, we still found the best block even though the other SAD values saturated. Now, consider the case if we use only 11 bits with saturation. The new set of SAD values will be {2047, 2047, 2047, 2047}. This is a difficult case because it can be seen that all the SAD values saturated and hence are same. In this case, finding the best matching block will be difficult. The encoder will pick one of these four blocks randomly and will use it for compression. If the selected block is not the best match, we lose a significant portion of video quality. The chance that the selected block is the perfect match is low in this case. So, this is the problem that was addressed in this paper.

4   Experiment and Results

The H.264 video coding standard used modular arithmetic for all SAD calculations. In our experiment, we rewrote the SAD calculation portions of the H.264 video coding standard by using saturation arithmetic. We varied the number of bits used to represent the SAD values and compared the performance of using modular arithmetic with the performance when saturation arithmetic was used. The H.264/AVC reference software used as a platform for implementation was the JM 15.0 version. When discussing the results, there are two important things that need to be noted, otherwise they might be confusing. Whenever the term "SAD value" is mentioned, it is the already found minimum SAD value from each of the individual matching block and is used in comparing the results of SAD from an entire encoding process instead of just a particular block. Another term is the "Individual SAD", which essentially means the SAD of an individual block and not considering all the SAD values in the whole picture. Noting these two differences will be beneficial to understand the results better. In experimenting with saturation arithmetic for SAD computation, six video input files were used namely: foreman_part_qcif.yuv, foreman_part_qcif_422.yuv, foreman_part_qcif_444.yuv, hall_qcif.yuv, news_qcif.yuv, and mother_daughter_qcif.yuv.

4.1   Snippet Code

The following code is the snippet that was modified in file *macroblock.c* of the H.264/AVC reference software:

```
{...
  for(j=0;j<4;j++)

//Absolute values are being accumulated
 current_intra_sad_2 += iabs(M4[j][i]);
  }

// Saturated value depending on n bits
  if(current_intra_sad_2 > Saturated_value)
  {
   current_intra_sad_2 = Saturated_value;
  }

// Find the best SAD value
  if(current_intra_sad_2 < best_intra_sad2)
   {
   best_intra_sad2=current_intra_sad_2;

// Update best intra mode
   *intra_mode = k; }
   }
// Printing SAD values for analysis
  fprintf(pFile,"Best SAD %d\n",best_intra_sad2);

  best_intra_sad2 = best_intra_sad2/2;
  return best_intra_sad2;
  fclose(pFile);
}
```

4.2   Results

The first experiment was implemented with the set of three video input files available from the H.264/AVC reference software.

   Table 1 shows the minimum and maximum of best SAD value for six video input files using the default number of bits required.

Table 1. Minimum and Maximum of best SAD values

Video Input File (.yuv)	Minimum of best SAD value	Maximum of best SAD value	Default number of bits required
foreman_part_qcif	402	31841	16
foreman_part_qcif_422	357	30941	16
foreman_part_qcif_444	200	31841	16
hall_qcif	924	33447	16
news_qcif	538	56665	16
mother_daughter_qcif	504	20151	16

The following figure shows the distribution of SAD values using modular arithmetic.
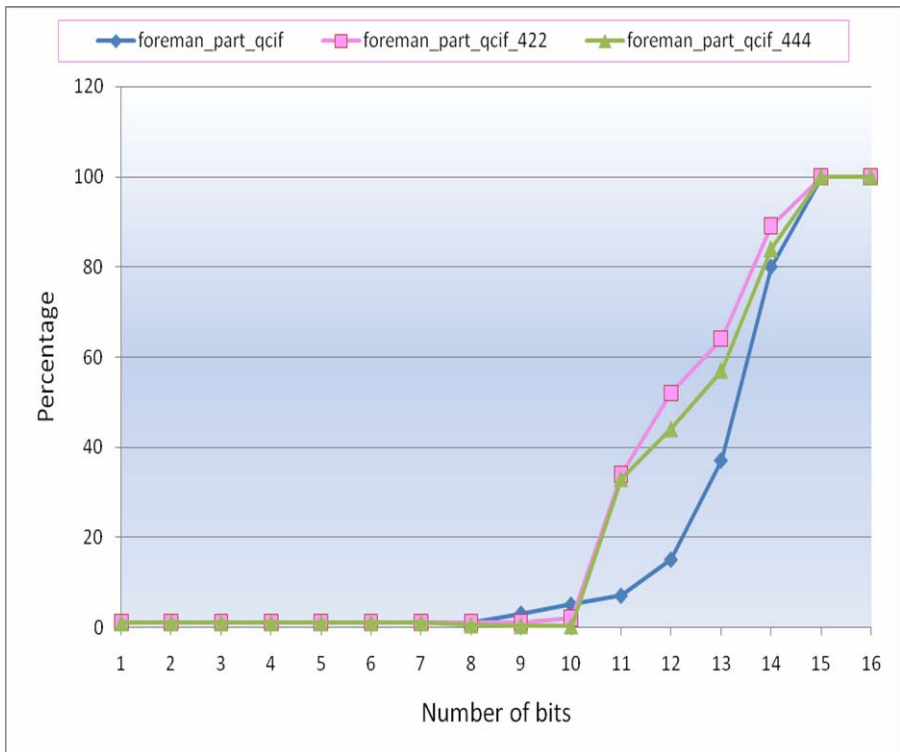
Fig. 1. Percentage of SAD values representation using 'n' bits without saturation (modular arithmetic)

These SAD values are from the whole encoding process, not for individual block matching. Essentially what this figure explains is that the percentage of SAD values that out of all the SAD found can be represented using a particular of bits. For example, let us consider the foreman_part_qcif.yuv video input file without using saturation arithmetic. It shows that for this particular input, 100% of all the minimum SAD values can be represented using 16 bits itself. As we reduce the number of bits that we use to represent, the number falls as shown. Only 80% of the minimum SAD can be represented using 14 bits and 37% of the minimum SAD can be represented using 13 bits. Therefore, this figure shows us what the range of the minimum SAD values is.
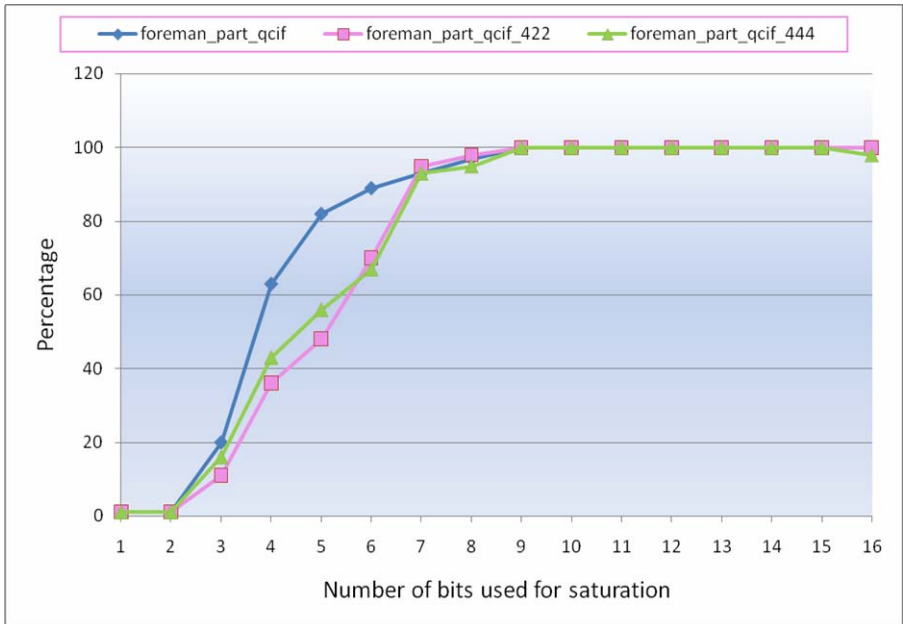
Fig. 2. Percentage of SAD values representation using 'n' bits saturation (saturation arithmetic)

If saturation arithmetic is used, the results are shown in Fig. 2. In this figure, we can see that instead of using 16 bits to represent 100% of the minimum SAD values, we can use only 9 bits to represent 100% of the minimum SAD values. It means that we can save 7 bits in representation of the SAD values and the computation complexity is actually reduced. The encoding process also performs faster with the smaller bits used while the video quality is still kept.

It can be also understood that if the minimum of SAD values saturates, then the rest of the individual SAD values will also saturate. This property was also observed at the output files from the H.264 encoder.

Another experiment was implemented with other set of three video input files: hall_qcif.yuv, news_qcif.yuv, and mother_daughter_qcif.yuv.

The following figure shows the percentage of SAD values represented with 'n'bits without using saturation arithmetic.
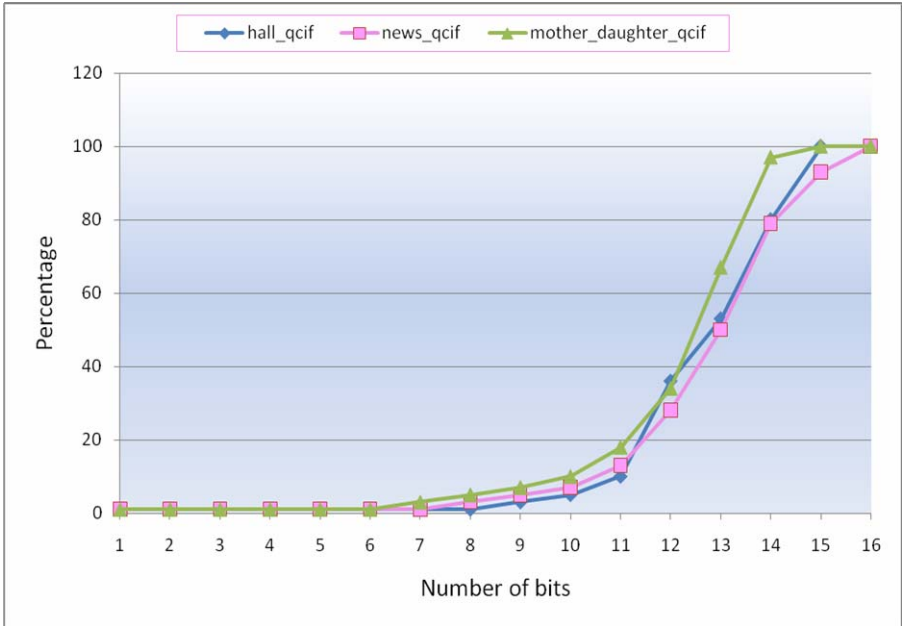
Fig. 3. Percentage of SAD values representation using modular arithmetic
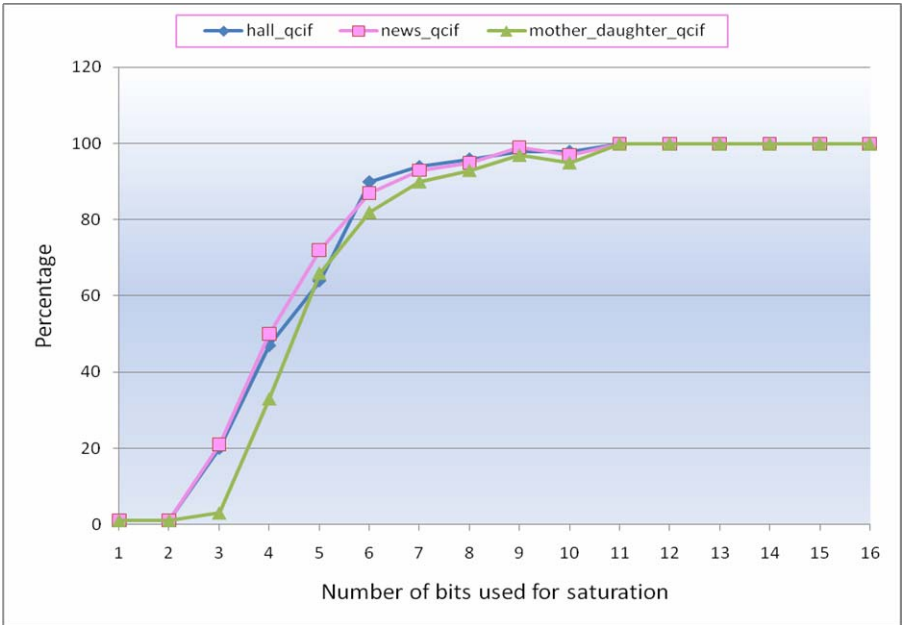
Fig. 4. Percentage of SAD values representation using saturation arithmetic

In Fig. 4, when the saturation arithmetic is used, the percentage of SAD values can be represented using 11 bits instead of 16 bits. It means that we can save 5 bits in representation of the SAD values of video sequences and the computation complexity in the encoding process is reduced. In this experiment, the number of saved-bits is only 5 bits in comparison with 9 bits of the first experiment because the video files used in this experiment are more complex than the ones in the first experiment. Therefore, the number of bits we can save by using saturation arithmetic depends on the complexity of video sequences.

5 Conclusions

Saturation arithmetic has started to gain popularity nowadays because the use of it can reduce computation complexity. In this paper, we present the effectiveness of saturation arithmetic on SAD calculation and motion estimation using the H.264 video coding standard as a platform for implementation. Experiment results show the effectiveness of using saturation arithmetic in video sequences processing. The complexity of SAD computation in H.264/AVC for finding the best matching block can be reduced by saving several bits of SAD value representation using saturation arithmetic whilst the video quality is the same in comparison with using of modular arithmetic. Therefore, saturation arithmetic enables efficient algorithms for many problems, particularly in digital signal and image processing.

Acknowledgements

References

1. Wiegand, T., Sullivan, G., Bjontegaard, G., Luthra, A.: Overview of the H.264/AVC Video Coding Standard. IEEE Transactions on Circuits and System for Video Technology 13, 560–576 (2003)
2. Guevorkian, D., Launiainen, A., Liuha, P., Lappalainen, V.: Architectures for the Sum of Absolute Differences operation. In: IEEE Workshop on Signal Processing Systems – Conference Proceedings, pp. 57–62 (2002)
3. Constantinides, G., Cheung, P., Luk, W.: Synthesis of Saturation Arithmetic Architectures. ACM Transactions on Design Automation of Electronic Systems 8, 334–354 (2003)
4. Yadav, N., Schulte, M., Glossner, J.: Parallel Saturating Fractional Arithmetic Units. In: Proceedings of Ninth Great Lakes Symposium on VLSI, pp. 241–217 (1999)
5. Gall, D.: MPEG: A Video Compression Standard for Multimedia Applications. Communications of the ACM 34, 46–58 (1991)
6. Mitchell, J., Pennebaker, W., Fogg, C., LeGall, D.: MPEG Video Compression Standard. Springer Press, Heidelberg (1997)
7. Marpe, D., Schwarz, H., Wiegand, T.: Context-adaptive binary arithmetic coding in the H.264/AVC video compression standard. IEEE Transactions on Circuits and System for Video Technology 13, 620–636 (2003)

8. Lee, B., Fiskiran, M.: PLX: A Fully Subword-Parallel Instruction Set Architecture for Fast Scalable Multimedia Processing. In: IEEE International Conference on Multimedia and Expo – Conference Proceedings, pp. 117–120 (2002)

9. Peleq, A., Weiser, U.: MMX Technology Extension to the Intel Architecture. Micro, IEEE, 42 – 50 (1996)

10. Luo, Z., Lee, R.: Cost-Effective Multiplication with Enhanced Adders for Multimedia Applications. In: Proceedings of IEEE International Symposium on Circuits and Systems, vol. 1, pp. 651–654 (2000)

A Parabolic Detection Algorithm Based on Kernel Density Estimation

Xiaomin Liu, Qi Song, and Peihua Li

School of Computer Science and Technology, Hei Long Jiang University
Harbin, Hei Long Jiang Province, 150080, China
peihualj@hotmail.com

Abstract. The traditional Hough transform needs the edge detection in advance, so the effect of edge detection influences the final fitting result. This paper proposes a new method of detecting parabolas using the kernel density estimate based on the theory of Rozenn Dahyot, and extends this method into the eyelid detection in noisy images and other images including parabolas. In our paper, the edge detection is not necessary. On one hand, we not only consider the current points on the parabola, but also ones around the parabola. Experiments demonstrate that the proposed algorithm is robust and insensitive to the noise.

Keywords: kernel density estimate, parabolic detection, eyelid detection, Hough transform.

1 Introduction

With the development of biotechnology, iris recognition has had an increasingly wide application in all fields, but the issue has arisen in the images which are captured in less constrained environment. Iris recognition technology involves iris preprocess and iris recognition. Iris preprocess mainly includes iris inner and outer boundary segmentation and removing the noise. At present, compared with iris inner and outer boundary segmentation, the method of removing the noise still has many problems to be solved, and particularly the eyelid affects the segmentation and recognition of the iris.

The approaches to detect eyelid include: the integro-differential operator based on the parabolic model [1], edge detection combined with the Hough transform [2] and shortest path method [3]. The three methods are mainly based on edge detection or corner point detection, so the edge points affect the final result of eyelid fitting severely.

Responding to their problems, this paper proposes a new method to detect eyelids with the kernel density estimate. Rozenn Dahyot used the kernel density estimate method to detect lines and achieved good results [4]. While determining whether the point falls into the area of a given line, we not only consider the current points, but also consider the contribution of its neighborhood, which avoids effects of the wrong edge points. This paper takes advantage of Rozenn Dahyot's

idea to fit the eyelids, which doesn't only consider the edge points, but their gradient. According to every point's gradient and orientations, we can determine every point's contribution. Furthermore, the experiments are performed on the images with lots of noises. Experiment results demonstrate that the proposed algorithm is robust and insensitive to noisy images.

2   Our Method

First of all, we introduce the concept of the kernel density function briefly. The kernel model of the parabola is proposed and we illustrate how to set the parameters of the model. At last the process of our method is introduced.

2.1   Introduction of Kernel Density Function

The kernel density function [5]: when we estimate one point, not only consider the current point, but also its neighborhood. Moreover the weight of every point in the neighborhood is different. The weight function given by a kernel function K satisfies the condition

$$\int_{-\infty}^{+\infty} K(x)\, dx = 1, \tag{1}$$

where K is a symmetric probability density function, and the kernel estimator with kernel function K is defined by

$$\hat{f}(x) = \frac{1}{Nh} \sum_{i=1}^{N} K\left(\frac{x - x_i}{h}\right), \tag{2}$$

where h is the window width, also called the smoothing parameter or bandwidth. Generally choose Gauss function as the kernel.

2.2   Kernel Function Model of Parabola

Given the equation of parabola

$$f(x) : y = a(x - b)^2 + c, \tag{3}$$

S is the data set and the kernel density function of the parabola is:

$$\hat{p}_{a,b,c}(a, b, c | S) = \sum_{i=1}^{N} K_\theta\left(\frac{\theta - \theta_i}{h_{\theta_i}}\right) K_x\left(\frac{x - x_i}{h_{x_i}}\right) K_y\left(\frac{y - y_i}{h_{y_i}}\right) \delta(f(x))\, p_i, \tag{4}$$

where $K_\theta\left(\frac{\theta - \theta_i}{h_{\theta_i}}\right)$ represents the kernel function of gradient, $K_x\left(\frac{x - x_i}{h_{x_i}}\right)$ represents the horizontal kernel function, $K_y\left(\frac{y - y_i}{h_{y_i}}\right)$ represents the vertical kernel function, and $\delta(f(x))$ represents the Dirac Function [6] of parabolic equation; p_i represents the probability of every data in data set S.

If by integration with respect to the spatial variables (x, y), the last-written could convert to the following formula:

$$\hat{p}_{a,b,c}(a, b, c|S) = \sum_{i=1}^{N} K_\theta \left(\frac{\theta - \theta_i}{h_{\theta_i}} \right) R_i(a, b, c) p_i, \tag{5}$$

where $R_i(a, b, c)$ is the Radon transform [7] of the spatial kernels:

$$R_i(a, b, c) = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} \delta(y - a(x - b)^2 - c) K_x \left(\frac{x - x_i}{h_{x_i}} \right) K_y \left(\frac{y - y_i}{h_{y_i}} \right) dx dy. \tag{6}$$

2.3   Parameters Setting

Because p_i is the probability of every data in data set S. Moreover for the images, the distribution of the pixels is uniform and we have $p_i = \frac{1}{N}$. Generally speaking, the kernel function is symmetrical, so we choose the Gauss function as the kernel:

$$K_\theta \left(\frac{\theta - \theta_i}{h_{\theta_i}} \right) = N(\theta_i, h_{\theta_i}^2) = \sqrt{2\pi} h_{\theta_i} \exp \left(\frac{-(\theta - \theta_i)^2}{2h_{\theta_i}^2} \right), \tag{7}$$

$$K_x \left(\frac{x - x_i}{h_{x_i}} \right) = N(x_i, h_{x_i}^2) = \sqrt{2\pi} h_{x_i} \exp \left(\frac{-(x - x_i)^2}{2h_{x_i}^2} \right), \tag{8}$$

$$K_y \left(\frac{y - y_i}{h_{y_i}} \right) = N(y_i, h_{y_i}^2) = \sqrt{2\pi} h_{y_i} \exp \left(\frac{-(y - y_i)^2}{2h_{y_i}^2} \right). \tag{9}$$

The Radon transform corresponding to the spatial kernels is:

$$R_i(a, b, c) = \frac{1}{2\pi h_{x_i} h_{y_i}} \int_{-\infty}^{+\infty} \exp \left(\frac{-(x - x_i)^2}{2h_{x_i}^2} \right) \exp \left(\frac{-(a(x - b)^2 + c - y_i)^2}{2h_{y_i}^2} \right) dx. \tag{10}$$

Because the result of Radon transform contains the improper integral, it could not obtain analytic expression. Thus we use Gauss-Hermite numerical method to solve the integral.

In the Eq. (10), the kernel function's bandwidth determines the scale of Gauss function and reflects how many neighborhood points are considered and how much their contributions are.

We can use the adaptive way to choose the horizontal bandwidth h_{x_i} and the vertical bandwidth h_{y_i} without any priori conditions. In the experiments, we set the spatial bandwidth as:

$$h_{x_i} = h_{y_i} = 1, \forall_i. \tag{11}$$

According to the gradient bandwidth h_{θ_i}, we adopt the self-adaptive variable bandwidth. According to the gradient of every point, we can choose the size of bandwidth. If the gradient is high, we consider that this point perhaps is the

edge point and we reduce the bandwidth, otherwise we increase the bandwidth and accept its neighborhood's contribution. The bandwidth is expressed as:

$$h_{\theta_i} = \frac{\sigma}{\|\nabla I_i\|}, \tag{12}$$

where σ is a priori value, represents the standard deviation of the noise distribution [4] and we can set it as 0, 20, 50 etc. $\|\nabla I_i\|$ represents the amplitude of the gradient of the point i and we can compute it as follows:

$$\|\nabla I_i\| = \|\nabla I(x_i, y_i)\| = \sqrt{I_x^2(x_i, y_i) + I_y^2(x_i, y_i)}, \tag{13}$$

where $I_x(x, y)$ and $I_y(x, y)$ represent separately the horizontal and vertical gradients of the pixel i [8].

2.4 Algorithm Flowchart

As shown in Fig. 1. First of all, we compute every pixel's horizontal and vertical gradients, and calculate each pixel's amplitude of the gradient $\|\nabla I_i\|$ and orientation θ_i. We utilize the kernel formula, search all values of a, b, c and calculate homologous probability of each different a, b, c. At last, we obtain the $3D$ probability density function's peaks, containing the parabolas which we want to detect.
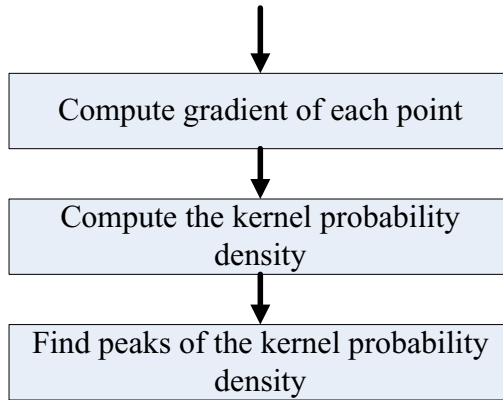
Fig. 1. Flowchart of our method

3   Experiment Results and Comparisons

This algorithm uses $C++$ language and runs on PC which is $1.8GHz$, Intel Pentium 4 and $2G$ memory, without any code optimization. This algorithm mainly contrasts with the traditional Hough transform and the detection effects of the proposed algorithm are better than it.

3.1   Upper Eyelid Detection

In the iris preprocess, the eyelid detection is very important. We propose a eyelid detection method based on the kernel density estimate, which has better performances than the traditional Hough transform detection.

Algorithm Implement. The proposed algorithm is performed on the sub-image, which is extracted with the eye detection method. After the inner and outer boundaries are located, we obtain the outer boundary center (x_0, y_0) and radius R. Meanwhile, the range of the three parameters a,b,c is determined, according to the parabolic equation $f(x) : y = a(x - b)^2 + c$. As shown in Fig. 2, $0.0025 \le a \le 0.3/R$, $x_0 - R/2 \le b \le x_0 + R/2$, $y_0 - R \le c \le y_0 - R/3$.
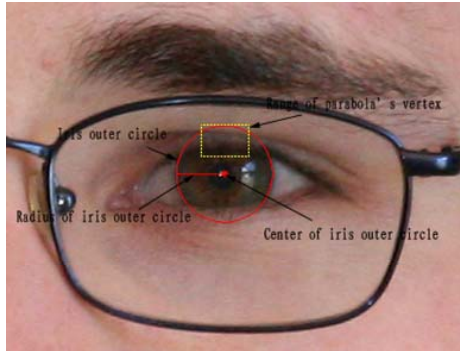
Fig. 2. Ranges of a,b,c

Next, we compute every pixel's horizontal and vertical gradients, and then calculate every pixel's amplitude of the gradient $\|\nabla I_i\|$ and orientation θ_i.

And then we utilize the kernel formula, search all values of a, b, c and calculate homologous probabilities of different a, b, c. Because of the heavy noise, we obtain several parabolas, such as, the double eyes and eyelashes etc. So we obtain top 4 parabolas, which contain the maximum $3D$ probability density function values. Because the changes of gray levels around the eyelid are apparent, we choose the parabola from the top 4 parabolas, whose difference of every point's gray level along the vertical direction of the parabola is maximum. Fig. 3 shows an example of our proposed algorithm, where the green parabolas are top 4 parabolas and the red one is final detection result.

Database and Truth Value. We use the UBIRIS.v2 iris database [9] and this database includes 500 iris images with lots of noises, where 28 iris images are close eyes or rotate eyes. Our experiment tests on the left 472 iris images. The size of the images is $400 * 300$.

The truth value is marked by hand and $L^2 - norm$ [10] is used to compare the test value with the truth value. The formula is as follows:

$$E_i = \|y_{t,i} - y_{o,i}\| \triangleq \left(\frac{1}{x_2 - x_1} \int_{x_1}^{x_2} (y_{t,i} - y_{o,i})^2 \, dx \right)^{\frac{1}{2}}, \qquad (14)$$

(a) Original image
(b) Sub image of the eye detection
(c) Top 4 (green parabolas) and final detection result (red parabola)
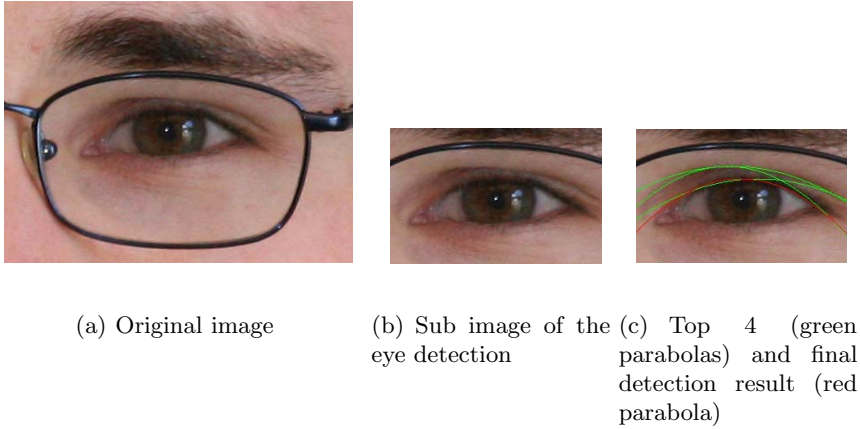
Fig. 3. The experimental results of the upper eyelid detection

where $y_{t,i}$ represents the truth value of image i and $y_{o,i}$ represents our test value of image i. The integrating range $[x_1, x_2]$ is based on the range of truth. The average error of 472 images is calculated as follows:

$$\overline{E} = \frac{1}{N} \sum_{i=1}^{N} \|y_{t,i} - y_{o,i}\|, \tag{15}$$

where N denotes the number of images.

Contrast Result. Our algorithm is compared with the improved Hough transform. The difference between the improved Hough transform and the traditional Hough transform is that the former parabola is chosen from the top 4 parabolas, whose difference of each point's gray level along the vertical direction of the parabola is maximum. In the Sobel operator, we choose the different percentage of the edge points to get the following detection results.

3.2 Other Applications

From the physics point of view, parabolic-type structure has favorable compressive capacity, so the parabolic-type structure is widely used in the construction

Table 1. Errors for upper eyelid localization

Method	Percentage	Average errors (pixels)
Our method		18.13
Sobel+Hough	0.6	20.14
Sobel+Hough	0.4	20.03
Sobel+Hough	0.2	20.47
Sobel+Hough	0.01	29.71
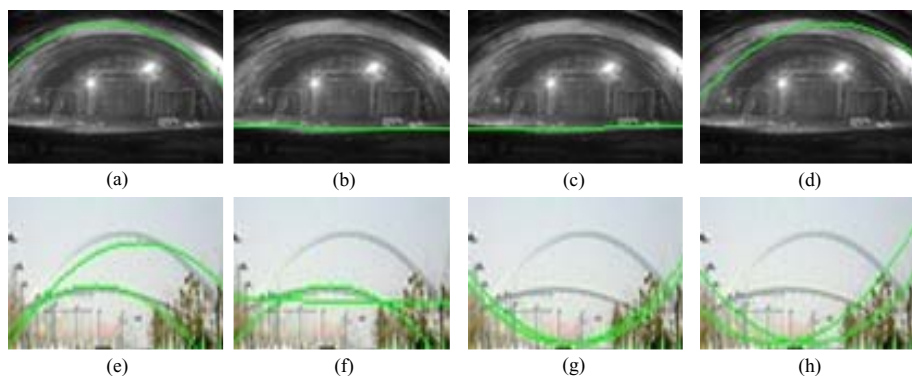
Fig. 4. The detection results of the kernel density estimate and Hough

design. Our experiments are performed on the images of the bridge arc and the tunnel, shown in Fig. 4, where (a), (e) are the results of the kernel density estimate, (b), (f) are the results of the Hough transform with 10 percent edge points, (c), (g) are the results of the Hough transform with 50 percent edge points and (d), (h) are the results of the Hough transform with 90 percent edge points.

4   Conclusion

This paper proposes a new method to detect parabolas with kernel density estimate, and we extend it to detect eyelids in the iris images with heavy noise. The experiment shows that this method is better than the traditional Hough transform. In the kernel density estimate, the edge detection is not necessary, so this method avoids the issues of coming from edge detection error or threshold picking. What's more, our method is general and is applied to many fields. And we can make use of the method to detect the eyelid in the iris preprocess, and also detect the circles, ellipses and conics. Our future work is to detect other conics with kernel density estimate and is used in practice.

Acknowledgements

References

1. Daugman, J.: How Iris Recognition Works. IEEE Transactions on Circuit and System for Video Technology 14(1), 21–30 (2004)
2. Wildes, R.: Iris Recognition: an Emerging Biometric Technologh. Proceeding of the IEEE 85(9), 1348–1363 (1997)
3. Pardas, M.: Extraction and Tracking of the Eyelids. In: IEEE International Conference on Acoustics, Speech, and Signal Processing, Istanbul, pp. 2357–2360 (2000)
4. Rozenn, D.: Statistical Hough Transform. IEEE Transactions on Pattern Analysis and Machine Intelligence (submission) (2008)
5. Silverman, B.W.: Density Estimation for Statistics and Data Analysis. In: Monographs on Statistics and Applied Probability, London (1986)
6. Radon Transform, http://mathworld.wolfram.com/RadonTransform.html
7. Deans, S.R.: Hough Transform from the Radon Transform. IEEE Transactions on Pattern Analysis and Machine Intelligence 3(2) (1981)
8. Dahyot, R., Wilson, S.: Robust Scale Estimation for the Generalized Gaussian Probability Density Function. Advances in Methodology and Statistics (Metodoloski zvezki) 3(1), 21–37 (2006)
9. UBIRIS. v2 Database. Website (2008), http://iris.di.ubi.pt/ubiris2.html
10. Blake, A., Isard, M.: Active Contours. Springer, London (1998)

An Intelligent Prediction Model for Generating LGD Trigger of IEEE 802.21 MIH

M. Yousaf, Sohail Bhatti, Maaz Rehan, A. Qayyum, and S.A. Malik

Center of Research in Networks and Telecommunications (CoReNeT),
M. A. Jinnah University, 44000 Islamabad, Pakistan
myousaf@ymail.com, rsmbhatti@yahoo.com, maazrehan@yahoo.com,
aqayyum@ieee.org, smalik@comsats.edu.pk

Abstract. IEEE recently standardized 802.21-2008 Media Independent Handover (MIH) standard. MIH is a key milestone toward the evolution of integrated heterogeneous 4G wireless networks. MIH provides number of link layer events in a unified way that facilitate upper layer protocols in making handover decisions. One such event is Link Going Down (LGD) trigger. LGD is a predictive event that is generated when link conditions are expected to degrade in near future. Traditionally such link quality degradations and connectivity losses are predicted on the basis of a single parameter only i.e. received signal strength. However, in varying wireless conditions, simple predictions relying on single link layer parameter may generate false LGD triggers. This false triggering may initiate unnecessary handovers that rather than facilitating upper layer mobility management protocols, may cause overhead and may degrade the overall network performance. In this paper, we present an intelligent model for generating MIH LGD trigger reliably. In our implementation, we used 'Time Delay Neural Networks (TDNN)' approach using multiple link layer parameters for LGD predictions. We also analyzed the prediction accuracy and the feasibility of using such intelligent technique for mobile devices.

Keywords: Mobility Management, Vertical Handovers, IEEE 802.21-2008 MIH, LGD Trigger, IEEE 802.11u, TDNN, FFNN.

1   Introduction

Multihomed mobile devices with different network interfaces have received significant attention. To enable users to move from one access network to another many mobility management protocols have been proposed. These protocols operate at layer-3 or above of the networking protocol stack and often use link layer information to make efficient handover decisions. IEEE 802.21-2008 Media Independent Handover (MIH) Standard provides this link layer intelligence to upper layer mobility management protocols in a media independent, seamless and efficient manner [1]. MIH provides the information through MIH event, command and information services. One such event is Link Going Down (LGD) event that helps to timely initiate the handover procedure thus reducing handover latency

and packet losses during handovers. MIH standard defines it as a predictive event that estimates the expected link behavior on the basis of past and current link behavior. However standard does not define any algorithm to predict this Link Going Down event. Mobility management protocols can initiate exchange of the handover control messages upon reception of LGD trigger. So false triggering of LGD event can initiate unnecessary handovers thus can cause overhead. So, there is a need to generate this LGD trigger intelligently.

Most of the research efforts regarding IEEE 802.21 MIH discuss the use of MIH services for handover decision making. So far little work has been done on exploring the issues and problems related to the MIH itself. Timely effective handover mechanism using MIH primitives have been discussed in [3]. Implementation and performance study of 802.21 has been described in [4]. Prediction of LGD event using least mean square technique has been discussed in [5]. To the best of our knowledge, so far no work has been done regarding LGD predictions using intelligent techniques. We used artificial neural network techniques to train our prediction module over varying link conditions and connectivity status. After learning the link behavior, trained neural networks are used for accurate and timely triggering of LGD trigger. We also evaluated the prediction accuracy as well as the feasibility of using such intelligent technique for the resource limited mobile devices.

Rest of the paper is organized as follows: Section 2 briefly describes the MIH standard, Link Going Down trigger, its advantages and related issues. Section 3 explains our intelligent prediction model for generating LGD trigger. Section 4 describes our experimentation setup. Section 5 evaluates the results of our experimental implementation and the feasibility of this intelligent model for mobile devices. In the end, section 6 concludes the paper.

2   IEEE 802.21-2008 MIH Standard and Link Going Down Trigger

IEEE 802.21-2008 MIH (Media Independent Handover) standard defines the mechanisms and necessary support to optimize the handover process of mobile nodes between IEEE 802 family of networks, 3GPP and 3GPP2 cellular networks. In order to provide link layer intelligence to upper layer mobility management protocols, MIH defines a logical entity *Media Independent Handover Function* (MIHF) in the protocol stack. MIHF receives information from link layers and provides it to the upper layers thus, hiding link specific complexities from the upper layers. Entity using this information is called the *MIH User*. MIH User can be a Handover Decision Algorithm [6] or any Layer-3 and above mobility management protocol, such as Mobile IP variants [7], TCP-Migrate [8], or EMF [9].

One of the services provided by MIHF is *Media Independent Event Service* that informs about changes in the local as well as remote link layer properties in the form of link events or triggers. These triggers include i) Link Detected, ii) Link Up, iii) Link Going Down, iv) Link Down, etc. Link Going Down (LGD) is

a predictive event that indicates the likelihood of link connectivity failure in near future due to degraded link conditions. Prediction of this possible disconnectivity is based on the past and present conditions of link parameters e.g. link quality, signal strength, etc. LGD event indicates that mobile node is moving away from the current access network and may soon be out of the coverage area of this network. Mobility management protocols such as mobile IP can significantly improve their performance using timely triggering of LGD event. This *make before break* can significantly reduce the handover latency, decrease packet loss ratio during handover and hence improve the user experience while on the move. An MIH compliant link can generate this event using some predictive algorithm that predicts the future link conditions and infers whether link connectivity will be maintained or not.

LGD trigger indicates the expectation that the link will go down in near future. However, link is not down yet. Handover is a costly process that involves exchange of multiple control messages at different layers of protocol stack. False LGD triggers may initiate unnecessary link layer authentication and association messages, IP address acquisition messages, mobility management control messages, etc. and thus may cause overhead. Therefore a mechanism is required that can generate LGD triggers intelligently thus minimizing the chances of false LGD triggering. Our proposed intelligent prediction model takes into account multiple link parameters, learn behavior of these parameters under varying link conditions. Then the model infers connectivity status for upper layer applications. If applications are expected to loose link connectivity under predicted link conditions, only then LGD trigger is generated.

MIH also defines new link layer service access points to get link specific information from the corresponding link technologies. MIH specific amendments for IEEE 802.11 (WiFi) are being defined as IEEE 802.11u [2]. Our proposed LGD prediction module resides in the local Station Management Entity (SME) of 802.11u management plane as depicted in figure 1. SME monitors local link conditions and generates *MS SME-Link-Down-Prediction.indication* message for the MAC State Generic Convergence Function (MSGCF). MSGCF then generates *MSGCF-ESS-Link-Going-Down.indication* message that is equivalent to the *Link-Going-Down.indication* message of link layer and passes it to the MIHF. MIHF on reception of this message checks whether there is some subscription for this event from any MIH user or not. If there is a subscription, then MIHF generates the *MIH-Link-Going-Down.indication* message to the MIH users that in turn, can take necessary handover decision.

3   Intelligent Prediction Model for LGD Trigger

Artificial Neural Networks (ANN) are information processing systems that are developed as generalization of mathematical models of human cognition. In our intelligent model, we used two types of neural networks i.e. Time Delay Neural Network (TDNN) and Feed-Forward Neural Network (FFNN). Both networks
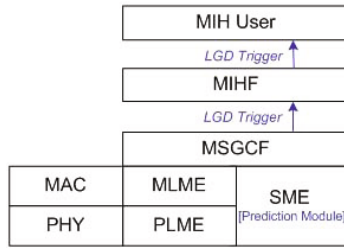
Fig. 1. LGD Prediction Module for IEEE 802.11u
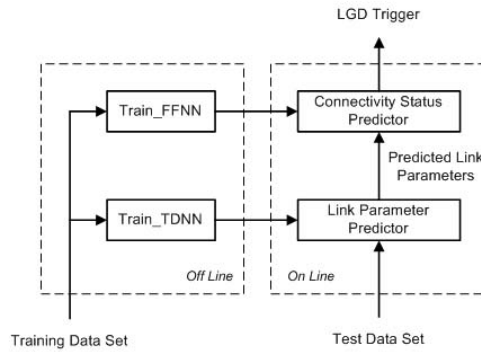
Fig. 2. Prediction Model for LGD

are first trained to learn the connectivity status under varying link conditions. Then TDNN is used for predicting future link conditions and FFNN is used to determine that with these predicted conditions, whether upper layer connectivity will be maintained or not. This model is depicted in figure 2.

3.1   Using Time Delay Neural Network

A TDNN is a temporal network with its input patterns successively delayed in time [10]. We use TDNN for the prediction of future conditions of link parameters based on the past and current values of these parameters. First, TDNN is trained over the training data consisting of real measurements of link parameters. The future link values are kept as the target for this training. To test the accuracy of trained TDNN, a window of past values of link parameters from test data set is input to the TDNN and future values of these parameters are predicted. These predictions can be termed as discrete in time because, we predict a certain number of samples of link parameters that are 200 msec apart from each other in time.

Output of TDNN $y(t)$ at time t depends upon the input values $x(t)$ at times *(t-1), (t-2), ..., (t-n)*. Here n is the number of delayed samples (window size) of link parameters.

$$y(t) = f(x(t-1), x(t-2), ..., x(t-n))$$
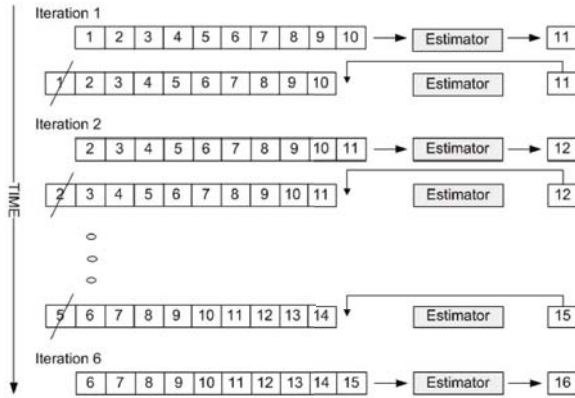
Fig. 3. Estimating future link conditions using TDNN

Similarly for predicting next value, we take currently predicted output *y(t)* as current input, coupled with *(n-1)* delayed samples, and predict the values of next sample *y(t+1)* of link parameters.

$$y(t + 1) = f(y(t), x(t - 1), x(t - 2), ..., x(t - n - 1))$$
$$y(t + 2) = f(y(t + 1), y(t), x(t - 1), x(t - 2), ..., x(t - n - 2))$$
....
$$y(t+m) = f(y(t+m-1), y(t+m-2), ..., y(t), x(t-1), x(t-2), ..., x(t-n-m))$$

This prediction process is depicted in figure 3. Now we have *m* number of predicted samples of future condition of link parameters. These predicted samples are input to the FFNN for deciding about the connectivity status.

3.2 Using Feed-Forward Neural Network

In FFNN, signals flow from input unit to output unit in forward direction. As part of training process, a back-propagation function is used to minimize the error [11]. For training the FFNN classifier, we have provided link parameter samples *x(t)*, and train FFNN to learn the connectivity status *C(t)* under these link conditions. In this way, FFNN network learns that under what link conditions upper layers experience the link connectivity and under what conditions upper layers loose the connectivity. In our experiments, a link parameter sample *x(t)* taken at time *t* is 4-tuple:

$$x(t) = \{v_1(t), v_2(t), v_3(t), v_4(t)\}$$

Here $v_1 = SignalLevel$, $v_2 = NoiseLevel$, $v_3 = LinkQuality$ & $v_4 = BitRate$

Let upper layer connectivity status *C(t)* at time *t* is the function of these link parameters:

$$C(t) = f(x(t)) = f(\{v_1(t), v_2(t), v_3(t), v_4(t)\})$$

We have used ICMP (Internet Control Message Protocol) traffic as upper layer traffic. Reception of ICMP message means that upper layer has the connectivity and drop of ICMP packet means that application has lost the connectivity. After learning the link behavior, we provide the predicted link parameters from TDNN as input to the FFNN to classify that under these predicted link conditions, whether connectivity $C(t)$ will be retained or not. For each predicted sample $y(t)$, FFNN classifies the connectivity as predicted connectivity status $C_p(t)$. Average of these predicted $C_p(t)$ decides whether the LGD event should be triggered or not.

$$C_{Avg}(t) = \frac{1}{m} \sum_{i=1}^{m} C_p(t_i)$$

$$C_{Avg}(t) = \begin{cases} Connectivity \qquad\qquad Lost; & C_{Avg}(t) > 0.5 \\ Connectivity\ Maintained; & C_{Avg}(t) \leq 0.5 \end{cases}$$

$$if(C_{Avg}(t) > 0.5),\ Then\ 'Trigger\_LGD\_Event',\ Else\ 'Ignore'$$

4   Experimental Setup

Following are the steps to evaluate the proposed prediction model:

1. Capture link parameter values along with upper layer connectivity status.
2. Preprocess the captured data. Filter out irrelevant details.
3. Train neural networks to learn link behavior and connectivity status in varying link conditions.
4. Deploy trained neural networks in prediction model. Using real time captured values, predict future values of link parameters. And predict whether under these predicted link conditions connectivity will be maintained or lost.

In these experiments, we have used ZyXel P-320W IEEE 802.11g Wireless Firewall Router that also acts as wireless Access Point (AP). Laptop with Fedora Core-9 Linux is used as mobile node with Intel Celeron 2.0GHz microprocessor and 512 MB DDR having AR 242x802.11abg WLAN Adapter with Madwifi driver. We capture real data of IEEE 802.11 link parameters using Wireless Tools (WT) for Linux. To associate the link parameters with upper layer connectivity, we have modified the source code of Linux Ping command. We periodically capture link conditions with period equal to the Ping interval. This way we have recorded link conditions under which upper layer packets are received as well as link conditions under which upper layer packets are dropped. We have simulated varying link conditions by changing mobile node's location and mobility patterns as depicted in figure 4.

All captured data is logged for further processing and analysis. Before providing this data to train the neural networks, we filter unnecessary parameters that either don't change their behavior during the experiments or increase monotonically such as sequence numbers and counter values. After filtering process, four link parameters i.e. signal level, noise level, link quality and bit rate along with
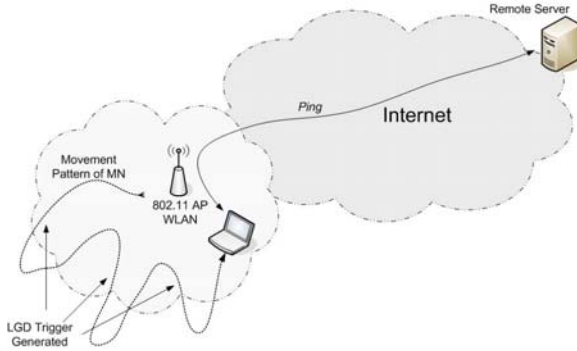
Fig. 4. Movement Pattern of MN

corresponding connectivity status of the link are chosen as input for learning and prediction modules.

Matlab R2008a is used to implement the neural networks. Captured data is divided into two sets; one is *training data set* and other is *test data set*. Training data set is used to train both the TDNN and FFNN. For testing the correctness of trained model we input test data to the TDNN. In different experiments, a variable number of samples (window size) of live data of link parameters is passed to the trained TDNN for predicting a certain number of future samples of link parameters. The output of this step is the predicted samples that are input to FFNN to decide the connectivity status.

We have used two hidden layers for each of the network with 5 neurons in layer-1 and 3 neurons in layer-2. We have used *trainlm* for our network's training function that updates weight and bias values using Levenberg-Marquardt optimization. Other functions that are used for these experiments are *learngdm* (Gradient Descent with Momentum) as learning function, *tansig* (Hyperbolic Tangent Sigmoid) as transfer function at hidden layer-1 and *purlin* (pure linear) as transfer function at hidden layer-2 and *mse* (Mean Squared Error) as performance function.

Considering the limited resources of mobile devices in terms of computation, memory space and battery life, we have performed the training of neural networks offline. Then trained neural networks are ported to mobile nodes for real time predictions. Predictions and classification using trained networks is less computation intensive and requires little memory space to make the decisions thus rendering this predictive model feasible for mobile nodes. Moreover we are using statically trained neural networks that also save the periodic computation cost of adaptive learning. Even with these statically trained neural networks we achieved moderate to high accuracy in our experiments.

5   Performance Analysis of Proposed Model

We have conducted large number of experiments with window sizes of 10, 20, 30, 40, 50, 60, 70, 80, 90, & 100 delayed samples and predicted next 2, 4, 6, 8, 10 samples for each of these window sizes and experimented with 5 different test data sets.

		Window Size									
		10	20	30	40	50	60	70	80	90	100
Predictions	2	0.8	0.7	0.7	0.8	0.8	0.8	1.0	0.7	0.7	0.8
	4	0.8	0.8	0.7	0.8	0.7	0.8	0.8	0.7	0.7	0.8
	6	0.8	0.7	0.7	0.8	0.8	0.8	0.8	0.7	0.7	0.7
	8	0.8	0.7	0.7	1.0	1.0	0.7	0.8	0.7	0.7	0.8
	10	0.8	0.7	0.7	1.0	1.0	0.7	0.7	0.7	0.8	1.0

Fig. 5. Prediction Accuracy Table

		8-Predictions	10-Predictions
Time (msec)	Estimation Cost	376	470
	Classification Cost	144	180
	Totol Prediction Cost	520	650
	Prediction For	1600	2000
	Time Gain	1080	1350

Fig. 6. Time Cost & Gain of LGD Prediction

5.1 Degree of Prediction Accuracy and Time Cost

Figure 5 depicts the summary of results of link going down prediction accuracy. Results show the dependence of prediction accuracy upon the number of delayed samples (window size) given as input for prediction and the number of future samples being predicted. Results show that a 100% accuracy in predicting the LGD trigger is achieved for window sizes of 40 & 50 samples. It means that a moderate number of previous samples (in our experiments 40 to 50 samples i.e. data of last 8 to 10 seconds respectively) is sufficient to predict the LGD trigger with a high degree of accuracy. Too small number of delayed values do not provide enough information to predict the LGD trigger correctly. Similarly too large number of delayed values also cause to effect the correct prediction of LGD trigger. As last 40 and 50 delayed samples both give 100% accuracy but computation cost and memory requirements for 50 delayed samples is higher therefore, we recommend to use the data of last 8 seconds for efficient and effective LGD predictions.

Prediction accuracy is highest when we make the decision on the basis of next 10 predicted samples (2 seconds time interval). However with window size 40 and 50 we have 100% accuracy for both next 8 samples and for next 10 samples. Although computation cost of making prediction for next 8 samples is less than that of making prediction for next 10 samples however, we recommend to use the prediction of next 10 samples. The reason for this recommendation is the time gain that a mobile node can have for the time interval of next 10 samples. Figure 6 shows the computation cost of parameter estimation and classification process for the next 8 and 10 samples for the case when 40 delayed samples (previous 8 seconds data) are used. Table shows that we can have a time gain of up to 1350 msec, which enables a mobile node to take handover decision with good reliability 1350 msec before the actual link down. This time is sufficient

for establishing link layer and network layer connectivity as well as exchanging appropriate mobility management control messages.

5.2   Runtime Memory Requirement for Prediction Process

Predictions using trained TDNN and the subsequent classification using trained FFNN involves simply the matrix multiplications. Now we determine the memory requirement for real time predictions using trained TDNN and FFNN for our model using following relationship:

$Number\ of\ variables\ required = v + a \times (v + 1) + b \times (a + 1) + c \times (b + 1)$
$Or \qquad \cdots \qquad\qquad = v \times (1 + a) + a \times (1 + b) + b \times (1 + c) + c$

Here v is the number of input parameters, a is the number of neurons in layer-1, b is the number of neurons in layer-2, & c is the number of outputs.

For prediction using TDNN: After experimental results, we recommend 40 delayed samples for TDNN and each sample is 4-tuple, thus $v = 4 \times 40 = 160$
a = 5, b = 3 and c = number of parameters in predicted sample = 4
Number of variables required $= 160 + 5(161) + 3(6) + 4(4) = 999 \times 4 = 3996\ bytes$

For classification using FFNN: v = number of input parameters to FFNN = numbers of output parameter of TDNN = 4
a = 5, b = 3 and c = connectivity status = 1
Number of variables required $= 4 + 5(5) + 3(6) + 1(4) = 51 \times 4 = 204\ bytes$

Hence total minimum memory required for our prediction model is 3996 + 204 = 4200 bytes. Considering the available memory in current mobile nodes i.e. laptops and smart phones this memory requirement is not too large, hence making our model feasible for mobile devices.

6   Conclusion

In this paper, we proposed an intelligent prediction model for generating IEEE 802.21-2008 Media Independent Handover (MIH) Link Going Down (LGD) trigger. We used Time Delay Neural Network (TDNN) for estimating the future values of link parameters and Feed-Forward Neural Network (FFNN) to identify whether link connectivity will be maintained or lost in near future. We performed experiments by capturing link parameters data from IEEE 802.11 interface along with the upper layer connectivity status. Then we trained neural networks on training data set. After training, we validated the performance of link parameter estimation and LGD prediction using test data set. Using the link parameters data of previous 8 to 10 seconds and then predicting for the next 2 second's link conditions, we have been able to predict the LGD event with high accuracy. We also analyzed the computation cost and memory requirement cost for the implementation of our intelligent prediction model in mobile devices.

References

1. IEEE 802.21-2008, Media Independent Handover Services, IEEE Std. 802.21-2008 (2009)
2. IEEE 802.11u Draft 5.0, Amendment 7, Interworking with External Networks, IEEE P802.11u/D5.0 (2009)
3. Yoo, S., Cypher, D., Golmie, N.: Timely Effective Handover Mechanism in Heterogeneous Wireless Networks. In: Proceedings of the Springer Wireless Personal Communications (2008), doi:10.1007S/11277-008-9633-8
4. Lim, W.S., et al.: Implementation and Performance Study of IEEE 802.21 in Integrated IEEE 802.11/802.16e Networks. Computer Communications 32(1), 134–143 (2009)
5. Yoo, S.J., Cypher, D., Golmie, N.: LMS Predictive Link Triggering for seamless handovers in heterogeneous wireless networks. In: Proceedings of MILCOM 2007, Orlando Florida, pp. 28–30 (2007)
6. Kassar, M.B., et al.: An Overview of Vertical Handover Decision Strategies in Heterogeneous Wireless Networks. Computer Communications Journal 31, 2607–2620 (2008)
7. Saha, D., et al.: Mobility Support in IP: A Survey of Related Protocols. IEEE Network 18(6), 34–40 (2004)
8. Atiquzzaman, M., Reaz, A.S.: Survey and Classification of Transport Layer Mobility Management Schemes. In: Proceedings of IEEE 16th International Symposium on Personal Indoor and Mobile Radio Communications (PIMRC), vol. 4, pp. 2109–2115 (2005)
9. Yousaf, M., Qayyum, A.: On End-to-End Mobility Management in 4G Heterogeneous Wireless Networks. In: IEEE International Networking and Communications Conference (INCC 2008), pp. 118–123 (2008)
10. Fausett, L.: Fundamentals of Neural Networks, Architectures, Algorithms and Applications. Prentice-Hall Inc., Upper Saddle River (1994)
11. Engelbrecht, P.: Computational Intelligence: An Introduction. John Wiley & Sons Ltd., The Artium (2002)

Cloud@Home: Bridging the Gap between Volunteer and Cloud Computing

Vincenzo D. Cunsolo, Salvatore Distefano, Antonio Puliafito, and Marco Scarpa

University of Messina,
Contrada di Dio, S. Agata, 98166 Messina, Italy
{vdcunsolo,sdistefano,apuliafito,mscarpa}@unime.it

Abstract. The ideas of using geographically distributed resources in a secure way (*Network-Internet/Grid computing*), providing self-management capabilities (*Autonomic computing*), quantifying and billing computing costs (*Utility computing*), in order to perform specific modular applications (*web services*), have been grouped altogether into the concept of *Cloud computing*.

Only commercial Cloud solutions have been implemented so far, offering computing resources and (web) services for renting. Some interesting projects, such as Nimbus, OpenNEbula, Reservoir, work on Cloud. One of their aims is to provide a Cloud infrastructure able to provide and share resources and services for scientific purposes. The encouraging results of Volunteer computing projects such as SETI@home and FOLDING@home and the great flexibility and power of the emergent Cloud technology, suggested us to address our research efforts towards a combined new computing paradigm we named *Cloud@Home*, merging the benefits and overcoming the weaknesses of both the original computing paradigms.

In this paper we present the Cloud@Home paradigm, describing its contribution to the actual state of the art on the topic of distributed and Cloud computing. We thus detail the functional architecture and the core structure implementing such a new paradigm, demonstrating how it is really possible to build up a Cloud@Home infrastructure.

1 Introduction and Motivation

Cloud computing is derived from the *service-centric perspective* that is quickly and widely spreading on the IT world. From this perspective, all capabilities and resources of a Cloud (usually geographically distributed) are provided to users *as a service*, to be accessed through the Internet without any specific knowledge of, expertise with, or control over the underlying technology infrastructure that supports them.

Cloud computing is strictly related to *service oriented science* [1], *service computing* [2] and *IT as a service* (ITAAS) [3], a generic term that includes: platform AAS, software AAS, infrastructure AAS, data AAS , security AAS, business process management AAS and so on.

It offers a user-centric interface that acts as a unique, user friendly, point of access for users' needs and requirements. Moreover, Cloud computing provides *on-demand service provision*, *QoS guaranteed offer*, and *autonomous system* for managing hardware, software and data transparently to users [4].

In order to achieve such goals it is necessary to implement a level of abstraction of physical resources, uniforming their interfaces and providing means for their management, adaptively to user requirements. This is done through *virtualizations*, *service mashups* (Web 2.0) and *service oriented architectures* (SOA). The development and the success of Cloud computing is due to the maturity reached by such technologies. These factors made realistic the L. Kleinrock outlook of computing as the 5th utility [5], like gas, water, electricity and telephone.

Virtualization allows to execute a software version of a hardware machine into a host system, in an isolated way. It "homogenizes" resources: problems of compatibility are overcome by providing heterogeneous hosts of a distributed computing environment (the Cloud) with the same virtual machine.

The Web 2.0 [6] provides an interesting way to interface Cloud services, implementing service mashup. It is mainly based on an evolution of JavaScript with improved language constructs (late binding, clousers, lambda functions, etc) and AJAX interactions.

The *Service Oriented Architecture* (SOA) is a paradigm that defines standard interfaces and protocols that allow developers to encapsulate information tools as services that clients can access without knowledge of, or control over, their internal workings [1].

A great interest on Cloud computing has been manifested from both academic and private research centers, and numerous projects from industry and academia have been proposed. In commercial contexts, among the others we highlight: Amazon Elastic Compute Cloud, IBMs Blue Cloud, Sun Microsystems Network.com, Microsoft Azure Services Platform, Dell Cloud computing solutions. There are also several scientific activities, such as: Reservoir [7], Nimbus-Stratus-Wispy-Kupa [8] and OpenNEbula [9]. All of them support and provide an on-demand computing paradigm, in the sense that a user submits his/her requests to the Cloud that remotely, in a distributed fashion, processes them and gives back the results. This client-server model well fits aims and scopes of commercial Clouds: the business. But, on the other hand, it represents a restriction for scientific Clouds, that have a view closer to *Volunteer computing*.

Volunteer computing (also called *Peer-to-Peer computing*, *Global computing* or *Public computing*) uses computers volunteered by their owners, as a source of computing power and storage to provide distributed scientific computing [10]. It is behind the *"@home"* philosophy of sharing network connected resources for supporting distributed computing.

We believe the Cloud computing paradigm is applicable also at lower scales, from the single contributing user, that shares his/her desktop, to research groups, public administrations, social communities, small and medium enterprises, which make available their distributed computing resources to the Cloud. Both free sharing and pay-per-use models can be easily adopted in such scenarios.

From the utility point of view, the rise of the "techno-utility complex" and the corresponding increase of computing resources demand, in some cases growing dramatically faster than Moore's Law as predicted by the Sun CTO Greg Papadopoulos in the *red shift theory* for IT [11], could bring, in a close future, towards an *oligarchy*, a lobby or a trust of few big companies controlling the whole computing resources market. To avoid such pessimistic but achievable scenario, we suggest to address the problem in a

different way: instead of building costly private *data centers*, that the Google CEO Eric Schmidt likes to compare to the prohibitively expensive cyclotrons [12], we propose a more "democratic" form of Cloud computing, in which the computing resources of single users accessing the Cloud can be shared with the others, in order to contribute to the elaboration of complex problems.

Since this paradigm is very similar to the Volunteer computing one, it can be named *Cloud@Home*. Both hardware and software compatibility limitations and restrictions of Volunteer computing can be solved in Cloud computing environments, allowing to share both hardware and software resources or *services*.

The Cloud@Home paradigm could be also applied to commercial Clouds, establishing an *open computing-utility market* where users can both buy and sell their services. Since the computing power can be described by a "long-tailed" distribution, in which a high-amplitude population (Cloud providers and commercial data centers) is followed by a low-amplitude population (small data centers and private users) which gradually "tails off" asymptotically, Cloud@Home can catch the *Long Tail* effect [13], providing similar or higher computing capabilities than commercial providers' data centers, by grouping small computing resources from many single contributors.

In the following we demonstrate how it is possible to make real all these aims through the Cloud@Home paradigm. Thus, in section 2 we describe the functional architecture of the Cloud@Home infrastructure, and in section 3 we characterize the blocks implementing the functions previously identified into the Cloud@Home core structure. With section 4 we conclude the paper recapitulating our work and discussing about challenges and future work.

2   Cloud@Home Overview

A possible Cloud@Home architecture is shown in Fig. 1, identifying three hierarchical layers: *frontend*, *virtual* and *physical*. A user can interact with the Cloud through the *consumer host* after authenticating him/herself into the system. The main enhancement of Cloud@Home is that a host can be at the same time both contributing and consumer host, establishing a symbiotic mutual interaction with the Cloud.
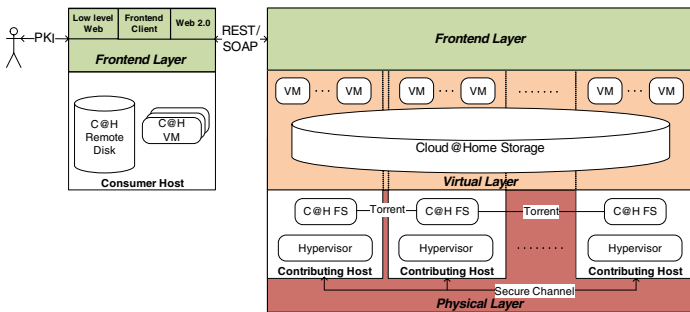
Fig. 1. Basic architecture of Cloud@Home

2.1   Frontend Layer

The Cloud@Home frontend layer is responsible for the resources and services management (enrolling, discovery, allocation, coordination, monitoring, scheduling, etc) from the global Cloud system's perspective. The frontend layer provides tools for translating end-user requirements into physical resources' demand, also considering QoS/SLA constraints, if specified by the user. Moreover, in commercial Clouds, it must be able to negotiate the QoS policy to be applied (SLA), therefore monitoring for its fulfillment and, in case of unsatisfactory results, adapting the computing workflow to such QoS requirements.

If the available Cloud's resources and services can not satisfy the requirements, the frontend layer provides mechanisms for requesting further resources and services to other Clouds, both open and/or commercial. In other words, the Cloud@Home frontend layer implements the interoperability among Clouds, also checking for services' reliability and availability. In order to improve reliability and availability of services and resources, especially if QoS policies and constraints have been specified, it is necessary introduce redundancy.

The frontend layer is split into two parts, as shown in Fig. 1: the server side, implementing the resources management and related problems, and the *light* client side, only providing mechanisms and tools for authenticating, accessing and interacting with the Cloud.

In a widely distributed system, globally spread around the world, the knowledge of resources' accesses and uses assumes great importance. To access and use the Cloud services a user first authenticates him/herself and then specifies whether he/she wants to make available his/her resources and services for sharing, or he/she only uses the Cloud resources for computing. The frontend layer provides means, tools and policies for managing users. The best mechanism to achieve secure authentications is the *Public Keys Infrastructure* (PKI) [14], better if combined with smartcard devices that, through a trusted certification authority, ensure the user identification.

Referring to Fig. 1, three alternative solutions can be offered by the frontend layer for accessing a Cloud: a) Cloud@Home frontend client, b) Web 2.0 user interface and c) low level Web interface (directly specifying REST or SOAP queries). These also provide mechanisms for customizing user applications by composing services (service mashup and SOA) and submitting own services.

2.2   Virtual Layer

The virtualization of physical resources offers end-users a homogeneous view of Cloud's services and resources. Two basic services are provided by the virtual layer to the frontend layer and, consequently, to the end-user: *execution* and *storage* services. The execution service is the tool provided by the virtual layer for creating and managing virtual machines. A user, sharing his/her resources within a Cloud@Home, allows the other users of the Cloud to execute and manage virtual machines locally at his/her node, according to policies and constraints negotiated and monitored at the frontend layer. In this way, a Cloud of virtual machine's executors is established, where virtual machines can migrate or can be replicated in order to achieve reliability, availability and

QoS targets. As shown in Fig. 1, from the end-user point of view an execution Cloud is seen as a set of virtual machines available and ready-to-use. The virtual machines' *isolation* implements protection and therefore security. This security is ensured by the hypervisor that runs the virtual machine's code in an isolated scope, similarly to a sandbox, without affecting the local host environment. The storage service implements a storage system distributed across the storage hardware resources composing the Cloud, highly independent of them since data and files are replicated according to QoS policies and requirements to be satisfied. From the end-user point of view, a storage Cloud appears as a locally mounted remote disk, similarly to a Network File System or a Network Storage. The tools, libraries and API for interfacing end-user and storage Clouds are provided to user by the frontend client, but are implemented at virtual and physical layers.

2.3   Physical Layer

The physical layer is composed of a *"cloud"* of generic nodes and/or devices geographically distributed across the Internet. They provide to the upper virtual layer both physical resources for implementing execution and storage services and mechanisms and tools for locally managing such resources. Cloud@Home negotiates with users that want to join a Cloud about his/her contribution. This mechanism involves the physical layer that provides tools for reserving physical execution and/or storage resources for the Cloud, and monitors these resources, such that constraints, requirements and policies thus specified are not violated. This ensures reliability and availability of physical resources, avoiding to overload the local system and therefore reducing the risk of crashes.

To implement the execution service in a generic device or to enroll it into an execution Cloud, the device must have a hypervisor ready to allocate and run virtual machines, as shown in Fig. 1. If a storage service is installed into the device, a portion of the local storage system must be dedicated for hosting the Cloud data. In such cases, the Cloud@Home file system has to be installed into the devices' shared storage space.

At physical layer it is necessary to implement data security (integrity and confidentiality) also ensuring that stored data cannot be accessed by who physically hosts them. We propose to combine the inviolability of the asymmetric cryptography and the performance of the symmetric one: data are firstly encrypted by the symmetric key, and then stored into the selected host with the symmetric key encrypted by the user private key. Since the data stored in a Cloud@Home storage are encrypted, a higher performance protocol, such as BitTorrent [15] can be used for transferring data. A secure channel such as SSH, TLS, IPSEC and the like is required for sending and receiving non-encrypted messages and data to/from remote hosts.

3   Cloud@Home Core Architecture

Once the functional architecture of Cloud@Home has been introduced, in Fig. 2 we characterize the blocks implementing the functions thus identified, the core structure of the overall system implementing the Cloud@Home server-side, subdivided into *management* and *resource subsystems*.
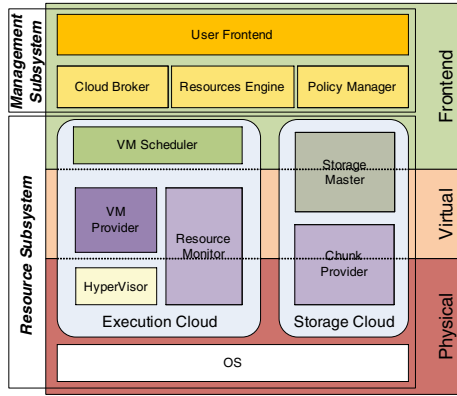
Fig. 2. Cloud@home Core Structure Organization

3.1   Management Subsystem

In order to enroll and manage the distributed resources and services of a Cloud, providing a unique point of access them, it is necessary to adopt a centralized approach that is implemented by the management subsystem. It is composed of four parts: the *user frontend* (UF), the *Cloud broker*, the *resource engine* and the *policy manager*.

The user frontend provides tools for Cloud@Home-User interactions. Incoming requests are transferred to the blocks composing the underlying layer (resource engine, Cloud broker and policy manager) for processing. An important task carried out by the user frontend is the *Clouds interoperability*, implemented by point-to-point connecting the user frontend of the Clouds wishing to interoperate. In case one of the Clouds does not have the Cloud@Home core structure of Fig. 2, it is necessary to translate the requests between Cloud@Home and foreign Clouds formats, task delegated by the user frontend to the Cloud broker. The Cloud broker collects and manages information about the available Clouds and the services they provide (both *functional* and *non-functional* parameters, such as QoS, costs, reliability, *request formats' specifications* for Cloud@Home-foreign Clouds translations, etc).

The policy manager provides and implements the Cloud's access facilities. This task falls into the security scope of identification, authentication, identity and permission management. To achieve this target, the policy manager uses an infrastructure based on PKI, smartcard devices and Certification Authority. The policy manager also manages the information about users' QoS policies and requirements.

The resource engine is the hearth of Cloud@Home. It is responsible for the resources' management, the equivalent of a Grid *resource broker* in a broader Cloud environment. To meet this goal, the resource engine applies a hierarchical policy. It operates at higher level, in a centralized way, indexing all the resources of the Cloud. Incoming requests are delegated to *VM schedulers* or *storage masters* that, in a distributed fashion, manage the computing or storage resources respectively, coordinated by the resource engine.
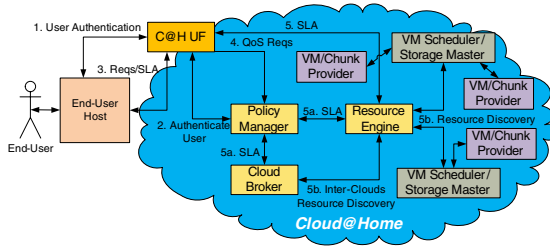
Fig. 3. Cloud@Home End-User Negotiation
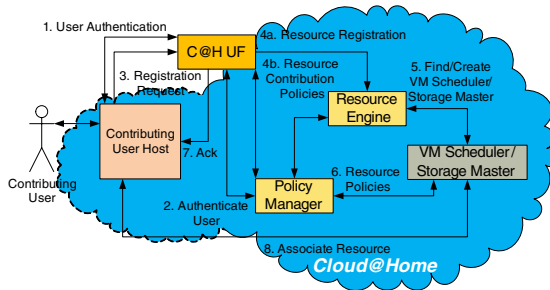
Fig. 4. Cloud@Home Resource Setup

In order to manage QoS policies and to perform the resources discovery, the resource engine collaborates with both Cloud broker and policy manager, as depicted in Fig. 3 showing the step-by-step interactions among such blocks. After authenticating into the system (steps 1 and 2), an end-user specifies his/her requirements (step 3), saved by the policy manager (step 4). Then, a negotiation between the two parties is triggered (step 5), iteratively interacting with the end-user till an agreement is met (SLA). This task is split into two parallel subtasks: the former (step 5a), performed by the policy manager under the supervision of the resource engine, estimates and evaluates the QoS requirements of the request; the latter (step 5b), performed by the resource engine, discovers resources and services to be used. Both subtasks can require the collaboration of the Cloud broker, that looks for other Clouds able to provide resources and services to satisfy SLA/QoS requirements.

Fig. 4 shows the interaction between a contributing user, that wants to provide his/her resources to a Cloud, and the Cloud@Home management system. A user, authenticated by the Cloud's policy manager (steps 1 and 2), sends a request for registering resources and services to the user frontend (step 3), also specifying policies for using them. It sorts the request at the resource engine (step 4a), and constraints and policies at the policy manager (step 4b). After that, the resource engine searches for a VM scheduler or a storage master to which such resources/services have to be assigned (step 5), collaborating with the policy manager. It can also create a new VM scheduler/storage master if the search results obtained do not satisfy the requirements.

Once the scheduler/master is identified the policy manager contacts it for exchanging policies and specifications of the resources. Then the resource engine sends the acknowledgement and the scheduler/master reference to the contributing host (step 7), that signals its availability and actual status (step 8).

3.2 Resource Subsystem

The resource subsystem contains all the blocks implementing the local and distributed management functionalities of Cloud@Home. This subsystem can be logically split into two parts according to the service offered: the *execution Cloud* and the *storage Cloud*. The management subsystem merges them providing a unique Cloud that can offer both execution and/or storage services. The execution Cloud provides tools for managing virtual machines according to users' requests and requirements coming from the management subsystem. It is composed of four blocks: *VM scheduler*, *VM provider*, *resource monitor* and *hypervisor*.

The VM Scheduler is a peripheral resource broker of the Cloud@Home infrastructure, to which the resource engine delegates the management of computing/execution resources and services of the Cloud. It establishes what, where, when and how allocate a VM, moreover it is responsible for moving and managing VM services. From the end-user point of view a VM is allocated somewhere on the Cloud, therefore its migration is transparent for the end-user that is not aware of any VM migration mechanism. The association between resources and scheduler is made locally, as shown in Fig. 4. Since a scheduler can become a bottleneck if the system grows, to avoid the congestion further decentralized and distributed scheduling algorithms (hierarchical, with replication, autonomic, etc) can be implemented.

The VM provider, the resource monitor and the hypervisor are responsible for managing a VM locally to a physical resource. A VM provider exports functions for allocating, managing, migrating and destroying a virtual machine on the corresponding host. The resource monitor allows to take under control the local computing resources, according to requirements and constraints negotiated in the setup phase with the contributing user. If during a virtual machine execution local resources crashes or becomes insufficient to keep running the virtual machine, the resource monitor asks the scheduler to migrate the VM elsewhere.
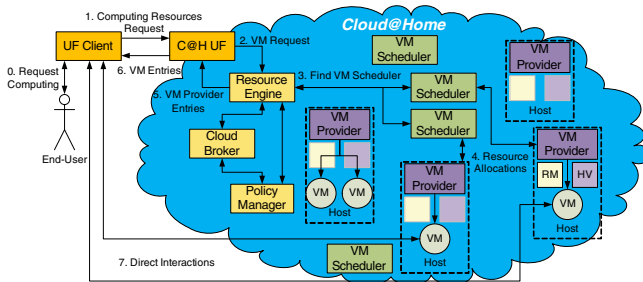
Fig. 5. User Computing Request Processing
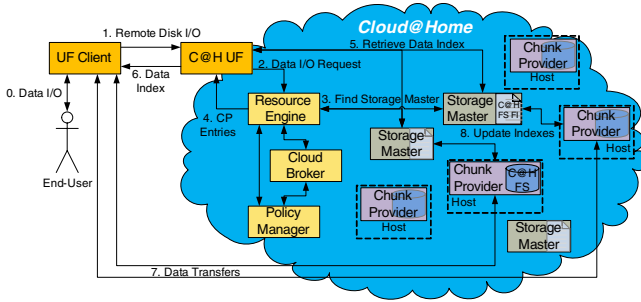
Fig. 6. User Remote Disk I/O Request Processing

Fig. 5 depicts the process of requesting and allocating computing resources in Cloud@Home environments. The overall process is coordinated by the resource engine that estimates requests and requirements submitted by the end-user (steps 0,1 and 2), and therefore evaluates and selects proper schedulers (step 3). Each of such schedulers, in its turn, allocates the physical resources that will host the VM (step 4). The access points of such resources are then fed back to the end-user (steps 5 and 6), and consequently the two parties get connected and can directly interact (step 7).

In order to implement the storage Cloud, we specify the *Cloud@Home file system* (FS) splitting data and files into *chunks* of fixed or variable size, depending on the storage resource available. The architecture of such file system is hierarchical: data chunks are physically stored on *chunk providers* and corresponding *storage masters* index the chunks through specific *file indexes* (FI). The storage master directly interfaces with the resource engine to discover the resources storing data. To improve the storage Cloud reliability, storage masters have to be replicated. Moreover, a chunk provider can be associated to more than one storage master. In order to avoid a storage master becoming a bottleneck, once the chunk providers have been located, data transfers are implemented by directly connecting end-users and chunk providers.

Chunk providers physically store the data, that, as introduced above, are encrypted in order to achieve the confidentiality goal. Data reliability can be improved by replicating data chunks and chunk providers, consequently updating the corresponding storage masters. In this way, a corrupted data chunk can be automatically recovered and restored through the storage masters, without involving the end-user. Similarly to the execution Cloud, the storage Cloud can be implemented as shown in Fig. 6: an end-user data I/O request to the Cloud (steps 0 and 1) is delivered to the resource engine (step 2), that locates the storage masters managing the chunk providers where data are stored or will be stored (step 3), and feeds back the list of chunk providers and data indexes to the end-user (step 4, 5 and 6). In this way the end user can directly interact with the assigned chunk providers storing his/her data (step 7).

4   Conclusions

In this paper we proposed an innovative computing paradigm merging volunteer contributing and Cloud approaches into Cloud@Home. This proposal represents a solution

for building Clouds, starting from heterogeneous and independent nodes, not specifically conceived for this purpose.

In this way Cloud@Home opens the Cloud computing world to scientific and academic research centers, as well as to communities or single users: anyone can voluntarily support projects by sharing his/her resources. On the other hand, it opens the utility computing market to the single user that wants to sell his/her computing resources. To realize this broader vision, several issues must be adequately taken into account: reliability, security, portability of resources and services, interoperability among Clouds, QoS/SLA and business models and policies.

References

1. Foster, I.: Service-oriented science. Science 308(5723) (May 2005)
2. Zhang, L.J.L.: EIC Editorial: Introduction to the Body of Knowledge Areas of Services Computing. IEEE Transactions on Services Computing 1(2), 62–74 (2008)
3. Foster, I., Tuecke, S.: Describing the Elephant: The Different Faces of IT as Service. Queue 3(6), 26–29 (2005)
4. Wang, L., Tao, J., Kunze, M., Castellanos, A.C., Kramer, D., Karl, W.: Scientific Cloud Computing: Early Definition and Experience. In: HPCC 2008, pp. 825–830 (2008)
5. Kleinrock, L.: A vision for the internet. ST Journal of Research 2(1), 4–5 (2005)
6. Tim O'Reilly: What is WEB 2.0 (September 2005), http://www.oreillynet.com/pub/a/oreilly/tim/news/2005/09/30/what-is-web-20.html
7. Reservoir Consortium: Reservoir Project (2009), http://www-03.ibm.com/press/us/en/pressrelease/23448.wss/
8. University of Chicago-University of Florida-Purdue University-Masaryk University: Nimbus-Stratus-Wispy-Kupa Projects (January 2009), http://workspace.globus.org/clouds/nimbus.html/, http://www.acis.ufl.edu/vws/, http://www.rcac.purdue.edu/teragrid/resources/#wispy, http://meta.cesnet.cz/cms/opencms/en/docs/clouds
9. Distributed Systems Architecture Research Group: OpenNEbula Project [URL]. Universidad Complutense de Madrid (2009), http://www.opennebula.org/
10. Anderson, D.P., Fedak, G.: The computational and storage potential of volunteer computing. In: CCGRID 2006, pp. 73–80 (2006)
11. Martin, R.: The Red Shift Theory. InformationWeek (August 20, 2007), http://www.informationweek.com/news/hardware/showArticle.jhtml?articleID=201800873
12. Baker, S.: Google and the Wisdom of Clouds. BusinessWeek (December 24, 2008), http://www.businessweek.com/magazine/content/07_52/b4064048925836.htm
13. Anderson, C.: The Long Tail: How Endless Choice Is Creating Unlimited Demand. Random House Business Books (July 2006)
14. Tuecke, S., Welch, V., Engert, D., Pearlman, L., Thompson, M.: Internet X.509 Public Key Infrastructure (PKI) Proxy Certificate Profile. RFC 3820 (Proposed Standard) (June 2004)
15. Cohen, B.: The BitTorrent Protocol Specification (2008), http://www.bittorrent.org/beps/bep_0003.html

Analysis and Improvement of an ID-Based Anonymous Signcryption Model

Mingwu Zhang[1], Yusheng Zhong[1], Bo Yang[1,⋆] and Wenzheng Zhang[2]

[1] Department of Computer Science and Engineering,
College of Informatics,
South China Agricultural University, Guangzhou, 510642, China
[2] National Laboratory for Modern Communications, Chengdu, 610041, China
byang@scau.edu.cn

Abstract. Ring signcryption, a cryptographic primitive to protect security and privacy, is an encryption and authentication scheme in a single logical step which allows a user to anonymously signcrypt a plaintext on behalf of a group of users that decrypter cannot know who is the actual signcrypter, which can be used to protect nodes or participants privacy in ubiquitous environments such RFID, WSN, Ad hoc etc. In 2009, Zhang, Gao, Chen and Geng proposed a novel anonymous signcryption scheme(denoted as the ZGCG scheme) which is more efficient in computational cost and ciphertext length than the related schemes. In this paper, however, we show that the ZGCG scheme has not anonymity secure for the receiver, and then we propose an improved anonymous signcryption scheme that remedies the weakness of the ZGCG scheme. Our proposed scheme satisfies the semantic security, unforgeability, signcrypter identity's ambiguity, and public authenticity. We also give the formal security proof in the random oracle model.

Keywords: anonymous signcryption, bilinear pairings, unforgeability, anonymity.

1 Introduction

The concept of signcryption was first proposed by Zheng [1] that is a cryptographic primitive to perform signature and encryption simultaneously, at a lower computational costs and communication overheads than the traditional the signature-then-encryption approach, which can be used to protect nodes or participants privacy in lighter-weight ubiquitous environment requirements such RFID, WSN, and Ad hoc etc.

Followed by the first constructions given in [1], a number of new schemes and improvements have been proposed [3,2,4,5]. Recently, a formal security proof model for signcryption scheme is formalized in [6]. To achieve simple and safe non-repudiation procedure, Bao and Deng [2] introduced a signcryption scheme

---

⋆ Corresponding author.

that can be verified by a sender's public key. A distinguishing property of ID-based cryptography is that a user's public key can be any binary string that can identify the user's identity, while private keys can be generated by the trusted Private Key Generator(PKG). Several ID-based signcryption schemes have been proposed [8,11,9,7,10].

Ring signcryption [15,13,12] is an important method to realize the signcrypter identities' ambiguity that motivated by ring signature [14]. The receiver only knows that the message is produced by one member of a designated group, but he cannot know more information about actual signcrypter's identity. To obtain that the signcrypter can authenticate the ciphertext was produced by himself, an authenticable anonymous signcryption was proposed in [15] which extend an authentication algorithm to let sender prove that the ciphertext is produced by himslef. In [16], Zhang et. al proposed a novel anonymous signcryption scheme(we called ZGCG scheme) that is more efficient. In this paper, however, we show that the ZGCG scheme is not anonymous for the decrypter nor public authenticable or verifiable for the third party. Furthermore, we propose an improved scheme that remedies the weakness of the ZGCG scheme. The improved scheme has the security notions such as confidentiality, unforgeability, signcrypter identities ambiguity, and public authenticity.

RoadMap. The rest of this paper is organized as follows: Section 2 gives a formal ID-based anonymous signcryption scheme and its security notions. The ZGCG scheme and its security analysis is described in 3. An improved scheme is proposed in section 4 and its security is given in section 5. At last the conclusion is drawn in section 6.

2   Formal Model of ID-Based Anonymous Signcryption Scheme

In this section, we will describe the outline and the security requirements of ID-based anonymous signcryption scheme. An ID-based anonymous signcryption scheme consists of four algorithms: Setup, KeyExtract, AnonySigncrypt, and Unsigncrypt.

- Setup: Take an input 1^k, where k is a security parameter, the algorithm generates a master key s and the system's public parameters $params$, which include a description of a finite message space together with a description of a ciphtertext space.
- KeyExtract: Given an identity string $ID \in \{0,1\}^*$, and system master key s, this algorithm outputs the private key associated with the ID, denoted by D_{ID}.
- AnonySigncrypt: If a user A identified by ID_A wishes to send a message m to B identified by ID_B ,this algorithm selects a group of n users' identities by L $= \bigcup ID_i (1 \leq i \leq n)$ including the actual signcrypter ID_A , and outputs the ciphertext C.

– UNSIGNCRYPT: When user B receives the cipertext C, this algorithm takes the ciphertext C, Ł, and B's private key D_B as input, and outputs plaintext m when unsigncryption is successful, otherwise it outputs $\perp$.

The algorithms must satisfy the standard consistency constraint of ID-based signcryption scheme as following

$$C = \text{ANONYSIGNCRYPT}(m, \text{Ł}, D_A, ID_B) \Rightarrow \text{UNSIGNCRYPT}(C, \text{Ł}, D_B) = m$$

2.1 Security Notions

The security of ID-based signcryption scheme was first defined by Malone-Lee [4,10] that satisfies *indistinguishable against adaptive chosen ciphertext attacks* and *unforgeability against adaptive chosen message attacks*. The anonymous signcryption scheme extends the security about *ciphertext anonymity against adaptive chosen ciphertext attacks*, and *public authenticity* and *public verifiability*.

Definition 1. (*Confidentiality*) *An ID-based anonymous signcryption scheme is indistinguishabe against adaptive chosen ciphertext attacks (IND-IDAS-CCA2) if no polynomially bounded adversary has a non-negligible advantage in IDAS game.*

We define the IDAS game played by a challenger $\mathcal{C}$ and an adversary $\mathcal{A}$ as following:

– Inital: The challenger $\mathcal{B}$ runs SETUP algorithm with security parameter k, keeps master key s and gives the *params* to the adversary.
– Phasei-I: The adversary $\mathcal{A}$ performs a series of following queries in an adaptive fashion:
  KEYEXTRACT queries: $\mathcal{A}$ produces an identity ID, $\mathcal{C}$ computes the private key $D_{ID} = \text{KEYEXTRACT}(ID)$ to respond to $\mathcal{A}$.
  ANONYSIGNCRYPT queries: $\mathcal{A}$ generates a group of n identities Ł $= \{ID_i\}$ (i=1,...,n), a plaintext m and a designated receiver ID_B. $\mathcal{B}$ randomly chooses a user $\mathcal{U}_i \in \{ID_i\}$, computes $D_i = \text{KEYEXTRACT}(ID_i)$ and generates ciphertext $C = \text{ANONYSIGNCRYPT}(m, \text{Ł}, D_i, ID_B)$ and sends C to $\mathcal{A}$.
  UNSIGNCRYPT queries: $\mathcal{A}$ chooses a group of identities Ł $= \{ID_i\}$ (i=1,..,n), a receiver identity ID_r and a ciphertext C. $\mathcal{B}$ first generates the privacy key $D_r = \text{KEYEXTRACT}(ID_r)$, computes $\text{UNSIGNCRYPT}(\{ID_i\}, C, D_r)$, then returns the result to $\mathcal{A}$. This result may be the $\perp$ if C is an invalid ciphertext for ID_r.
– Challenge: $\mathcal{A}$ chooses two plaintexts $m_0, m_1 \in \mathcal{M}$, a group of identities Ł$^* = ID_i^*$ (i=1,...,n), and a designated receiver ID_B^* on which he wishes to be challenged. The challenger $\mathcal{B}$ picks a random $b \in \{0, 1\}$ and computes $C^* = \text{ANONYSIGNCRYPT}(m_b, \text{Ł}^*, ID_B^*)$ and sends C^* to $\mathcal{A}$.
– Phase-II: $\mathcal{A}$ can ask a series number of queries adaptively again as in the first stage with the restriction that he cannot make the KEYEXTRACT query on group Ł* member nor ID_B^*, and he cannot make the UNSIGNCRYPT query on ciphertext C^*.
– Output: Finally, $\mathcal{A}$ outputs a bit b' and wins the game if $b' = b$.

The adversary $\mathcal{A}$'s advantage is defined as $Adv(\mathcal{A}) = |2Pr[b' = b] - 1|$.

Definition 2. *(Anonymity) An ID-based anonymous signcryption scheme is un-conditional anonymous if for any group of n members with identities $L = \bigcup ID_i(1 \le i \le n)$, any adversary cannot identify the actual signcrypter with probability better than random guess's.*

That is, $\mathcal{A}$ outputs the identity of actual signcrypter with probability $1/n$ if he is not the member of Ł, and with probability $1/(n-1)$ if he is the member of Ł.

Definition 3. *(Unforgeability) An ID-based anonymous signcryption scheme is existentially unforgeable against adaptive chosen-message attacks and adaptive chosen-identity attacks(EUF-IDAS-CMIA) if no polynomially bounded adversary has a non-negligible advantage in the following game:*

- The challenger $\mathcal{C}$ runs the SETUP algorithm with a security parameter k and gives the public parameters to adversary $\mathcal{A}$.
- $\mathcal{A}$ performs a polynomially bounded phase-I queries in IDAS game.
- Finally, $\mathcal{A}$ outputs a cipertext C^* and wins the game if: (1)The C^* is a valid cipertext under the group users Ł and receiver ID^* such that the result of the UNSIGNCRYPT(C^*, L^*, ID^*) is not the $\perp$ symbol; (2)C^* was not produced by ANONYSIGNCRYPT oracle; (3)Group Ł identities were not performed KEYEXTRACT queries.

Definition 4. (Public verifiability) *An ID-based anonymous signcryption scheme is publicly verifiable if given a plaintext m and ciphertext C, and possibly some additional information provided by the receiver, anyone can verify that C is a valid message of the sender without knowing the receiver's private key.*

Definition 5. (Public authenticity) *An ID-based anonymous signcryption scheme is publicly authenticable if anyone can verify that the validity and the origin of the ciphertext without knowing the content of the message and getting any help from the receiver.*

3   Analysis of the ZGCG Scheme

In this section, we review the ZGCG scheme [16], and demonstrate that the ZGCG scheme is neither anonymous in sender identity for the decrypter nor public authenticable or verifiable for a third party.

3.1   Review of the ZGCG Scheme

The ZGXCG scheme is described as follow four algorithms.

1. SETUP Given a security k, the PKG chooses bilinear map groups (G_1, G_2) of order $q > 2^k$, bilinear map $e : \mathbb{G}_1 \times \mathbb{G}_1 \to \mathbb{G}_2$. Let P be a generator of G_1. It randomly chooses a master key $s \in Z_q^*$ and computes $P_{pub} = sP$ as the corresponding public key. Next, PKG chooses cryptography hash functions: $H_1 : \{0,1\}^* \to \mathbb{G}_1^*$, $H_2 : \mathbb{G}_2 \to \{0,1\}^n$, $H_3 : \{0,1\}^n \times G_2 \to Z_q$, $H_4 : \{0,1\}^* \times G_1 \times \{0,1\}^* \to G_1$. The system public parameters are
   $params = \{G_1, G_2, e, P, P_{pub}, H_1, H_2, H_3, H_4\}$.

2. KEYEXTRACT Given an identity ID, PKG computes user public key $Q_{ID} = H_1(ID)$ and corresponding secret key $D_{ID} = sQ_{ID}$.
3. ANONYSIGNCRYPT Let L $= \bigcup\{\mathcal{U}_i\}$ $(i=1,...,n)$ be a set of users including the actual signcrypter $ID_S(S \in [1,n])$. To signcrypt a message m on behalf of the group L to receiver ID_B, the signcrypter ID_S executes as follows:
   - For $i = 1, ..., n(i \neq s)$, randomly picks $x_i \in_R Z_q^*$ to computes $R_i = x_iP$;
   - Randomly chooses $x_s \in_R Z_q^*$ to compute $\omega = e(P_{pub}, \sum_{i=1}^n x_iQ_B)$, and sets $R_s = x_sP - \sum_{i=1,i\neq s}^n (H_3(R_i,\omega)Q_i + R_i)$;
   - Computes $c = H_2(\omega) \oplus m$, and $U = \sum_{i=1}^n x_iP$;
   - Computes $S = \sum_{i=1}^n x_iH_4(L,U,m) + x_sP_{pub} + H_3(R_s,\omega)D_s$;
   - Finally, outputs the ciphertext of message m as $C = (c, S, U, R_1, ..., R_n)$.
4. UNSIGNCRYPT Upon receiving the cipertext $C = (c, S, U, R_1, ..., R_n)$, ID_B uses his secret key D_B to recover and verify the massage as follows:
   - Computes $\omega = e(U, D_B)$, and $m = c \oplus \omega$;
   - Accepts the message iff the following equation holds:
   $e(S,P) = e(U, H_4(L,U,m))e(P_{pub}, \sum_{i=1}^n (R_i + H_3(R_i,\omega)Q_i))$

3.2 Security and Anonymity Analysis

We now show the ZGCG scheme is neither anonymous for the ciphertext unsigncrypter nor publicly authenticable or verifiable for a third party.

1.Anonymity analysis. Only S and R_s contain signcrypter ID_S and group L users identity information in ciphertext $C = (c, S, U, R_1, ..., R_n)$. We show that $S = \sum_{i=1}^n x_iH_4(L,U,m) + x_sP_{pub} + H_3(R_s,\omega)D_s$ leaks the actual signcrypter identity ID_S. We have
$e(S,P) = e(\sum_{i=1}^n x_iH_4(L,U,m) + x_sP_{pub} + H_3(R_s,\omega)D_s, P)$
$= e(\sum_{i=1}^n x_iH_4(L,U,m),P)e(x_sP_{pub} + H_3(R_s,\omega)D_s, P)$
$= e(U, H_4(L,U,m))e(x_sP + H_3(R_s,\omega)Q_s, P_{pub})$
The designcrypter ID_B can get the value ω by $e(U, D_B)$, and check whether the user $\mathcal{U}_j(1 \leq j \leq n)$ is the actual signcrypter by checking the following equation:
$e(x_jP + H_3(R_j,\omega)Q_j, P_{pub}) = e(S,P)e(U, H_4(L,U,m))^{-1}$
It is only the actual signcrypter ID_S who can pass through the above checking equation because it has the generated equation $R_s = x_sP - \sum_{i=1,i\neq s}^n (H_3(R_i,\omega)Q_i + R_i)$ in ANONYSIGNCRYPT algorithm. The other user in L cannot pass through the checking equation because his x_i is randomly picked from Z_q^*, and $R_i = x_iP$ is uniformly distributed in G_1.

2. Public verifiability and authenticity analysis. We show that the ZGCG scheme is neither public verifiable nor public authenticable. The verification equation $e(S,P) = e(U, H_4(L,U,m))e(P_{pub}, \sum_{i=1}^n (R_i + H_3(R_i,\omega)Q_i))$ needs receiver's decrypting agreeing key ω. If a third party want to check the equation, he must obtain the ω. It cannot provide public verifiability. If the decrypting receiver sends ω to a third party, it cannot provide the confidentiality in this scheme because the third party can decrypt the plaintext by $m = c \oplus \omega$. So the decrypting receiver cannot leak ω to any third party so that the scheme is not public authenticable.

4   Improved Anonymous Signcryption Scheme

To overcome the weakness of the ZGCG scheme, we improve the anonymous signcryption in this section.

1. SETUP Given a security k, the PKG chooses groups G_1 and G_2 of prime order $q > 2^k$ (with G_1 additive and G_2 multiplicative), bilinear map $e : \mathbb{G}_1 \times \mathbb{G}_1 \to \mathbb{G}_2$, a generator P of G_1. It randomly picks a master key $s \in Z_q^*$ and computes $P_{pub} = sP$. Next, PKG chooses hash functions: $H_1 : \{0,1\}^* \to \mathbb{G}_1^*$, $H_2 : \mathbb{G}_2 \to \{0,1\}^n$, $H_3 : \{0,1\}^l \times G_1 \to Z_q^*$, where n and l is plaintext and ciphertext length. The PKG keeps the master key s and public system parameters
   $params = \{G_1, G_2, e, P, P_{pub}, H_1, H_2, H_3, H_4\}$.
2. KEYEXTRACT Given an identity ID, PKG computes user public key $Q_{ID} = H_1(ID)$ and corresponding secret key $D_{ID} = sQ_{ID}$.
3. ANONYSIGNCRYPT Let $L = \bigcup\{\mathcal{U}_i\}$ $(i=1,...,n)$ be a set of users including the actual signcrypter ID_S. To signcrypt a message m on behalf of the group L to receiver ID_B, ID_S executes as follows:
   - For $i = 1, ..., n(i \neq s)$, randomly picks $x_i \in Z_q^*$ to computes $R_i = x_i P$;
   - Randomly picks $x_s \in Z_q^*$ to compute $\omega = e(P_{pub}, \sum_{i=1}^n x_i Q_B)$, and sets $c = H_2(\omega) \oplus m$;
   - Computes $R_s = x_s Q_s - \sum_{i=1, i\neq s}^n (H_3(c, R_i)Q_i + R_i)$, and $U = \sum x_i P$;
   - Computes $S = (x_s + H_3(c, R_s))D_s$;
   - Finally, outputs the ciphertext $C = (c, S, U, R_1, ..., R_n)$.
4. UNSIGNCRYPT Upon receiving the cipertext $C = (c, S, U, R_1, ..., R_n)$, ID_B uses his secret key D_B to recover and verify the massage as follows:
   - Checks whether $e(S, P) = e(P_{pub}, \sum_{i=1}^n (R_i + H_3(c, R_i)Q_i))$. If the equation holds, computes $\omega' = e(U, D_B)$, then recovers plaintext $m = c \oplus H_2(\omega')$; otherwise outputs $\perp$ as failure.

5   Correctness and Security Analyzes

5.1   Correctness

If the ciphertext C is generated in the way described as above algorithm, it has
$\omega' = e(U, D_B) = e(\sum_{i=1}^n x_i P, D_B) = e(sP, \sum_{i=1}^n x_i Q_B) = \omega$
Furthermore,
$$e(S, P) = e((x_s + H_3(c, R_i))D_s, P) = e((x_s Q_s + H_3(c, R_i)Q_s, P_{pub})$$
$$= e(\sum_{i=1, i\neq s}^n (H_3(c, R_i)Q_i + R_i) + R_s + H_3(c, R_i)Q_s, P_{pub})$$
$$= e(\sum_{i=1}^n (H_3(c, R_i)Q_i + R_i), P_{pub})$$

5.2   Security

Theorem 1. *(Confidentiality) In the random oracle model, if there is an IND-IDAS-CCA2 adversary $\mathcal{A}$ who can distinguish ciphertexts from the users set*

$\bigcup\{\mathcal{U}_i\}$ with an advantage ϵ when running in at most q_{H_i} queries to $H_i(1 \le i \le 3)$ hashes, at most q_E key extract queries, q_S signcryption queries, q_U unsigncryption queries. Then, there exists another algorithm $\mathcal{B}$ that can solve a random instance of the DBDH problem with an advantage $Adv(\mathcal{B}) \ge (\epsilon - \frac{q_U}{2^k})/q_{H_0}^2$.

Proof. Let the distinguisher $\mathcal{B}$ receives a random instance (P, aP, bP, cP, h) of the DBDH problem whose goal is to decide whether $h = e(P, P)^{abc}$ or not. In order to solve this problem, $\mathcal{B}$ runs $\mathcal{A}$ as a subroutine and act as $\mathcal{A}$'s challenger in the *IDAS* game. We assume that: (1)$\mathcal{A}$ will ask for $H_1(ID)$ before ID is used in any other queries; (2)$\mathcal{A}$ never makes an UNSIGNCRYPT query on a ciphertext obtained from the ANONYSIGNCRYPT oracle, and he can only make UNSIGNCRYPT queries for or guessed ciphertext.

Setup: At first, $\mathcal{B}$ sets $P_{pub} = cP$ as system public key and sends *params* to $\mathcal{A}$ after running the SETUP algorithm with parameter k. The value c is unknown to $\mathcal{B}$ and is used as the role of the PKG's master key.

Queries-I: For the key extraction and the signcryption/unsigncryption on the message m, $\mathcal{B}$ simulates the hash oracles(H_1, H_2, H_3), KEYEXTRACT oracle, ANONYSIGNCRYPT oracle, and UNSIGNCRYPT oracle. $\mathcal{A}$ can perform its queries adaptively in which every query may depend on the answers according to the previous ones.

H_1 queries. To response these queries, $\mathcal{B}$ maintains the list L_1 of tuples (ID, b). When $\mathcal{A}$ queries the oracle H_1, $\mathcal{B}$ chooses a random number $j \in \{1, ..., q_{H_1}\}$. At the jth H_1 query, $\mathcal{B}$ answers by $H_1(ID_j) = bP$, otherwise for queries $H_1(ID_e)$ with $e \ne j$, $\mathcal{B}$ chooses $b_e \in_R Z_q^*$, answers $H_1(ID_e) = b_e P$ and accords the pair (ID_e, b_e) in list L_1.

H_2, H_3 queries. When $\mathcal{A}$ asks queries on these hash values, $\mathcal{B}$ checks the corresponding lists. If an entry for the query is found, the same answer will be given to $\mathcal{A}$; otherwise, a randomly generated value will be used as an answer to $\mathcal{A}$, the query and the answer will then be recorded in the lists.

KEYEXTRACT queries. When $\mathcal{A}$ asks a query KEYEXTRACT(ID_i), $\mathcal{B}$ first searches the corresponding tuple (ID_i, b_i) in L_1. If $ID_i = ID_j$, $\mathcal{B}$ fails and stops. Otherwise, $\mathcal{B}$ computes the secret key $D_i = b_i P_{pub} = cb_i P$ and returns D_i to $\mathcal{A}$.

ANONYSIGNCRYPT queries. $\mathcal{A}$ can perform a ANONYSIGNCRYPT queriy for a plaintext m, a user group $L = \bigcup\{\mathcal{U}_i\}$ and a designated receiver with identity ID_B.

- $\mathcal{B}$ randomly chooses a user $\mathcal{U}_A \in L$ whose identity is $ID_A(ID_A \ne ID_j)$. $\mathcal{B}$ can compute $\mathcal{U}_A$'s secret key $D_A = b_A P_{pub}$ where b_A is in the corresponding tuple (ID_A, b_A) in L_1;
- $\mathcal{B}$ runs ANONYSIGNCRYPT (m, L, D_A, ID_B) to signcrypt a message m on behalf the group L using $\mathcal{U}_A$'s private key D_A;
- At last, $\mathcal{B}$ returns the result C to $\mathcal{A}$.

UNSIGNCRYPT queries. At any time, $\mathcal{A}$ can perform an UNSIGNCRYPT query for a ciphertext $C = (c, S, U, R_1, ..., R_n)$ between the group L and the receiver ID_B.

– If $ID_B = ID_j$, $\mathcal{B}$ always returns $\mathcal{A}$ that the ciphertext is invalid, because $\mathcal{B}$ does not know ID_j's secret key in KEYEXTRACT oracle. If this ciphertext is a valid one, the probability that $\mathcal{A}$ will find is no more than 2^{-k};

– If $ID_B \neq ID_j$, the equation $e(S, P) = e(P_{pub}, \sum_{i=1}^{n}(U_i + h_iQ_i))$ holds, $\mathcal{B}$ computes $\omega = e(U, D_B)$, $m' = c \oplus H_2(\omega)$ and returns m'. Otherwise $\mathcal{B}$ notifies $\mathcal{A}$ that the ciphertext is invalid with symbol $\perp$;

  For all q_U UNSIGNCRYPT queries, the probability to reject a valid ciphertext does not exceed $q_U/2^k$.

Challenge: After performing a series number of queries-I, $\mathcal{A}$ chooses two message $m_0^*, m_1^* \in \mathcal{M}$, n users $L^* = \{ID_1^*, ..., ID_n^*\}$ and a receiver ID_B^*. If $ID_B^* \neq ID_j$, $\mathcal{B}$ fails and stops. $\mathcal{B}$ chooses $b \in_R \{0,1\}$ and let $U^* = aP, \omega = h(h$ is $\mathcal{B}$ candidate for the DBDH problem). Then $\mathcal{B}$ signcrypts the message m_b^* as described in the ANONYSIGNCRYPT request and sends the ciphertext $C^* = (c^*, S^*, U^*, R_1^*, ..., R_n^*)$ to $\mathcal{A}$.

$\mathcal{A}$ performs a second series of queries just like in queries-I. In this stage, he can query neither the secret key of any user in the group L^* nor ID_B^*, and he cannot make the UNSIGNCRYPT oracle to the ciphertext C^*. At the end of the simulation, he produces a bit b' for which he believes the relation C^*=ANONYSIGNCRYPT (m_b^*, L^*, ID_j) holds and sends b' to $\mathcal{B}$. At this moment, if $b' = b$, $\mathcal{B}$ answers 1 as a result of DBDH problem because his selection h satisfying $h = e(U^*, D_j) = e(ap, cbP) = e(P, P)^{abc}$. If $b' \neq b$, $\mathcal{B}$ answers 0.

Success probability: Now we analyze $\mathcal{B}$'s success probability. The probability that $\mathcal{B}$ does not fail during the key extraction queries is greater than $1/q_{H_0}$. Furthermore, with a probability $1/q_{H_0}$, $\mathcal{A}$ chooses to be challenge on the ID_j to solve DBDH problem if $\mathcal{A}$ wins the IND-IDAS-CCA2 game. we have

$p_1 = Pr[b' = b]$ANONYSIGNCRYPT$(m_b^*, D_A^*, ID_j)] = \epsilon + \frac{1}{2} - \frac{q_U}{2^k}$

$p_2 = Pr[b' = i | h \in_R G_2] = 1/2$ $(i=0,1)$

$Adv(\mathcal{B}) = \frac{p_1 - p_2}{q_{H_0}^2} = (\epsilon - \frac{q_U}{2^k})/q_{H_0}^2$

Theorem 2. *(Anonymity) The improved anonymous signcryption scheme is full anonymous.*

Proof. Given a ciphertext $C = (c, S, U, R_1, ..., R_n)$, we know that $c, U, R_i(i \neq s)$ cannot leak any identity information about group identity L. It remains to consider whether R_s and S leaks information about the actual signcrypter. It has

$e(S, P) = e((x_s + H_3(c, R_i))D_s, P)) = e((x_s + H_3(c, R_i))Q_s, P_{pub})$
$= e(x_sQ_s, P_{pub})e(H_3(c, R_s)Q_s, P_{pub})$
$= e(R_s + \sum_{i=1, i \neq s}^{n}(R_i + H_3(c, R_i)Q_i), P_{pub})e(H_3(c, R_s), P_{pub}))$
$= e(\sum_{i=1, i \neq s}^{n}(R_i + H_3(c, R_i)Q_i), P_{pub})e(H_3(c, R_s) + R_s, P_{pub}))$

It seems that it can check whether ID_i is the actual signcrypter by
$e(S, P) = e(\sum_{i=1, i \neq j}^{n}(R_i + H_3(c, R_i)Q_i), P_{pub})e(H_3(c, R_j) + R_j, P_{pub}))$
However, it is no use in leaking signcrypter information because the above equality not only holds when $i = j$, but also $\forall \in \{1, ..., n\} \setminus \{j\}$.

$$e(R_i + \sum_{j=1,j\neq i}^{n}(R_j + H_3(c, R_j)Q_j), P_{pub})e(H_3(c, R_i)Q_i, P_{pub})$$
$$= e(R_i + \sum_{j=1,j\neq i}^{n}(R_j + H_3(c, R_j)Q_j) + H_3(c, R_i)Q_i, P_{pub})$$
$$= e(\sum_{j=1}^{n}(R_j + H_3(c, R_j)Q_j), P_{pub}) = e(\sum_{j=1}^{n}(x_j + H_3(c, R_j)D_j), P)$$
$$= e(S, P)$$

Theorem 3. *(*Unforgeability*) The improved anonymous signcryption scheme is existentially unforgeable against adaptive chosen-message and adaptive chosen-identity attacks (EUF-IDAS-CMIA).*

Proof. The improved scheme is unforgeable against adaptive chosen-message and chosen-identity attacks that can be derived directly from the security of Chow's ID-based ring signature scheme under the CDH assumption. If an adversary can forge a valid message of the proposed scheme, then he must be able to forge a valid Chow's ring signature. That is if $\mathcal{A}$ can forge a valid ciphertext on message m, say $C = (c, S, U, R_1, ..., R_n)$ of a user group Ł and a designated receiver ID_B, then $\sigma^* = (S, R_1, ..., R_n)$ can be viewed as the Chow's ID-based ring signature on message $m = c$ of the ring Ł.

Theorem 4. *(*Public authenticity *The improved anonymous signcryption scheme is public authenticable.*

Proof. When obtains the ciphertext $C = (c, S, U, R_1, ..., R_n)$, anyone can check the ciphertext C's origin group without knowing the content of the message and getting any help of the receiver by the following equation:

$$e(S, P) = e(P_{pub}, \sum_{i=1}^{n}(R_i + H_3(c, R_i)Q_i))$$

6   Conclusion

We have showed that the ZGCG scheme proposed in 2008 cannot provide neither anonymity for unsigncrypter nor public variability or public authencity for a third party. We also proposed an improved anonymous signcryption scheme that satisfying confidentiality, unforgeability, signcrypter anonymity and public authenticity in the random oracle model.

Acknowledgment

References

1. Zheng, Y.L.: Digital signcryption or how to achieve cost (Signature & encryption) << cost(Signature) + cost(Encryption). In: Kaliski Jr., B.S. (ed.) CRYPTO 1997. LNCS, vol. 1294, pp. 165–179. Springer, Heidelberg (1997)
2. Bao, F., Deng, R.H.: A signcryption scheme with signature directly verifiable by public key. In: Imai, H., Zheng, Y. (eds.) PKC 1998. LNCS, vol. 1431, pp. 55–59. Springer, Heidelberg (1998)
3. Yum, D.H., Lee, P.J.: New signcryption schemes based on KCDSA. In: Kim, K.-c. (ed.) ICISC 2001. LNCS, vol. 2288, pp. 533–547. Springer, Heidelberg (2002)
4. Malone-Lee, J., Mao, W.: Two birds one stone: Signcryption using RSA. In: Joye, M. (ed.) CT-RSA 2003. LNCS, vol. 2612, pp. 211–226. Springer, Heidelberg (2003)
5. Chung, K.L., Yang, G.M., Wong, D.S., Deng, X.T., Chow, S.S.M.: An efficient signcryption scheme with key privacy. In: López, J., Samarati, P., Ferrer, J.L. (eds.) EuroPKI 2007. LNCS, vol. 4582, pp. 78–93. Springer, Heidelberg (2007)
6. Joonsang, B., Ren, S., Zheng, Y.L.: Formal proofs for the security of signcryption. Journal of cryptology 20(1), 203–235 (2007)
7. Chen, L., Malone-Lee, J.: Improved identity-based signcryption. In: Vaudenay, S. (ed.) PKC 2005. LNCS, vol. 3386, pp. 362–379. Springer, Heidelberg (2005)
8. Yu, Y., Yang, B., Sun, Y., Zhu, S.L.: Identity based signcryption scheme without random oracles. Computer standard & interfaces 31(1), 56–62 (2009)
9. Barreto, P.S.L.M., Libert, B., McCullagh, N., Quisquater, J.-J.: Efficient and provably-secure identity-based signatures and signcryption from bilinear maps. In: Roy, B. (ed.) ASIACRYPT 2005. LNCS, vol. 3788, pp. 515–532. Springer, Heidelberg (2005)
10. Chow, S.S.M., Yiu, S.M., Hui, L.C.K., Chow, K.P.: Efficient forward and provably secure ID-based signcryption scheme with public verifiability and public ciphertext authenticity. In: Lim, J.-I., Lee, D.-H. (eds.) ICISC 2003. LNCS, vol. 2971, pp. 352–369. Springer, Heidelberg (2004)
11. Tan, C.H.: Analysis of improved signcryption scheme with key privacy. Information Processing Letters 99(4), 135–138 (2006)
12. Yuen, T.H., Wei, V.K.: Fast and proven secure blind identity-based signcryption from pairings. In: Menezes, A. (ed.) CT-RSA 2005. LNCS, vol. 3376, pp. 305–322. Springer, Heidelberg (2005)
13. Huang, X.Y., Su, W., Yi, M.: Identity-based ring signcryption scheme: cryptographic primitives for preserving privacy and authenticity in the ubiquitious world. In: 19th International conference on Advance Information Networking and Applications, pp. 649–654 (2003)
14. Rivest, R.L., Shamir, A., Tauman, Y.: How to leak a secret. In: Boyd, C. (ed.) ASIACRYPT 2001. LNCS, vol. 2248, pp. 552–565. Springer, Heidelberg (2001)
15. Zhang, M.W., Yang, B., Zhu, S., Zhang, W.Z.: Efficient secret authenticatable anonymous signcryption scheme with identity privacy. In: Yang, C.C., Chen, H., Chau, M., Chang, K., Lang, S.-D., Chen, P.S., Hsieh, R., Zeng, D., Wang, F.-Y., Carley, K.M., Mao, W., Zhan, J. (eds.) ISI Workshops 2008. LNCS, vol. 5075, pp. 126–137. Springer, Heidelberg (2008)
16. Zhang, J.H., Gao, S.N., Chen, H., Geng, Q.: A Novel ID-Based Anonymous Signcryption Scheme. In: Li, Q., et al. (eds.) APWeb/WAIM 2009. LNCS, vol. 5446, pp. 604–610. Springer, Heidelberg (2009)

Some Distributed Algorithms for Quantized Consensus Problem*

Jianping He, Wenhai Chen, and Lixin Gao

Institute of Operations Research and Control Science,
Wenzhou University, Zhejiang, 325027, China
lxgao@wzu.edu.cn

Abstract. In this paper, we propose some distributed algorithms for quantized consensus. These algorithms are used to study the distributed averaging problem on arbitrary connected graphs and arbitrary connected weighted graphs, with the additional constraint that the weight value at each node is an integer. These algorithms can guarantee the system achieve consensus with some moderate assumptions and can use to solve several application problems, such as averaging in a network with finite capacity channels and load balancing in a processor network, which can be modeled as distributed averaging problem.

Keywords: Quantization, Distributed detection, Consensus.

1 Introduction

Consensus problems have a long history in computer science and form the foundation of the field of distributed computing. In the last few years, the consensus problem of multi-agent systems has received major attention within the control community. This is partly due to broad applications of multi-agent systems in many areas including cooperative control of unmanned air vehicles, flocking of birds, schooling for underwater vehicles, distributed sensor networks, attitude alignment for cluster of satellites, and congestion control in communication networks [1].

A distributed averaging algorithm is a procedure using which the agents can exchange messages and update their values iteratively, so that eventually, each agent is able to compute the average of all initial values. The computation of the average is important in many different contexts, such as information fusion in sensor networks [3], load balancing in processor networks [9], and multi-agent coordination and flocking [1]. Averaging with integer constraints has been studied extensively in the load balancing literature. Load balancing algorithms can be classified into dimension-exchange algorithms and diffusion algorithms, depending on whether a processor is allowed to exchange load with only one or all, respectively, of its neighbors. Constraints on communication resources are a key factor in the design of a distributed averaging algorithm. Each agent may be able to communicate with only a small subset of all agents. The communication links between agents may not be reliable and

---

may fail over the time-scale of the computation. It is therefore of interest to design distributed averaging algorithms in which each agent needs to communicate only with its immediate neighbors, and does not need to know any further information about the global structure of the network.

Several such algorithms have been studied in [4, 6, 7, 8]. Especially in [6] study a discrete version of the distributed averaging problem that models such quantization. The paper design of a class of simple distributed algorithm, which they call quantized gossip algorithm, that converge to the set of quantized consensus distributions for an arbitrary initial vector $x(0)$ and arbitrary connected graph G. In this paper, we propose some algorithms for quantized consensus. In particular, some algorithms are used to solve the distributed averaging problem arbitrary connected weighted graphs with the additional constraint that the value at each node is an integer. So these algorithms have more practicality and also have applications to load balancing in processor networks.

The paper is organized as follows. In Section II, we give a formulation of the problem with the help of graph theory. Then in Section III, the main results of the quantized consensus and weighted quantized consensus are obtained, while in section IV, some examples are presented. Finally, the concluding remarks are given in Section V.

2   Preliminary and Problem Setup

We consider a network of N nodes, the connections between which are specified by an undirected connected graph $G = (V, E)$, where $V(G) = \{v_i : i = 1, 2, ..., N\}$ the set of nodes, $E(G) \subset \{e_{ij} = (v_i, v_j) : v_i, v_j \in V(G)\}$ is the set of edges. The set of neighbors of node v_i is denoted by $N_i = \{v_j \in v : (v_i, v_j) \in E\}$. A sub-graph G_s of G is a graph such that the vertex set $V(G_s) \subset V(G)$, and the edge set $E(G_s) \subset E(G)$. If $V(G_s) = V(G)$, we call G_S a spanning sub-graph of G. A path in graph G is a finite sequence $V_{i_1}, ..., V_{i_j}$ of vertices such that $e_{i_k i_{k+1}} \in E(G)$ for $k = 1, ..., j-1$. The digraph G is said to be connected if any two distinct nodes of the graph can be connected via a path. There is an integer value associated with each node. Time is assumed to be discrete. We denote the value at node i at time t by $x_i(t)$, and the vector of values in the network by $x(t) = [x_1(t), \cdots, x_N(t)]$. Let $S = \sum_i x_i(0)$, where $x(0)$ is the vector of initial values.

We describe algorithms in which all nodes update their values using the values of their neighbors in G in such a way that eventually, the value of each node converges to an integer approximation of the average of the initial values, $(1/N)\sum_1^N x(0)$, under the further constraints (see [6]) that:

(1) The value at each node is always an integer.

(2) The sum of values in the network does not change with time: $\sum_i x_i(t) = S$, for all time t.

Let S be written as $NL + R$, where L and R are integers with $0 \leq R < N$. We accept both L and $L+1$ as integer approximations of the true average S/N. Thus, the quantized consensus can be denoted as follows.

Definition 1 (Quantized consensus [6]). We say that a vector x has a quantized consensus distribution if $x \in \psi$ where

$$\psi = \{x \mid x_i \in \{L, L+1\}, i = 1, ...N, \sum_1^N x_i = S\} \qquad (1)$$

Similarly, we say that the network has reached quantized consensus when the vector of values $x(t)$ lies in the set ψ.

For example, in a three node network, in which initial values $x(0) = [x_1(0), x_2(0),$

$x_3(0)] = [2, 3, 5]$ the vectors which have reach quantized consensus distributions are given by $[3, 3, 4]$, $[3, 4, 3]$ and $[4, 3, 3]$, for $x(0)=[x_1(0), x_2(0), x_3(0)] =[2,3,1]$, the only such vector is $[2, 2, 2]$.

There is a lot of weighted network, such as a computer networks, the ability of each computer to deal with information is different, so different computer should be given different weight when we use them to deal with information. Therefore, we need to consider the consensus problems of a weighted network.

Definition 2 (Weighted quantized consensus). We say that a vector x has a quantized consensus distribution if $x \in \psi_w$ where

$$\psi_w = \{x \mid \left| \frac{x_i(t)}{a_i} - \frac{x_j(t)}{a_j} \right| \leq 1, \ i, j = 1, ...N, \sum_1^N x_i = S\} \qquad (2)$$

Similarly, we say that the weighted network has reached quantized consensus when the vector of values $x(t)$ lies in the set ψ_w.

For example, in a three node network with the weight $[a_1, a_2, a_3] = [1, 2, 3]$, and in which initial values $x(0) = [x_1(0), x_2(0), x_3(0)] = [2, 3, 2]$ the vectors which have reach weighted quantized consensus distributions are given by $[1, 2, 4]$, $[1, 3, 3]$ and $[2, 2, 3]$, when initial values are $x(0)=[x_1(0), x_2(0), x_3(0)] = [2, 3, 1]$, the only such vector is $[1, 2, 3]$.

Obviously, definition of weighted quantized consensus is compatible with quantized consensus definition 1. From the definition 2, it is easy to know that we can get the quantized consensus as the definition 1 when every agent with weight 1.

Lemma 1 [6]. Consider a distributed algorithm in which, for any graph G, in addition to the integer-values and constant-sum constrain, the following conditions are met:

A1. For any given initial condition $x(0)$, at any time during the execution of the algorithm, the value of $x(t)$ lies in some finite set Ω, which may depend on $x(0)$.

A2. For any state $x(t) = x$, there exists a finite time t_x such that $\Pr[x(t + t_x) \in \psi \mid x(t) = x] > 0$, where we recall from (1) that ψ is the set of all vectors which have the quantized consensus distribution.

A3. If $x[t] \in \psi$, then $x(t') \in \psi$, for any $t' \geq t$.

Such an algorithm converges to quantized consensus for any graph and any initial condition $x(0)$.

3 Propose Algorithms and Main Results

In this section, we will consider the quantized consensus problem in two cases: networks with general topology and networks with weighted topology. We give a node random quantized algorithm for the networks with general topology and two algorithms for the networks with weighted topology, which achieve consensus in the presence of discrete networks.

3.1 Node Random Quantized Algorithm

Consider a class of distributed average algorithms, which we call node random quantized algorithm. In a node random quantized algorithm, at each time, one node is selected at random, independent from earlier instants, from the set V of nodes of G, and the values of this node and its neighbors are updated. Node random quantized algorithm is completely described by the method of updating values on the selected node and its neighbors, and the probability distribution over V according to which nodes are selected. We require that the method used to update the values satisfies the following properties.

Node random quantized algorithm: Say node v_i is selected at time t, then let $D_i[t] = \left\lfloor \dfrac{x_i(t) + \sum_{j \in N_i(t)} x_j(t)}{|\{j \mid j \in N_i(t)\}| + 1} \right\rfloor$, where $\lfloor \cdot \rfloor$ is down integral, such as $\lfloor 3.5 \rfloor = 3$ and $\lfloor 3.2 \rfloor = 3$. Let $d_i(t) = x_i(t) + \sum_{j \in N_i(t)} x_j(t) - \left(|\{j \mid j \in N_i(t)\}| + 1 \right) D_i[t]$, where $|\{j \mid j \in N_i(t)\}|$ is the neighbors' number of the node v_i. Then, if $d_i[t] = 0$, then the values of the nodes unchanged, i.e. $x_i(t+1) = D_i[t]$ and $x_j(t+1) = D_i[t]$, where $j \in N_i(t)$. If $d_i[t] \neq 0$, we require that

(P1) $x_i(t+1) + \sum_{j \in N_i(t)} x_j(t+1) = x_i(t) + \sum_{j \in N_i(t)} x_j(t)$;

(P2) $x_j(t+1) \in [D_i[t], D_i[t]+1]$ $\forall j \in N_i(t)$;

(P3) $x_i(t+1) = D_i[t] + 1$.

To achieve consensus, we require that the probability distribution used to select nodes be such that it assigns a positive probability to all nodes. The graph $G(t)$ can be switched at any time. When a node is selected, all edges connected with this node in $G(t)$ are also selected. Then, we assume that any edge is selected with probability. To achieve consensus, we always assume that the graph, which has node set V and all its edges with positive probability, is connected. It is easy to see that the connection condition is necessary. Then, we can get following result.

Theorem 1. For any given initial vector $x(0)$, if the values $x(t)$ are updated using node random quantized algorithm, then $\lim_{t\to\infty} \Pr[x(t)\in\psi]=1$, where ψ is denoted by (1).

Proof. From the lemma 1, if the algorithm satisfies conditions *A1*, *A2*, *A3* then we have $\lim_{t\to\infty} \Pr[x(t)\in\psi]=1$, so we just need to prove a node random quantized algorithm satisfies the conditions *A1*, *A2*, *A3*.

Let $m[t]=\min_i x_i(t)$ and $M[t]=\max_i x_i(t)$, it is easy to see that $m[t]$ is non-decreasing and $M[t]$ is non- increasing, at the same time, from

$$D_i[t]=\left\lfloor \frac{x_i(t)+\sum_{j\in N_i(t)} x_j(t)}{\left|\{j\mid j\in N_i(t)\}\right|+1}\right\rfloor, \text{ and } \min\{x_i[t],x_j[t]\}\le D_i[t]\le \max\{x_i[t],x_j[t]\}$$

where $j\in N_i(t)$, we get $m[t]\le D_i[t]\le M[t]$.

Therefore, we have $x_i(t+1),x_j(t+1)\in\left[D_i[t],D_i[t]+1\right]$, which implies that $m[t]\le x_i[t+1]\le M[t]+1$. Therefore, the values of nodes $x_i(t)\in\left[m[0],M[0]+1\right]$ for $\forall i\in(1,2,\cdots,n)$ at any time t, that is, there can be at most $M[0]-m[0]+1$ different values in the network at any time. As a result, a trivial upper bound on the size of Ω in this case is $(M[t]-m[t]+1)^N<\infty$. Also, it is easy to see that in a node random quantized algorithm, if $x(t)\in\psi$, then $x(t')\in\psi$ for any $t'\ge t$. A node random quantized algorithm satisfies conditions *A1* and *A3*.

Let $N_{\max}[t]=\left|\{i\mid x_i(t)=M[t]\}\right|$, $N_{\min}[t]=\left|\{i\mid x_i(t)=m[t]\}\right|$ be the number of nodes with the maximum value and the minimum value in the network at time t. For any time t, if $M[t]-m[t]=1$, then that $x[t]\in\psi$; If $M[t]-m[t]>1$, then there are at least a node's value is $M[t]$ and a node's value is $m[t]$. Let $u(t)$ be the set of nodes which have value $M[t]$ and $w(t)$ be the set of nodes which have value $m[t]$, it is easy to see that $u(t)$ and $w(t)$ are non-empty at time t. Select a pair of nodes from $w(t)$ and $u(t)$ respectively, such that a path between them is a shortest path between $w(t)$ and $u(t)$. Let this path be $P=V_1-V_2-\cdots-V$, where $x_1,x_2,\cdots,x_p$ are their values and let $x_1(t)=M[t]$, $x_p(t)=m[t]$. Then, we have others nodes on the path P have value between $m[t]+1$ and $M[t]-1$. Further each node on the path has a positive probability of being selected at any time sequentially, starting with the node V_2. From *(P2)* in a node random quantized algorithm, we can see that the value of the selected node is lager or equal to the value of its neighbors, then we have $x_2(t+1)=M[t]-1$, $x_1(t+1)\le M[t]-1$, in this case, $N_{\max}[t]-1\ge N_{\max}[t+1]$; or the value $x_2(t+1)=M[t]$, $x_1(t+1)\le M[t]-1$. As this, at the next time $t+2$, the node V_3 have positive probability be selected, and it have same two case as the V_2, until the node V_{p-1} is selected at time $t+p-1$, we have $x_{p-1}(t+p-1)=M[t]-1$ and $N_{\max}[t+p-1]\le N_{\max}[t+p-2]-1\le\cdots\le N_{\max}[t]-1$; or the value $x_{p-1}(t+p-1)=M[t]$,

at this case, the node's value satisfy $x_p(t + p - 1) = M[t] - 1 > m[t]$ and the minimum value satisfy $N_{min}[t] > N_{min}[t + p - 1]$. Therefore, we have proved that

$$N_{min}[t] + N_{max}[t] > N_{min}[t + p - 1] + N_{max}[t + p - 1]$$

with a positive probability.

If $N_{min}[t'] + N_{max}[t'] = 1$, then, the reasoning exactly as above, we see that there is a positive probability that $M[t] - m[t] > M[t'] - m[t']$, therefore we have proved the $M[t] - m[t]$ is strictly decrease when $M[t] - m[t] > 1$.

Using the above claim, it is easy to see that $M[t + t_x] - m[t + t_x] = 1$ with a positive probability, that is $x(t + t_x) \in \psi$, a node random quantized algorithm satisfies conditions A2, thus completes the proof of Theorem 1.

The convergence time is a random variable defined for each initialization of the network as

$$T_{con}(x) = \inf\{t \mid x(t) \in \psi\}, \text{ given that } x(0) = x.$$

Remark 1. The convergence time of a node random quantized algorithm is smaller than the given random quantized algorithm in [6], especially when the graph of network has more well connection. Such as, for a fully connected of N nodes, the values $x(t)$ are updated using a node random quantized algorithm, then $T_{con}(x) = 1$, but the values $x(t)$ are updated using random quantized algorithm given in [6], the $T_{con}(x) \geq \frac{n(n-1)}{2}$ (see lemma 6 [6])

3.2   Weighted Network Quantized Algorithms

Consider quantized consensus in the weighted network. There is a lot of weighted network, such as a computer networks, the ability of each computer to deal with information is different, so different computer should be given different weight when we use them to deal with information. Then we will introduce weighted network quantized consensus algorithm, we give two weighted algorithms as following:

Algorithm 1 (select an edge at random)**:** Say edge $\{i, j\}$ is selected at time t, and let $D_{ij}[t] = \left| \dfrac{x_i(t)}{a_i} - \dfrac{x_j(t)}{a_j} \right|$. Then, if $D_{ij}[t] < 1$, we leave the values of nodes unchanged; if $D_{ij}[t] \geq 1$, we require that

1. $x_i(t+1) + x_j(t+1) = x_i(t) + x_j(t)$;
2. If $D_{ij}[t] > 1$,   then $D_{ij}[t+1] < D_{ij}[t]$;
3. If $D_{ij}[t] = 1$, then if $\dfrac{x_i(t)}{a_i} > \dfrac{x_j(t)}{a_j}$,

$$\begin{cases} x_j(t+1) = \left\lfloor \dfrac{a_j(x_i(t)+x_j(t))}{a_i+a_j} \right\rceil \\ x_i(t+1) = \left\lfloor \dfrac{a_i(x_i(t)+x_j(t))}{a_i+a_j} \right\rfloor \end{cases} \text{ or else } \begin{cases} x_j(t+1) = \left\lfloor \dfrac{a_j(x_i(t)+x_j(t))}{a_i+a_j} \right\rfloor \\ x_i(t+1) = \left\lceil \dfrac{a_i(x_i(t)+x_j(t))}{a_i+a_j} \right\rceil \end{cases}.$$

where $\lceil \cdot \rceil$ is up integral, such as $\lceil 3.5 \rceil = 4$ and $\lceil 2.1 \rceil = 3$.

We require the probability distribution used to select edges be such that it assigns a positive probability to all edges on some spanning sub-graph of G. (However, since we prove convergence for arbitrary graphs G, there is no loss of generality in assuming that all edges of G have a positive probability of being selected, and we assume this in the remainder of this section).

Remark 2. In this weighted network quantized consensus algorithm, we assume that every node has integer weight a_i, $\forall i \in \{1,2,\cdots,n\}$, when the weight $a_i = a_j = 1, \forall i, j$ $\in \{1,2,\cdots,n\}$ at this case, this weighted network quantized consensus algorithm is equivalent to the given quantized algorithm in [6].

Remark 3. At fact, the node with the weight a_i can be seen as a_i nodes which are fully connected. As this, the weight network can be seen as a special general network.

Theorem 2. For any given initial vector $x(0)$, if the values $x(t)$ of a weighted network are updated using weighted Algorithm 1, then

$$\lim_{t\to\infty} \Pr\left[x(t)\in\psi_w\right]=1,$$

where ψ_w is denoted by (2).

Proof. Through the remark 2, 3 and the proof of theorem 1 in [6], theorem 2 can be proved similarly. Thus, the proof is omitted.

Algorithm 2 (select a node at random)**:** Say node V_i is selected at time t, and let

$$D_i[t] = \left\lfloor \dfrac{a_i x_i(t) + \sum_{j\in N_i(t)} a_j x_j(t)}{a_i + \sum_{j\in N_i(t)} a_j} \right\rfloor \text{ and } d_i[t] = \dfrac{a_i x_i(t) + \sum_{j\in N_i(t)} a_j x_j(t)}{a_i + \sum_{j\in N_i(t)} a_j} - D_i[t],$$

If $d_i[t] = 0$ then leave $x_i(t+1) = a_i D_i[t]$; $x_j(t+1) = a_j D_i[t]$, where $j \in N_i(t)$.
If $d_i[t] \neq 0$ then we require that

1. $x_i(t+1) + \sum_{j\in N_i(t)} x_j(t+1) = x_i(t) + \sum_{j\in N_i(t)} x_j(t)$;
2. $x_j(t+1) \in \left[a_j D_i[t], a_j D_i[t] + a_j \right] \ \forall \ j \in N_i(t)$;
3. $x_i(t+1) = D_i[t] + \lceil a_i d_i[t] \rceil$.

Remark 4. We require the probability distribution used to select nodes be such that it assigns a positive probability, and this weighted network quantized consensus algorithm, we also assume that every node has integer weigh $a_i = 1$, $\forall i \in \{1,2,\cdots,n\}$,

when $a_i = a_j = 1, \forall i, j \in \{1, 2, \cdots, n\}$, this weighted network quantized consensus algorithm equal to the node random quantized consensus algorithm and have the values of nodes

$$x_i(t+1) = D_i[t] + \lceil a_i d_i[t] \rceil = D_i[t] + 1 \quad x_j(t+1) \in [D_i[t], D_i[t]+1] \quad \forall \, j \in N_i(t) .$$

Theorem 3. For any given initial vector $x(0)$, if the values $x(t)$ of a weighted network are updated using weighted Algorithm 2, then

$$\lim_{t \to \infty} \Pr[x(t) \in \psi_w] = 1 .$$

where ψ_w is denoted by (2).

Proof. In algorithm 2, condition 1 ensure that the sum of values in the network does not change with time, the node V_i is selected at time t , through the condition 2, 3, we have

$$\left| \frac{x_i(t+1)}{a_i} - \frac{x_j(t+1)}{a_j} \right| \leq 1$$

then form the equality that

$$\left\{ a_j D_i[t], a_j D_i[t] + a_j \right\} = \left\{ a_j D_i[t], a_j D_i[t] + \lceil a_j d_i[t] \rceil \right\}$$

we can get $x_i(t+1) \geq x_j(t+1)$ for $\forall \, j \in N_i(t)$. It is easy to see that the conditions 1, 2, 3 in algorithm 2 have same action with the condition *P1, P2, P3* in the node random quantized algorithm as above, then follow the prove of in theorem 1 we can get that

$$\lim_{t \to \infty} \Pr\left[\left| \frac{x_i(t)}{a_i} - \frac{x_j(t)}{a_j} \right| \leq 1 \right] = 1, \forall i, j \in \{1, 2, \cdots, n\} .$$

That is

$$\lim_{t \to \infty} \Pr\left[\max_{i, j \in \{1, 2, \cdots n\}} \left| \frac{x_i(t)}{a_i} - \frac{x_j(t)}{a_j} \right| \leq 1 \right] = 1 .$$

Then we have completed the proof of theorem 3.

4 Examples

In this section, we give some examples of quantized consensus algorithms.

A linear network express in Fig 1, and the numbers in the circles are indices of nodes.
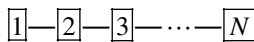
Fig. 1.

By using node random quantized algorithm, state-transition diagram of the Markov chain corresponding to the example linear network is expressed in Fig 2. We consider the network of 5 agents, each node with probability p is selected, and each directed edge denotes a transition in both direction with a probability that upper on it. If the probabilities of two sides are different, then the probability from left side to right side is upper on directed edge, and the probability from right side to left side is below in directed edge, from the following graph, we can see that if the values of the network $x(t) = (0,1,1,1,2)$ at any time t, then the values of the network will have a positive probability to be $(1,1,1,1,1)$ and will change until reach quantized consensus distribution.
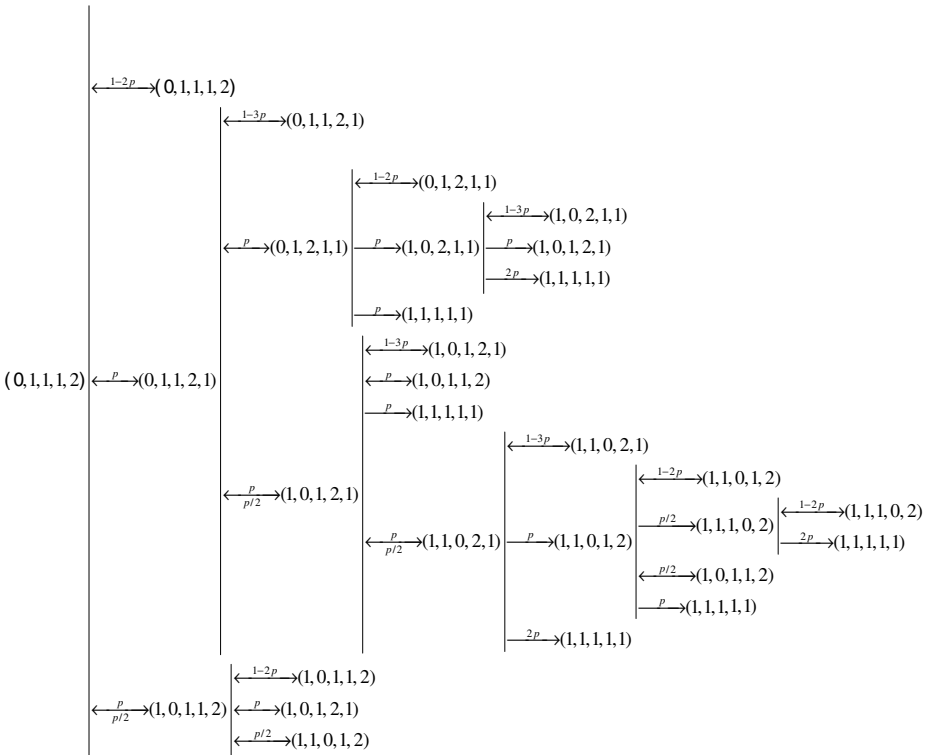
Fig. 2.

For weighted network Algorithm 1, if edge $\{i, j\}$ is selected at time t and $\dfrac{x_i(t)}{a_i} \geq \dfrac{x_j(t)}{a_j}$, then we update the node values as follows:

$$
\begin{cases}
x_j(t+1) = \left\lceil \dfrac{a_j(x_i(t) + x_j(t))}{a_i + a_j} \right\rceil \\[3mm]
x_i(t+1) = \left\lfloor \dfrac{a_i(x_i(t) + x_j(t))}{a_i + a_j} \right\rfloor
\end{cases}
$$

For example, consider the three node linear network 1-2-3 , with $x(t)=(1,2,3)$ and the weighted $a=(a_1,a_2,a_3)=(3,2,1)$. If the edge $\{1,2\}$ is selected at time t , then $x(t+1)=(2,1,3)$ .If the edge $\{2,3\}$ is selected at time t , then $x(t+1)=(1,4,1)$. If the edge $\{1,2\}$ is selected at time t and the edge $\{2,3\}$ at time $t+1$, and the edge $\{1,2\}$ is selected again at time $t+2$, then $x(t+3)=(3,2,1)$, the network reach weighted quantized consensus.

5   Conclusion and Future Work

In this paper, we have proposed some distributed algorithms, which achieve consensus in the presence of discretization. First, we consider a class of distributed average algorithms, which we call node random quantized algorithm, this algorithm may convergence faster than the similar algorithm proposed in [6]. Second we consider quantized consensus in the weighted network and give Algorithm 1 which select an edge at random and Algorithm 2 which select a node at random. Our future works involve in designing fast convergence quantized consensus algorithms and designing quantized algorithms for some detected connected graph case.

References

1. Olfati-Saber, R., Murray, R.M.: Consensus problems in networks of agents with switching topology and time-delays. IEEE Trnsactions on Automatic Control 49(9), 1520–1533 (2004)
2. Akar, M., Shorten, R.: Time synchronization for wireless sensor networks. In: Proceedings of MTNS, pp. 221–225 (2006)
3. Boyd, S., Ghosh, A., Prabhakar, B., Shah, D.: Gossip algorithms Design, analysis and applications. In: Proceedings of IEEE Infocom, pp. 1653–1664 (2005)
4. Giridhar, A., Kumar, P.R.: Distributed clock synchronization over wireless networks: Algorithms and analysis. In: Proceedings of IEEE conference on decision and control, pp. 4915–4920 (2006)
5. Herlihy, M., Tirthapura, S.: Self-stabilizing smoothing and balancing networks. Distributed Computing 18(5), 345–357 (2005)
6. Akshay, K., Tamer, B.: Quantized consensus. IEEE Transactions On Automatic (2007)
7. Houle, M.E., Symvonis, A., Wood, D.R.: Dimension-exchange algorithms for token distribution on tree-connected architectures. Journal of Parallel and Distributed Computing 64, 591–605 (2004)
8. Ghosh, B., Leighton, F.T., Maggs, B.M., Muthukrishnan, S., Plaxton, C.G., Rajaraman, et al.: Tight analyses of two local load balancing algorithms. SIAM Journal of Computing 29(1), 29–64 (1999)

Jobs Run-Time Scheduling in a Java Based Grid Architecture

Cataldo Guaragnella, Andrea Guerriero, Ciriaco C. Pasquale, and Francesco Ragni

Politecnico di Bari Bari, Italy
guaragnella@poliba.it, guerriero@poliba.it,
c.pasquale@poliba.it, ragnifrancesco@gmail.com

Abstract. Grid computing provides infrastructure for solving distributed problem by sharing, selection and aggregation of distributed resources at runtime depending on their availability, performance, cost and user's quality of service requirements. Utilization of this powerful technology is mainly conditioned by tricky management of different architectures and environments and by the difficulty to identify an efficient resource selection to map tasks into grid resources that dynamically vary their features. Resources selection needs of intelligence based automatic workflow generation to predict optimal run-time jobs allocation. In this paper we propose a dynamic job run-time scheduling system based on Java and fuzzy technology to manage Grid resources and minimize human interaction in scheduling grid jobs.

Keywords: Distributed computing, grid, fuzzy-clustering, similarity measurement.

1 Introduction

Cloud and Grid computing technologies are the dominant IT approaches to High performance computing (HPC). These network based computing technologies allow users to have access on demand to a wide range of heterogeneous and dynamic resources and services geographically distributed.

Well know examples of network computing approach are Condor [1], GridMPI [2], and Globus [3]. Several Java-based implementations have been also proposed and developed, according to Chandy's [4] and Fox's [5] infrastructures, to integrate computer resources into a distributed architecture and allow parallel applications execution on different O.S. environments.

In this paper we propose a middleware to build a Java based high-performance network computing infrastructure. High performances are related to the independent platform approach and to the fuzzy jobs run time scheduling to manage distributed resources and minimize human interaction to identify an efficient resources selection.

Java overcomes the issue of maintaining different binary codes, multiple execution environments, and complex architectures. It offers the basic infrastructure to integrate computer resources connected to Internet into a distributed computational resource and allows parallel application execution on different O.S. environments.

In previous works [6-7] we presented an agent based grid data monitoring and discovering system (MDS) and resources management; in this paper we discuss how fuzzy clustering and static/dynamic grid and jobs parameters can be used to perform efficient job scheduling.

The proposed MDS collects and processes several node and network parameters such as Cpu load, benchmark results, free memory, upload and download network bandwidth. All these parameters describe an important knowledge base in a distributed and collaborative environment and they should be used to identify an efficient job scheduling on grid resources.

Artificial intelligence components are necessary to improve the agent proposed system, aim of this research is to integrate resource management system and minimize human interaction in scheduling grid job. In literature several approaches have been proposed for grid resource selection, a large number of works evaluate fuzzy similarity measures by data mining technique based on clustering [6,8-11]. The most effective classification is obtained by dynamic fuzzy clustering [9].

Dynamic fuzzy clustering requires the following steps: data standardization, fuzzy similar matrix computation and by transitive closure, computation of the fuzzy equivalent matrix.

Fuzzy clustering time depends mainly on computation of similar matrix transitive closure that is characterized by a complexity of $O(n^3)$ [10]. According with [9-11], in order to improve the computational efficiency of traditional algorithms we propose and evaluate a different method that make the approach suitable also for grids with a very large number of nodes

Some software systems have an architecture allowing interconnections of several software modules but results in a too complicated system structure [12-14] and addition of new functions is difficult.

The proposed middleware is based on multi-agent technology and on a collaborative networked knowledge-based platform. The innovation of this approach is related onto an intelligence based automatic workflow generation that identifies an efficient jobs scheduling on available resources .

This paper is structured as follows: in the next section the basic elements of the distributed environment and services provided by autonomous agents is described. In section 3 the elements necessary to create a collaborative and secure middleware is presented. An efficient resource selection approach, based on fuzzy similarity measures to find the appropriate resource that meets task needs is finally described and discussed in section 4. Conclusion and future works close the paper.

2   An Infrastucture Based on Java Agent Technology

The design goal is to have an architecture in which various components or agents cooperate with each other effectively.

The architecture is chosen to be distributed because these components could reside on different systems. Moreover, interaction among components, developed by different programmer using Java languages, must be supported. Also the use of the architecture should be intuitive and simple. For this reason we have deployed different autonomous intercommunicating entities, using a common ontology and sharing the same communication protocol.

Java Agent (JA) is the basic entity of environment and is able to solve different simple or complex jobs, in relation with its ability and availability.

JA is the fundamental architecture component of our distributed system, and it is composed in every distributed node of our architecture by a Java engine and some Java applications and services. These agents are autonomous and interactive components able to communicate each others by asynchronous messages. The agents are characterized by: services, deployed by users using Java class applications; a container, to store and communicate Java applications available in the grid node, using a local monitoring and discovery system (MDS); a Java-engine, to schedule and manage applications.

Every grid user can create services that can be shared within grid community. Each node has an embedded container service. Planning philosophy is like common application i.e. Tomcat [10] or JBoss [11]. The developed Java applications are stored in a defined directory monitored by the container service.

Hardware resources, available Java applications, developed services and jobs queue define the "status" of every grid node. This great deal of information is the knowledge base of an agent that measures and periodically tests and updates its own performance, recording the results in a local MDS. This approach is necessary because every agent is reactive to environmental changes.

The core of every JA is the Java-engine, that allows management and standalone application execution. Using multi-thread approach, this engine is based on a scheduler that collects environment information in a local database (*it organizes its knowledge*), manages applications execution (*it manages itself*) and exchange messages and information to identify itself inside the community of agents (*it communicates in an open environment*).

Every agent implements a priority-based scheduling among local threads, creating different queues in relation with users priority policy and environmental condition. This service can manage each type of job such as I/O tasks or computational tasks, even if real management is demanded to the Operative System.

3   Distributed Environment

In the proposed multi-level architecture [18], different Java Agents are introduced in order to satisfy specific and well defined goals. We have specified two classes of
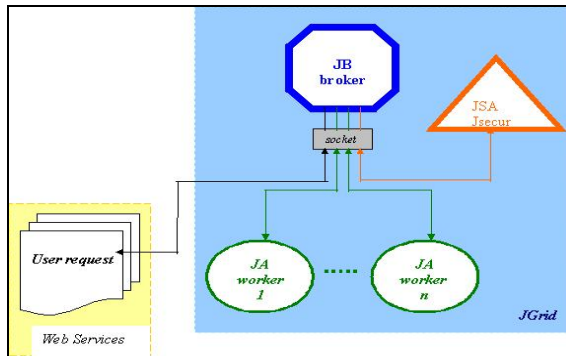
Fig. 1. Architectural schema

agents: a low level agent class composed by "JA workers", whose java-engine manages job execution in the single node; a high level agent class composed by "JB broker" (JB) and "J-Security" (JS), that manage JA worker jobs as shown in figure 1.

Each developed JA is characterized by the following services:

-Info MDS (Management and Discovering Service): this service is a variable-time synchronous application that characterizes node resources in the grid. Each node collects static parameters: "Cpu number", "Cpu frequency", "Ram on board", "Hd capacity". Moreover other static parameters are evaluated by benchmarks, applying standard test such as: FFT, Gauss-Seidel relaxation, Sparse matrix-multiply, Monte Carlo integration, Dense LU factorization. The results are measured in Mflops (Number of Mega FLoating point Operations Per Second). Dynamic parameters are considered such as: "cpu load", "Available ram", "Available capability storage", "Number of active jobs", "Network bandwidth". The agent synchronizes local MDS info with global grid MDS info, sending hardware and software information to a global grid information system stored in a broker agent. Information updating rate depends on variation of the work load in the node. These information are collected in a database to allow JB to define resources and jobs classification in order to apply the best distribution policy, using different technique. In the next section we'll show the result applying a fuzzy approach;

-Job Run Service: goal of this asynchronous service is to handle and execute JB requests, that are coded by an ontology to grant information around job, user and proxy. Local Job Manager can handle different queues in relation with user priority policy. This service can manage each type of job such as I/O tasks or computational tasks;

-Job Status Service: goal of this synchronous service is to monitor status of active jobs, in order to obtain information about partial/complete results of every single application;

-LRC (Local Replica Catalog) service: goal of this synchronous service is to mapping local file resources. The LRC, installed in each storage node, associates the physical locations name (PLN) to logical file name (LFN). Every agent of the system has the same structure, but in order to manage a community of agents, the proposed grid architecture is based on two 'specialized' elements.

The JB-broker agent coordinates resources, services and requests; it is a set of services needed to monitor and analyze available resources and to define management policy, using Info MDS. This agent can be also a generic node that executes jobs in conjunction with the other nodes of the GRID and gets information on its own status and available resources.

JB-broker has the same structure of JA-worker, with other services allowing the management of the distributed environment.

Services added to JB agents are:

-Resources and Jobs Classification Service: the goal is to evaluate resource and job fuzzy similarity measures;

-Job Scheduler Service: the goal of this asynchronous service is to manage job execution requests submitted by *Resources* and Jobs Classification Service;

-RLS (Replica Location Service): its goal is to map distributed file resources and manage their replicas to assure redundancy and optimize the scheduling performance.

The Replica Location Service, used to associate Logical File Name (LFN) with Physical File Name (PFN), tracks in the grid platform the location of data and possible existing replicas. This service is based on above-mentioned Local Replica Catalog (LRC) and on the JB agent creation of Replica Location Index (RLI).

JB Java-engine applies job scheduling policy to itself and also to the whole architecture, distributing and collecting works for each JA worker. JB agent-engine manages local services, Network Service and Grid Service (jobs that users submit using a Web interface). Aim of this agent is to organize and plan the information and application interchange among the network nodes, maximizing resource utilization, and minimizing job runtime.

The J-Security agent provides the security requirements, generates valid certificates and verifies proxy. It includes a certification system with valid PKI/GSI certificate and a proxy service allowing certified users to use GRID resources.

Users and computational resource credentials are stored in a security server gridnode in order to assure a high security level. Its principal services are:

-Security Service: an asynchronous service allowing access to users by means of a public key infrastructure; this approach allows association between grid node MAC address (public key) and the user private key;

-CA Service (Certification Authority Service): the goal of this asynchronous service is to verify user access using a proxy time. After proxy authentication, user can work in the grid environment.

Three main levels of user are managed to access into Grid Environment: root (web services administrator), user (generic customer that can submit applications, use available grid jobs or demo applications), guest (generic customer that can only execute demo applications).

4   Resources and Efficient Jobs Scheduling by Fuzzy Classification

The goal of JB-broker is to associate grid processing resources to applications. An accurate and effective resource selection requires evaluation of several parameters: application preferences and static and dynamic grid resource features.

A preference-based fuzzy clustering method to improve grid resource selection performance is proposed. It applies fuzzy clustering according to application preference *W* to create resource clusters, then the mean cluster performances are evaluated to find the appropriate resources that match job needs.

Since resources are classified by different application preferences, this method can also avoid excessive work loads concentration on only few resources so as to improve load balance in grid environment. The JB-broker fuzzy module service analyzes collected local MDS information.

Typically, Grid resources are geographically distributed and heterogeneous; they are owned by different individuals or organizations, resources are also characterized by dynamically varying load and availability. These features changing dynamically, make grid resources selection very difficulty.

Different grid applications have different preferences too. Fuzzy clustering evaluation is performed wherever there is a job request avoiding excessive work loads concentration on only few resources.

The performance of the proposed preference-based fuzzy clustering method for grid resource selection, has been evaluated considering the number of grid nodes. To explain the evaluation method, in the following example we have considered ten grid nodes and seven parameters among those described above: "Cpu frequency", "Ram on board", "Storage capacity", "Cpu load", "Available ram", "Available storage capacity" and "Network Bandwidth".

Resource selection is made in two steps: a selection of grid nodes meeting the job needs is identified by dynamic fuzzy clustering; the workflow evaluation on selected nodes allows job run time scheduling.

In the dynamic clustering step, a limited numbers of static and dynamic parameters are considered, in order to reduce the computational complexity $O(n^3)$. The workflow evaluation is carried out on a limited numbers of resources, on the cluster identified by the first step. In this way the approach is suitable also for grid characterized by a very large numbers of nodes.

In the first step, according to the difference in application preferences, suitable logical resource clusters are selected separately by using the λ matrix method [17]. For example, a computational-intensive application needs high power processors and therefore nodes with high Cpu frequency. Figure 2 shows how grid nodes are clustered in relation with a computational intensive job, using a threshold value λ=0,75.

	CPU [MHz]	Ram [MB]	HD [GB]	Net BW [Mb/s]	CPU Free [%]	Ram Free [%]	HD Free [%]
#1	12000	16000	200	20	80	50	70
#2	8000	12000	220	50	70	50	20
#3	3000	2000	300	50	25	45	80
#4	3000	2000	300	50	68	50	49
#5	3000	2000	300	50	90	69	70
#6	1000	512	800	100	90	66	90
#7	1200	512	800	100	87	66	20
#8	900	256	500	100	90	45	71
#9	900	256	500	80	66	20	45
#10	900	256	500	10	50	49	45

Application preference

w=[1.0   0.0   0.0   0.0   ]

Cluster: 1,2   Performance avg.: 10000,000
Cluster: 3,4,5   Performance avg.: 3000,000
Cluster: 6,7,8,9,10   Performance avg.: 980,000

Fig. 2. Example of clustering for computational-intensive applications

Fuzzy clustering results identify three main clusters. The first cluster has a high performance average value, the second cluster has a sufficient value and the third has a very low performance value.

Fuzzy clustering results are different under different preferences, even if the λ values are the same. By changing the value of λ, different clusters can be estimated in order to satisfy the requirements of high performance computing applications.

In our test we have set threshold λ =0,75; an example of resources clustering according to a job characterized by Cpu and Ram preferences is shown in figure 3. In this case only two grid nodes are able to solve computationally intensive applications.

	CPU [MHz]	Ram [MB]	HD [GB]	Net BW [Mb/s]	CPU Free [%]	Ram Free [%]	HD Free [%]
#1	12000	16000	200	20	20	70	71
#2	8000	12000	220	50	80	50	45
#3	3000	2000	300	50	49	45	36
#4	3000	2000	300	50	68	49	49
#5	3000	2000	300	50	66	20	70
#6	1000	512	800	100	50	66	90
#7	1200	512	800	100	20	66	91
#8	900	256	500	100	80	20	50
#9	900	256	500	80	10	80	45
#10	900	256	500	10	45	49	49

Application preference

w=[0.0   0.0   0.1   0.9   ]

Cluster: 2,3,4,5,6,7,8,9   Performance avg.: 71,700
Cluster: 1,10   Performance avg.: 18,800

Fig. 3. Example of clustering for computational-intensive application

	CPU [MHz]	Ram [MB]	HD [GB]	Net BW [Mb/s]	CPU Free [%]	Ram Free [%]	HD Free [%]
#1	12000	16000	200	20	50	70	70
#2	8000	12000	220	50	70	50	20
#3	3000	2000	300	50	80	45	80
#4	3000	2000	300	50	68	66	49
#5	3000	2000	300	50	90	69	70
#6	1000	512	800	100	90	90	90
#7	1200	512	800	100	87	66	91
#8	900	256	500	100	78	50	71
#9	900	256	500	80	20	20	45
#10	900	256	500	10	66	49	36

Application preference

w=[0.50   0.35   0.15   0.0   ]

Cluster: 1,2   Performance avg.: 8418,000
Cluster: 3,4,5,6,7,8,9,10   Performance avg.: 968,413

Fig. 4. Example of clustering according to high network bandwidth preference job

Best Cluster:

[1 2] Performance AVG 10000.00

Job time evaluation:

	Data time [s]	Appl time [s]	Stat job exec time [s]	Job queue [n]	Free CPU [%]	Free RAM [%]
#1	0	15	200	15	50	70
#2	125	13	210	25	70	50

Best Grid Node is: #1

Fig. 5. Example of fuzzy node selection from cluster identified in figure 2

Figure 4 shows resources clustering according to a high network bandwidth prefer-
ence job, for example a job for the visualization of a large amount of data stored in
one or more different nodes. In this case the amount of clusters decreases, while
the resource amount in some clusters increases. In the second step, after available

resources fuzzy clustering, an automatic workflow generation is used to predict run-time the best resources selection. Workflow generation uses the following parameters: data transfer time; application transfer time; statistic job execution time; number of jobs queue; available cpu and ram.

Dynamic selection is performed on this new set of parameters and cluster nodes to identify the more effective job scheduling.

Figure 5 shows the results of this second step related to the example reported in figure 2. In the following tables this two steps approach is compared to the single step approach. Table 1 reports clustering time versus number of nodes for the single step approach.

Table 2 reports clustering time versus number of nodes for the first step of the two steps approach. Table 3 reports workflow generation time versus number of nodes for the second step (workflow prediction) of the two steps approach. Tables shows how the two steps approach improves the clustering efficiency compared to the single step clustering.

Table 1. Clustering time versus number of nodes for the single step approach

Grid nodes	Parameters	Time [ms]
10	7	16
20	7	22
50	7	108
100	7	1215
150	7	5825
200	7	18425
500	7	Memory Overflow
1000	7	Memory Overflow

Table 2. Clustering time versus number of nodes for the first step of the two steps approach

Grid nodes	Parameters	Time [ms]
10	4	12
20	4	15
50	4	90
100	4	973
150	4	4659
200	4	15247
500	4	1004693
1000	4	Memory Overflow

Table 3. Step two, workflow generation time versus number of nodes for the two steps approach.

Cluster dimension	Workflow parameters	Time [ms]
10	3	4
20	3	6
50	3	15
100	3	231
150	3	1122
200	3	3066
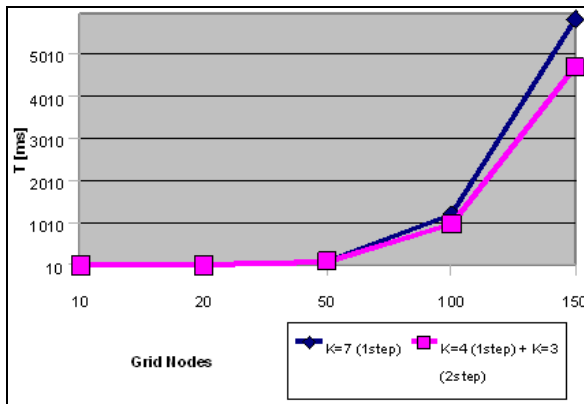
Fig. 6. Diagrams of clustering with two approaches
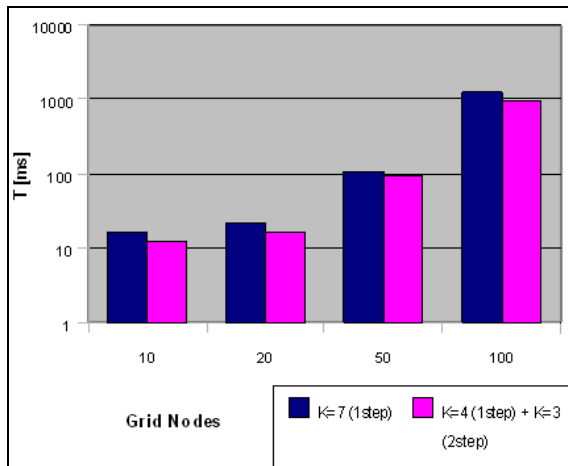
Fig. 7. Histogram of resource selection time using two approaches

These tests were performed using JVM v1.6 standard configuration with 64 MB ram allocated and Intel Centrino Duo 1.6 GHz CPU. Figure 6 shows resource selection time of the two approaches versus number of grid nodes. For larger number of grid nodes, default parameters setting of Java memory allocation, must be be increased to avoid memory overflow.

5   Conclusion and Future Work

In this paper a two step fuzzy dynamic clustering approach for an effective grid resource selection especially suitable for grid with a very large number of nodes is presented. Several static and dynamic parameters are evaluated in order to obtain an accurate and effective resource selection. Results show that the dynamic clustering of these parameters allows accurate resources selection and also avoids excessive work loads concentration on only few resources so as to improve load balancing in grid environment. The proposed two steps clustering approach allows effective workflow prediction on a reduced set of nodes, only on nodes that meet the job preferences. Identification of the most effective available resource features and job preferences and definition of the most efficient rules for job scheduling in grid environment remain open questions that will be dealt, for specific applications, i.e. image processing, in future works.

References

1. http://www.cs.wisc.edu/condor/
2. http://www.gridmpi.org/gridmpi.jsp
3. http://www.globus.org/
4. Chandy, K., Dimitron, B., Le, H., Mandleson, J., Richardson, M., Rifkin, S.A., Tawaka, P.W., Weisman L.: A world-wide distributed system using java and the internet. In: Proceedings of the 5th IEEE Internation Symposium on High Performance Distributed Computing. IEEE HPDCS (1996)
5. Fox, G., Formaski, W.: Towards web/java based high performance distributed computing – and evolving virtual machine. In: Proceedings of the 5th IEEE Internation Symposium on High Performance Distributed Computing. IEEE HPDCS (1996)
6. Guerriero, A., Pasquale, C., Ragni, F.: Resources and Jobs Fuzzy Classification in a Java Based Grid Architecture. In: Proceeding of CIMSA 2009 - IEEE International Conference on Computational Intelligence for Measurement System and Applications, Hong Kong, China, May 11-13 (2009)
7. Guerriero, A., Pasquale, C., Ragni, F.: Java based architecture for Grid Application. In: Proceeding of VECIMS 2009 - IEEE International Conference on Virtual Environments, Human-Computer Interfaces, and Measurement Systems May, Hong Kong, China, pp. 11–13 (2009)
8. Guo, D., Hu, L., Jin, S., Guo, B.: Applying Preference-based Fuzzy Clustering Technology to Improve Grid Resources Selection Performance - Fourth International Conference on Fuzzy Systems and Knowledge Discovery, FSKD 2007 (2007)
9. Hu, L., Guo, D., Che, X.: A Fast Resource Selection Approach for Grid Applications Based on Fuzzy Clustering Technology. In: The 10th IEEE International Conference on High Performance Computing and Communications

10. Li, F., Qi, D.: Research on Grid Resource Allocation Algorithm Based on Fuzzy Clustering. In: Second International Conference on Future Generation Communication and Networking
11. Lu, B., Chen, J.: Grid Resource Scheduling Based on Fuzzy Similarity Measures. In: 3rd IEEE International Conference on Cybernetics and Intelligent Systems (CIS), June 3- 6 (2008)
12. Han, J., Kamber, M.: Data Mining: Concepts and Techniques. Morgan Kaufmann, San Fransisco (2006)
13. Everitt, B.S., Landau, S., Leese, M.: Clustering Analysis. Oxford University Press, Oxford
14. Dabhi, V.K., Prajapati, H.B.: Soft Computing Based Intelligent Grid Architecture. In: Proceedings of the International Conference on Computer and Communication Engineering, Kuala Lumpur, Malaysia, May 2008, pp. 13–15 (2008)
15. http://tomcat.apache.org
16. http://www.jboss.org
17. Gui, X.L., Wang, Q.J., Gong, W.Q., Qiang, D.P.: Study of a Machine Selection Algorithm for Grid Computing. Journal of Computer Research and Development, Peking 2004-12, 2189–2194 (2004)
18. Di Lecce, V., Pasquale, C., Piuri, V.: Agent-Based Communication for an Environmental Monitoring Application. In: 14th International Conference on Computer Theory and Application, ICCTA 2004, Alexandria Egypt, September, CD-Rom version, pp. 28–30 (2004)

Pseudorandom RFID Tag Arrangement for Improved Mobile Robot Localization

Sungbok Kim

School of Electronics and Information Engineering,
Hankuk University of Foreign Studies, Gyeonggi-do 449-791, Korea
sbkim@hufs.ac.kr

Abstract. In passive RFID environment, this paper presents a pseudorandom tag arrangement for improved performance of RFID based mobile robot localization. It is assumed that a mobile robot travels along a series of linear segments, each at a constant velocity, and the number of tags sensed at one time is at most one. First, using spatial and temporal information during tag traversing, a simple but effective mobile robot localization method is developed. Second, four repetitive tag arrangements, including square, parallelogram, tilted square, and equilateral triangle, are examined. For each tag arrangement, the difficulty in tag installation and the problem of tag invisibility are discussed and compared. Third, inspired from the Sudoku puzzle, a pseudorandom tag arrangement is proposed, which is superior to conventional deterministic tag arrangement in terms of tag invisibility and tag installation.

Keywords: Mobile robot, localization, RFID, tag arrangement, pseudorandom.

1 Introduction

Recently, the RFID system has been used for mobile robot localization, which consists of tags, antenna, and reader. There have been two different research groups of RFID based mobile robot localization, where a set of tags storing the absolute positional information are deployed throughout a navigation environment. In one group, either active or passive tags are installed along the wall and they are used as beacons or landmarks to guide the navigation of a mobile robot [1]-[8]. However, in the other group, passive tags are installed on the floor and they are used to indicate the current position of a mobile robot [9]-[11]. This paper belongs to the latter group.

When an antenna senses a tag on the floor, there involves the positional uncertainty within the sensing range, which degrades the performance of RFID based mobile robot localization. One simple way of alleviating such a limitation may be to increase the tag distribution density on the floor. If more than one tag is sensed by an antenna at one instant, the current position of a mobile robot can be estimated more accurately by utilizing multiple tag readings [11]. However, the increased tag distribution density may be accompanied by the economical problem of high tag installation cost and the technical problem of incorrect tag readings.

For a given tag distribution density, the performance of RFID based mobile robot localization is affected by how a set of tags are arranged over the floor. There have

been a variety of tag arrangements considered so far, which can be categorized into three repetitive arrangements, including square, parallelogram, and tilted square. Depending on the localization method, the tag arrangement can be optimized for improved localization performance. It is claimed that the triangular pattern is optimal in [11], and the tilted pattern is optimal in [10], for example.

In this paper, we present a pseudorandom RFID tag arrangement for improved mobile robot localization. This paper is organized as follows. Section 2 describes a mobile robot localization method using spatial and temporal information. Section 3 examines four repetitive tag arrangements, including square, parallelogram, tilted square and equilateral triangle, in terms of tag installation and tag invisibility. Inspired from the Sudoku puzzle, Section 4 proposes the pseudorandom tag arrangement for reduced tag invisibility without increased installation difficulty. Finally, the conclusion is made in Section 5.

2   Mobile Robot Localization

In RFID based mobile robot localization, a mobile robot equipped with an antenna at the bottom navigates over the floor covered with a set of tags. As a mobile robot moves around, an antenna often senses tags that are located within the sensing range. In this paper, we make the following assumptions. First, a mobile robot travel along a trajectory consisting of a series of linear segments, each line segment at a constant velocity. Second, the number of tags sensed by an antenna at one instant is at most one. Third, the sensing range of an antenna is circular and the shape of a tag is a point. For explanation, it is convenient to think that a point form antenna passes through the circular tag sensing range. This is illustrated in Fig. 1.
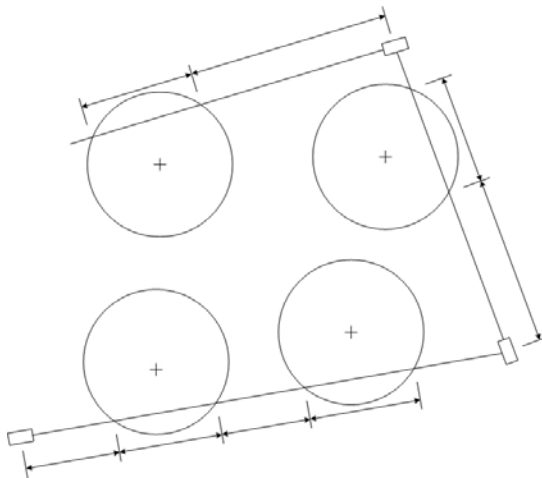
Fig. 1. Mobile robot trajectory over the floor covered with tags

2.1 Velocity Estimation

Fig. 2 depicts the situation where a mobile robot initially standing at a priori known position moves straight across the sensing range of a tag at a constant speed. Suppose that a pair of temporal information on the traverse of a mobile robot across the sensing range are given, including the elapsed time from starting to entering and the elapsed time from entering and exiting. Given these two timing information, the velocity of a mobile robot, that is, the steering angle and the forwarding speed, can be determined. Note that there are two constraints for two unknowns.

For convenience, the local coordinate system is introduced, in such a way that the tag position is defined as the coordinate origin, $O = [0\,0]^t$, and the starting position is defined at $A = [-l\,0]^t$, as shown in Fig. 2. Let r be the radius of the circular sensing range centered at a tag. Let t_1 be the elapse time during which a mobile robot starts to move and then reaches the sensing range. Let t_2 be the elapse time during which a mobile robot enters into the sensing range and then exits out of it. Let $\theta(= \angle OAB)$ be the steering angle of a mobile robot, and v be the forwarding speed along the linear segment. Let us denote $\overline{OA} = l$, $\overline{OB} = \overline{OC} = r$, $\overline{OF} = c$, $\overline{AB} = a(= vt_1)$, and $\overline{BC} = b(= vt_2)$.
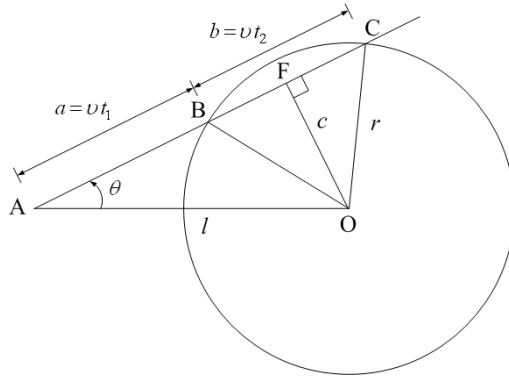
Fig. 2. Mobile robot traversing across the tag sensing range

First, from $\triangle OAF$ and $\triangle OBF$, using Pythagoras' theorem,

$$l^2 = c^2 + \left(a + \frac{b}{2}\right)^2 , \tag{1}$$

and

$$r^2 = c^2 + \left(\frac{b}{2}\right)^2 . \tag{2}$$

respectively. From (1) and (2), we can have

$$a(a+b) = l^2 - r^2, \tag{3}$$

so that the forwarding speed, v, of a mobile robot can be obtained by

$$v^2 = \frac{l^2 - r^2}{(t_1 + t_2)t_1}, \tag{4}$$

where $a = vt_1$ and $b = vt_2$ are used.

Once v is known using (4), applying the law of cosines to ΔOAB, the steering angle, θ, of a mobile robot can be determined:

$$\cos\theta = \frac{(vt_1)^2 + l^2 - r^2}{2(vt_1)l}, \tag{5}$$

which leads to

$$\theta = \text{atan2}(\pm\sqrt{1 - \cos^2\theta}, \cos\theta). \tag{6}$$

Seen from (6), there are two solutions of θ, however, only one of them can be physically true as the velocity of a mobile robot. This solution duplicity should be resolved to uniquely determine the velocity of a mobile robot. One way of resolving the solution duplicity is to utilize the information from the encoders that are readily available. For instance, the estimated steering angle using the encoder readings can be used as the reference to choose the true solution out of two possible solutions.

2.2 Position Estimation

At each sampling instant, the current position of a mobile robot will be updated using the velocity information obtained at the previous sampling instant. Unfortunately, this implies that the RFID based mobile robot localization proposed in this paper suffers from the positional error accumulation, like conventional encoder based localizations. However, in the case of RFID based localization, the positional error does not keep increasing over time but is bounded within the sensing range at each tag traversing. Under a normal floor condition, RFID based localization will work better than encoder based localization in term of positional uncertainty, while the reverse is true in terms of positional accuracy.

3   Repetitive Tag Arrangements

The performance of RFID based mobile robot localization is heavily dependent on how densely tags are distributed over the floor and how they are arranged over the floor. As the tag distribution density increases, more tag readings can be used for mobile robot localization, leading to better accuracy of localization. However, the increased tag distribution density may cause the economical problem of excessive tag installation cost as well as the technical problem of duplicated tag readings.

For a given tag distribution density, the tag arrangement over the floor affects the performance of RFID based mobile robot localization. Several tag arrangements have been considered so far, however, they can be categorized into four repetitive arrangements, including square, parallelogram, tilted square and equilateral triangle. For a given tag distribution density, it is claimed that the tag arrangement can be optimized for improved mobile robot localization, which depends on the localization method used [10],[11].

3.1   Tag Installation

One important consideration in determining the tag arrangement should be how easily a set of tags can be installed over the floor. Practically, it is very difficult or almost impossible to precisely attach many tags right on their respective locations one by one. To alleviate the difficulty in tag installation, two-step procedure can be suggested. First, attach each group of tags on a square or rectangular tile in a designated pattern. Then, place the resulting square tiles on the floor in a certain repetitive manner.

First, consider the case in which a group of four tags are placed on a square tile of side length of $2s\,(\geq 4r)$, where r is the radius of the circular tag sensing range, under the restriction that all four sensing ranges lie within a square tiles without overlapping among them. Note that the maximum number of tags sensed at one instant is assumed to be one in this paper. Fig. 3 shows three square tag patterns, including square, parallelogram, and tilted square. Fig. 3a) shows the square pattern, where four tags are located at the centers of four quadrants of a square tile.

Fig. 3b) shows the parallelogram pattern, which can be obtained from the square pattern shown in Fig. 3a) by shifting upper two tags to the right and lower two tags to the left, respectively. The degree of slanting, denoted by h, is the design parameter of
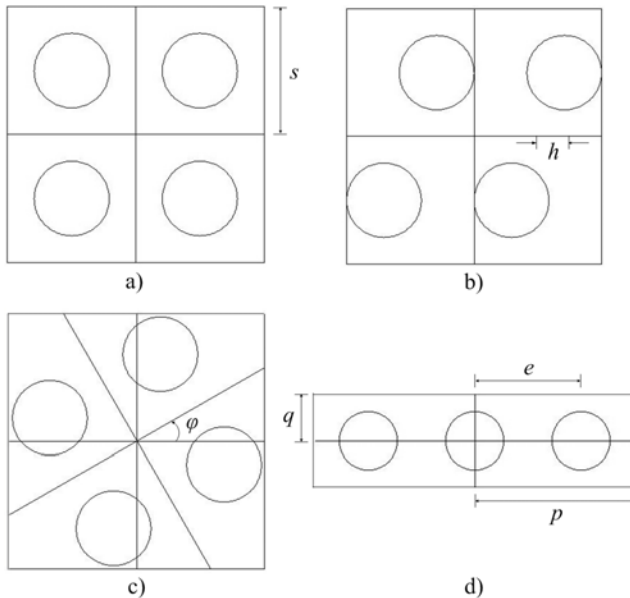
Fig. 3. Four tag patterns: a) square, b) parallelogram, c) tilted square and d) line

the parallelogram pattern. In the case of $h = \dfrac{s}{4}$, the parallelogram pattern becomes an isosceles triangular pattern [11]. And, in the case of $h = 0$, the parallelogram pattern reduces to the square pattern.

Fig. 3c) shows the tilted square pattern [10], which can be obtained by rotating the square pattern shown in Fig. 3a). The angle of rotation, denoted by φ, is the design parameter of the tilted square pattern. Note that the tilted square pattern returns to the square pattern in the case of $\varphi = 0, \dfrac{\pi}{2}$.

Next, consider the case in which a group of three tags are placed in a line on a rectangular tile of two side lengths of $2p\,(\geq 6r)$ and $2q\,(\geq 2r)$, under the same restriction imposed on three square tag patterns above. Fig. 3d) shows the line tag pattern. For later use in equilateral triangular pattern generation, we set

$$2p = 3e, \quad 2q = \frac{\sqrt{3}}{2}e, \tag{7}$$

where e denotes the tag spacing, that is, the distance between two adjacent tags. For the line pattern to have the same tag distribution density as three square patterns,

$$4s^2 : 4pq = 4 : 3. \tag{8}$$
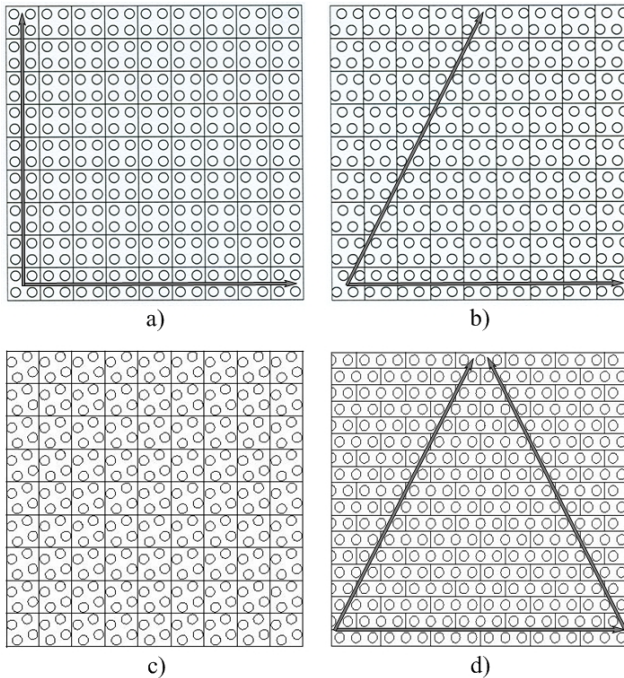
Fig. 4. Four repetitive tag arrangements: a) square, b) parallelogram, c) tilted square and d) equilateral triangle

From (7) and (8), it can be obtained that

$$e^2 = \frac{2}{\sqrt{3}} s^2 .$$

(9)

Fig. 4 shows four different tag arrangements, each of which results from placing the corresponding tag pattern in a certain repetitive manner.

3.2 Tag Invisibility

In RFID based mobile robot localization, it may happen that an antenna cannot have a chance to sense any tag during navigation, referred here to as the tag invisibility. If the tag invisibility persists for a long time, it may lead a mobile robot astray, resulting in the failure of RFID based localization. The tag invisibility should be one critical factor that needs to be taken into account in determining the tag arrangement. For a given tag distribution density, it will be desirable to make the tag visibility, which is the reverse of tag invisibility, evenly for all directions rather than being biased in some directions.

In the case of square arrangement, shown in Fig. 4a), tags cannot be sensed at all while a mobile robot moves along either horizontal or vertical directions. The smaller sensing radius compared to the tag spacing, the more serious the problem of tag invisibility. In the case of parallelogram arrangement, shown in Fig. 4b), the problem of tag invisibility is somewhat alleviated along the vertical direction, while the tag invisibility along the horizontal direction remains the same. One the other hand, in the case of tilted square tag arrangement, shown in Fig. 4c), the situation gets better along both horizontal and vertical directions. Finally, in the case of equilateral triangular tag arrangement, shown in Fig. 4d), the problem of tag invisibility occurs along three equiangular directions, however, the range of tag invisibility becomes smaller compared to the cases of both square and the parallelogram arrangements.

4   Pseudorandom Tag Arrangement

To significantly reduce the tag invisibility in all directions, the random tag arrangement seems to be best. Due to highly expected installation difficulty, however, it is hard to select the random tag arrangement in practice.

Taking into account both tag invisibility and installation difficulty, a pseudorandom tag arrangement is proposed using a set of different tilted squares that have different angles of rotation, shown in Fig. 3c). It is expected that the proposed pseudorandom tag arrangement exhibit randomness to some extent without increasing the difficulty in installation.

First, let us define a set of nine different tilted square tag patterns as follows. Since the rotation by 90° makes the resulting tilted pattern back to the original one, we propose to use the set of discrete angles of rotation, given by

$$\phi_K = (K - 1) \frac{\pi}{2} \frac{1}{9} = (K - 1) \frac{\pi}{18}, \quad K = 1, \cdots, 9 ,$$

(10)

where $K=1$ corresponds to the square pattern shown in Fig. 3a). Fig. 5 shows the set of nine different tilted square patterns, given by (10). After making nine copies of each set of nine different tilted square tag patterns, we place them on the floor side by side, according to the number placement in the Sudoku puzzle. In the Sudoku puzzle, the numbers '1' through '9' should be placed in a 9×9 array without any duplication along horizontal & vertical directions and also within 3×3 sub array.
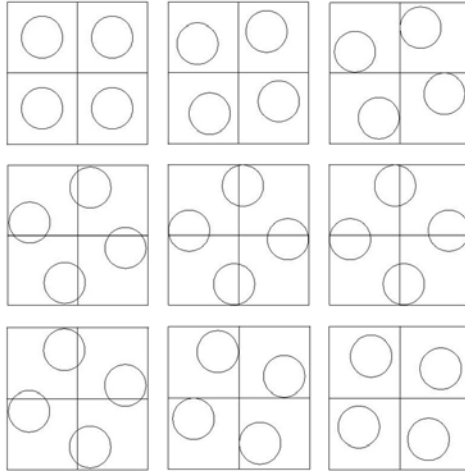
Fig. 5. The set of nine different tilted square patterns

a)

b)
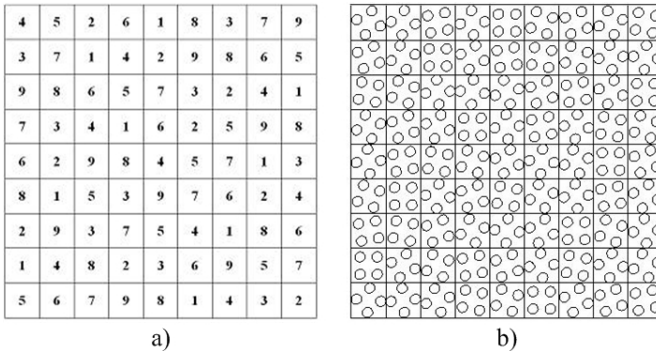
Fig. 6. Pseudorandom tag arrangement: a) one solution to the Sudoku puzzle and b) the corresponding tag arrangement

Fig. 6 shows one solution to the Sudoku puzzle and the corresponding tag arrangement. Compared to conventional deterministic tag arrangements shown in Fig. 4, it can be observed that the tag arrangement shown in Fig. 6b) exhibits randomness successively, which is called the pseudorandom tag arrangement.

5   Conclusion

This paper presented a pseudorandom RFID tag arrangement for improved perform-ance of mobile robot localization. First, using temporal as well as spatial information across tag traversing, we developed a simple but effective mobile robot localization method. Second, we examined four repetitive tag arrangements, including square, parallelogram, tilted square, and equilateral triangle, in terms of tag installation and tag (in)visibility. Third, we proposed a pseudorandom tag arrangement inspired from the Sudoku puzzle, which results in satisfactory tag visibility without increased difficulty in tag installation.

Acknowledgments

References

1. Kubitz, O., Berger, M.O., Perlick, M., Dumoulin, R.: Application of Radio Frequency Identification Devices to Support Navigation of Autonomous Mobile Robots. In: Proc. IEEE Vehicular Technology Conf., pp. 126–130 (1997)
2. Kantor, G., Singh, S.: Preliminary Results in Read-only Localization and Mapping. In: Proc. IEEE Int. Conf. Robotics and Automation, pp. 1818–1823 (2002)
3. Hahnel, D., Burgard, W., Fox, D., Fishkin, K., Philipose, M.: Mapping and Localization with RFID Technology. In: Proc. IEEE Int. Conf. Robotics and Automation, pp. 1015–1020 (2004)
4. Kulyukin, V., Gharpure, C., Nicholson, J., Pavithran, S.: RFID in Robot-assisted Indoor Navigation for the Visually Impaired. In: Proc. IEEE/RSJ Int. Conf. Intelligent Robots and Systems, pp. 1979–1984 (2004)
5. Penttila, K., Sydanheimo, L., Kivikoski, M.: Performance Development of a High-speed Automatic Object Identification Using Passive RFID Technology. In: Proc. IEEE Int. Conf. Robotics and Automation, pp. 4864–4868 (2004)
6. Yamano, K., Tanaka, K., Hirayama, M., Kondo, E., Kimura, Y., Matsumoto, M.: Self-localization of Mobile Robots with RFID System by Using Support Vector Machine. In: Proc. IEEE/RSJ Int. Conf. Intelligent Robots and Systems, pp. 3756–3761 (2004)
7. Kim, B.K., Tomokuni, N., Ohara, K., Tanikawa, T., Ohba, K., Hirai, S.: Ubiquitous Local-ization and Mapping for Robots with Ambient Intelligence. In: Proc. IEEE/RSJ Int. Conf. Intelligent Robots and Systems, pp. 4809–4814 (2006)
8. Vorst, P., Schneegans, S., Yang, B., Zell, A.: Self-localization with RFID Snapshots in Densely Tagged Environments. In: Proc. IEEE/RSJ Int. Conf. Intelligent Robots and Sys-tems, pp. 1353–1358 (2008)
9. Bohn, J., Mattern, F.: Super-distributed RFID Tag Infrastructure. In: Markopoulos, P., Eg-gen, B., Aarts, E., Crowley, J.L. (eds.) EUSAI 2004. LNCS, vol. 3295, pp. 1–12. Springer, Heidelberg (2004)
10. Choi, J.-W., Oh, D.-I., Kim, S.-W.: CPR Localization Using the RFID Tag-floor. In: Yang, Q., Webb, G. (eds.) PRICAI 2006. LNCS (LNAI), vol. 4099, pp. 870–874. Springer, Hei-delberg (2006)
11. Han, S., Lim, H., Lee, J.: An Efficient Localization Scheme for a Differential-driving Mobile Robot Based on RFID System. IEEE Trans. Industrial Electronics 54(6), 3362–3369 (2007)

Dynamic Model Identification of 2-Axes PAM Robot Arm Using Neural MIMO NARX Model

Kyoung Kwan Ahn[1] and Ho Pham Huy Anh[2]

[1] School of Mechanical and Automotive Engineering, University of Ulsan, Korea
Tel.: +82-52-259-2282
kkahn@ulsan.ac.kr
[2] Ho Chi Minh City University of Technology, Viet Nam
Tel.: +84-08-394-90415
hphanh@hcmut.edu.vn

Abstract. In this paper, a novel Forward Dynamic MIMO Neural NARX model is used for simultaneously modeling and identifying both joints of the 2-axes PAM robot arm's forward dynamic model. The contact force variation and highly nonlinear cross effect of both links of the 2-axes PAM robot arm are modeled thoroughly through a Forward Neural MIMO NARX Model-based identification process using experiment input-output training data. The results show that the novel Forward Dynamic Neural MIMO NARX Model trained by Back Propagation learning algorithm yields outstanding performance and perfect accuracy.

Keywords: Pneumatic artificial muscle (PAM), 2-axes PAM robot arm, Forward Dynamic Neural MIMO NARX model, Modeling and identification.

1 Introduction

The nonlinear n-DOF robot arm used in rehabilitation robot is belonged to highly nonlinear systems where perfect knowledge of their parameters is unattainable by conventional modeling techniques because of the time-varying inertia, external force variation and cross-effect between the. To guarantee a good force/position tracking performance, lots of researches have been carried on. Jatta et al. in [1] applied hybrid force/velocity control of industrial manipulator considering friction compensation. In [2], Khayati et al. realized a Multistage Position/Force Control for constrained robotic system using multi-objective observer. Katsura et al. in [3] installed an adaptive model used as a disturbance observer for force control of robotic manipulator and so on. Unfortunately, all such results seem still lack of robust and accuracy to be applied in force/position control of rehabilitation robot.

Thanks to their universal approximation capabilities, neural networks provide a promising implementation tool for modeling the complex input-output relations of the multiple *n*-DOF robot arm dynamics being able to solve problems like variable-coupling complexity and external force variation. During the last decade several neural network models and learning schemes have been applied to offline and online

learning of manipulator dynamics [4],[5]. Ahn and Anh in [6],[7] have successfully optimized a NARX fuzzy model of the PAM robot arm using genetic algorithm and in [8] have identified the PAM manipulator based on recurrent neural networks. The drawback of all these results is considered the *n*-DOF robot arm as *n* independent decoupling joints and the external force variation like negligible effect. Consequently, all intrinsic cross-effect features of the *n*-DOF robot arm have not represented in its recurrent neural model.

To overcome these disadvantages, in this paper, a new approach of intelligent dynamic model, namely Forward Dynamic Neural MIMO NARX model, firstly utilized in simultaneously modeling and identification both joints of the 2-axes pneumatic artificial muscle (PAM) robot arm system used as 2-axes PAM-based wrist and elbow rehabilitation robot. The results show that the novel Forward Dynamic Neural MIMO NARX Model trained by Back Propagation learning algorithm yields outstanding performance and perfect accuracy.

2 Back Propagation (BP) Learning Algorithm in Forward MIMO NARX Model Identification

A fully connected 3-layer feed-forward MLP-network with *n* inputs, *q* hidden units (also called "nodes" or "neurons"), and *m* outputs units is shown in Fig. 1.
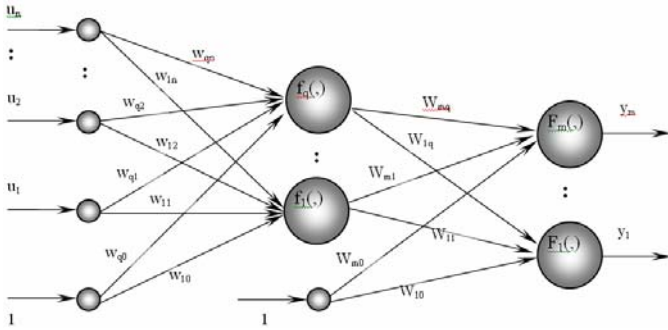
Fig. 1. Structure of feed-forward MLPNN

In Fig.1, w_{10},.., w_{q0} and W_{10},.., W_{m0} are weighting values of Bias neurons of Input Layer and Hidden Layer respectively.

The class of MLPNN-networks considered in this paper is furthermore confined to those having only one hidden layer and using sigmoid activation function:

$$\hat{y}_i(w,W) = F_i\left(\sum_{j=1}^{q} W_{ij} O_j(w) + W_{i0}\right) = F_i\left(\sum_{j=1}^{q} W_{ij}.f_j\left(\sum_{l=1}^{n} w_{jl}z_l + w_{j0}\right) + W_{i0}\right) \tag{1}$$

The objective of training is then to determine a mapping from the set of training data to the set of possible weights: $Z^N \rightarrow \hat{\theta}$ so that the network will produce

predictions $\hat{y}(t)$, which in some sense are "*closest*" to the true joint angle outputs $y(t)$ of PAM robot arm.

Concretely, the weights W_{ij} and w_{jl} of weighting vector of neural NARX are then updated as:

$$W_{ij}(k+1) = W_{ij}(k) + \Delta W_{ij}(k+1)$$
$$\Delta W_{ij}(k+1) = \lambda.\delta_i.O_j \tag{2}$$
$$\delta_i = \hat{y}_i(1-\hat{y}_i)(y_i - \hat{y}_i)$$

with δ_i is search direction value of ith neuron of output layer $(i=[1 \rightarrow m])$; O_j is the output value of jth neuron of hidden layer $(j=[1 \rightarrow q])$; y_i and $\hat{y}_i$ are truly real output and predicted output of ith neuron of output layer $(i=[1 \rightarrow m])$.

3 Experiment Configuration of the 2-Axes PAM Robot Arm

A general configuration and the schematic diagram of the 2-axes PAM robot arm and the photograph of the experimental apparatus are shown in Fig.2.

The hardware includes an IBM compatible PC (Pentium 1.7 GHz) which sends the voltage signals $u_1(t)$ and $u_2(t)$ to control the two proportional valves (FESTO, MPYE-5-1/8HF-710B), through a D/A board (ADVANTECH, PCI 1720 card) which changes digital signals from PC to analog voltage $u_1(t)$ and $u_2(t)$ respectively. Consequently, the both of joints of the 2-axes PAM robot arm will be rotated to follow the desired
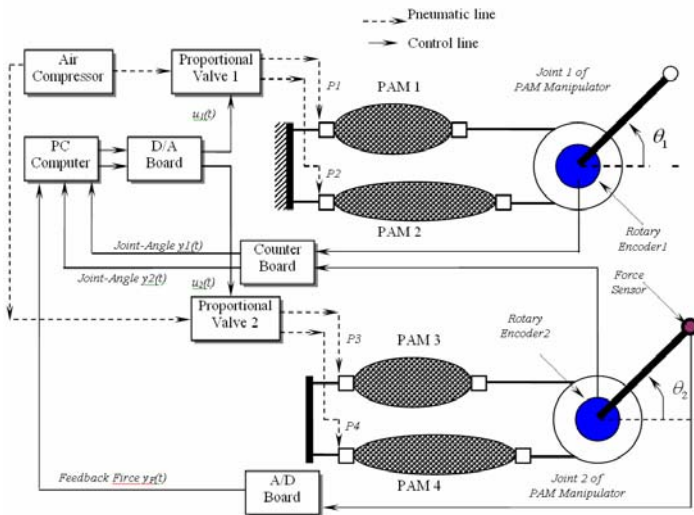
Fig. 2. Block diagram for working principle of 2-axes PAM robot arm

joint angle references ($Y_{REF1}(k)$ and $Y_{REF2}(k)$) respectively. The joint angles, θ_1[deg] and θ_2[deg], are detected by two rotary encoders (METRONIX, H40-8-3600ZO) and fed back to the computer through a 32-bit counter board (COMPUTING MEAS-UREMENT, PCI QUAD-4 card) which changes digital pulse signals to joint angle values $y_1(t)$ and $y_2(t)$. Simultaneously, through an A/D board (ADVANTECH, PCI 1711 card) which will send to PC the external force value which is detected by a force sensor CBFS-10. The pneumatic line is conducted under the pressure of 5[bar] and the software control algorithm of the closed-loop system is coded in C-mex program code run in Real-Time Windows Target of MATLAB-SIMULINK environment.

4   Results of Forward Neural MIMO NARX Model Identification

The procedure which must be executed when attempting to identify a dynamical system consists of four basic steps: Step 1 (Getting Training Data); Step 2 (Select Model Structure); Step 3 (Estimate Model) and Step 4 (Validate Model).

To realize Step 1, Fig.3 presents the PRBS input applied simultaneously to the 2 joints of the tested 2-axes PAM robot arm and the responding joint angle and feedback force outputs. This experimental PRBS input-output data is used for training and validating the Forward neural MIMO NARX model (see Fig.1) of the whole dynamic 2-joint structure of the 2-axes PAM robot arm.
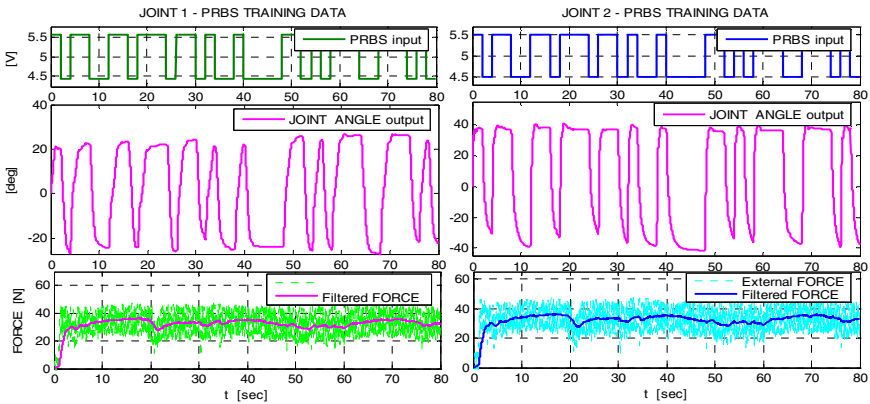
Fig. 3. Forward Neural MIMO NARX Model Training data

The 2nd step relates to select model structure. A nonlinear neural NARX model structure is attempted. The full connected Multi-Layer Perceptron (MLPNN) network architecture composes 3 layers with 5 neurons in hidden layer is selected (results derived from Ahn *et al.*, 2007 [12]). The final structure of proposed Forward neural MIMO NARX11 used in proposed neural MIMO NARX FNN-PID hybrid force/position control scheme is shown in Fig.4.

2-AXES PAM ROBOT ARM NEURAL MIMO FORWARD NARX MODEL
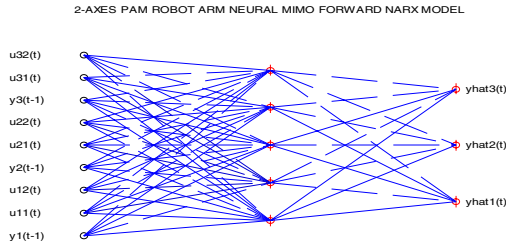
Fig. 4. Structure of Forward neural MIMO NARX11 models of 2-axes PAM robot arm
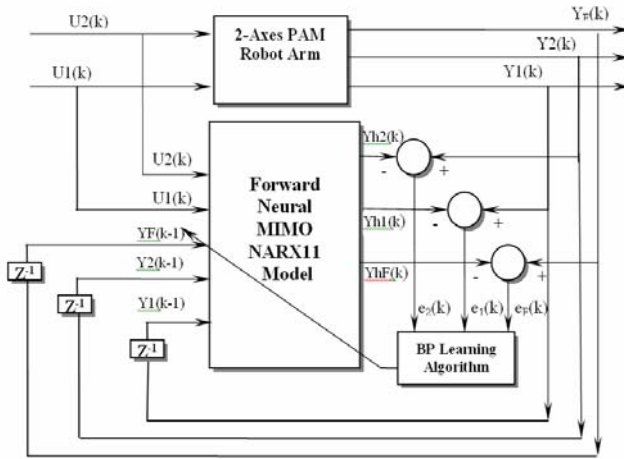
Fig. 5. Block diagram of Forward Neural MIMO NARX model identification
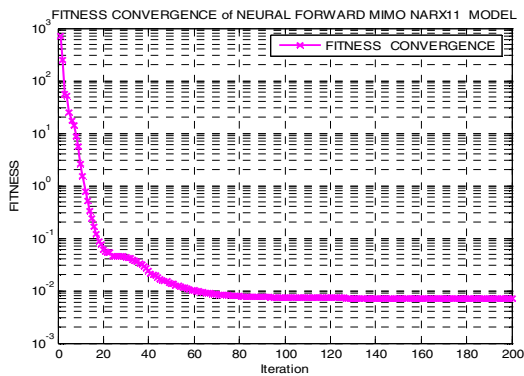
Fig. 6. Fitness convergence of Forward neural MIMO NARX Model

Fig.5 represents the block diagram for identifying Forward neural MIMO NARX11 model of the 2-axes PAM robot arm. The weighting values of proposed Forward neural MIMO NARX models will be trained and optimized with the maximum 100 iterations. The Back Propagation BP learning method is chosen as training algorithm. Consequently, the resulting Forward neural MIMO NARX models will be designed as to learn as precise as possible all dynamic and nonlinear features of the 2-axes PAM robot arm with varied external force.

The 3rd step estimates trained Forward neural MIMO NARX11 model. A good minimized convergence is shown in Fig.6 with the minimized Mean Sum of Scaled Error (MSSE) value is equal to 0.003659 after number of training 200 iterations with the proposed forward neural MIMO NARX model.
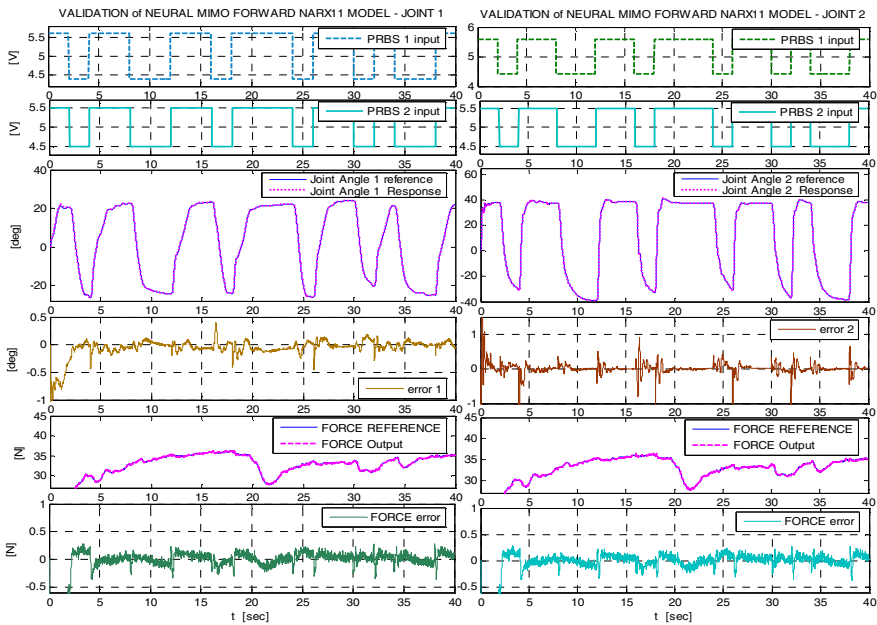
Fig. 7. Validation of proposed Forward neural MIMO NARX Model

The last step relates to validate resulting nonlinear neural Forward MIMO NARX models. Applying the same experimental diagram in Fig.5, an excellent validating result, which proves the performance of resulted Forward Neural MIMO NARX model, is shown in Fig.7. A good minimized convergence is shown in Fig.7 with the minimized Mean Sum of Scaled Error (MSSE) value is equal to 0.005577 after a number of training equal 200 iterations.

5   Conclusions

In this paper, a new approach of Forward Dynamic Neural MIMO NARX model firstly utilized in modeling and identification of the prototype 2-axes pneumatic artificial muscle (PAM) robot arm which has overcome successfully the cross-effect,

contact force variation and nonlinear characteristic of the 2-axes PAM robot arm and resulting neural Forward Dynamic MIMO NARX model surely improve the control performance of the 2-axes PAM robot arm. This resulting proposed intelligent model is quite suitable to be applied for the modeling, identification and control of various MIMO plants, including linear and nonlinear MIMO process without regard greatly change of external environments.

References

1. Jatta, F., Legnani, G., Visioli, A.: Friction compensation in hybrid force/velocity control of industrial manipulators. IEEE Trans. on Industrial Electronics 53(2), 604–613 (2006)
2. Khayati, K., Bigras, D.L.A.: A Multistage Position/Force Control for Constrained Robotic Systems with Friction: Joint-Space Decomposition, Linearization, and Multiobjective Observer/Controller Synthesis Using LMI Formalism. IEEE Trans. on Industrial Electronics 53(5), 1698–1712 (2006)
3. Katsura, S., Matsumoto, Y., Ohnishi, K.: Modeling of Force Sensing and Validation of Disturbance Observer for Force Control. IEEE Trans. on Industrial Electronics 54, 530–538 (2007)
4. Jin, M., Kang, S.H., Chang, P.H.: A new MIMO approach of modeling and identification of the 2-axes pneumatic artificial muscle (PAM) manipulator using recurrent neural networks. IEEE Trans. on Industrial Electronics 55(1), 261–269 (2008)
5. Sung, J.Y., Jin, B.P., Yoon, H.C.: Adaptive Dynamic Surface Control of Flexible-Joint Robots Using Self-Recurrent Wavelet Neural Networks. IEEE Trans. on Syst. Man and Cybern. 36(6), 1342–1355 (2006)
6. Ahn, K.K., Anh, H.P.H.: System modeling and identification of the two-link pneumatic artificial muscle (PAM) manipulator optimized with genetic algorithm. In: Proceedings 2006 IEEE-ICASE Int. Conf., Busan, Korea, pp. 4744–4749 (2006)
7. Anh, H.P.H., Ahn, K.K.: Identification of the pneumatic artificial muscle manipulators by MGA-based nonlinear NARX fuzzy model. IFAC Journal of MECHATRONICS 19(1), 106–133 (2009)
8. Anh, H.P.H., Ahn, K.K.: A new approach of modeling and identification of the pneumatic artificial muscle (PAM) manipulator based on neural networks. Proc. IMechE, Part I: Journal of Systems and Control Engineering 221(I8), 1101–1122 (2007)

Natural Language Human-Robot Interface Using Evolvable Fuzzy Neural Networks for Mobile Technology

Wojciech Kacalak and Maciej Majewski

Koszalin University of Technology, Faculty of Mechanical Engineering
Raclawicka 15-17, 75-620 Koszalin, Poland
wojciech.kacalak@tu.koszalin.pl, maciej.majewski@tu.koszalin.pl

Abstract. In this paper, a human-robot speech interface for mobile technology is described which consists of intelligent mechanisms of human identification, speech recognition, word and command recognition, command meaning and effect analysis, command safety assessment, process supervision as well as human reaction assessment. A review of selected issues is carried out with regards to recognition and evaluation of speech commands in natural language using hybrid neural networks. The paper presents experimental results of automatic recognition and evaluation of spoken commands of a manufacturing robot model simulating the execution of laser processing tasks in a virtual production process.

Keywords: Human-robot interaction, Voice communication, Speech interface, Artificial intelligence, Hybrid neural networks, Mobile technology.

1 Introduction

The advantages of intelligent human-robot speech communication for mobile technology in a production process include the following:

- More robustness against human's errors and more efficient supervising of the process with the chosen level of supervision automation [1].
- Improvement of the co-operation between a human and a robot in respect to the richness of communication [2].
- Achievement of a higher level of organization of a technological process that is equipped with an intelligent two-way speech communication system, which is relevant for its efficiency and production humanization.
- Technological decision and optimization systems can be remote elements with regard to a technological process. There is no need of a human to be present at the work station of the manufacturing robot.

Commands produced in continuous speech are processed to text with a Tablet PC after identification and authentication of the human by the mobile system shown in abbreviated form on Fig. 1. The recognized text is analyzed by the meaning analysis subsystem performing word and command recognition. This
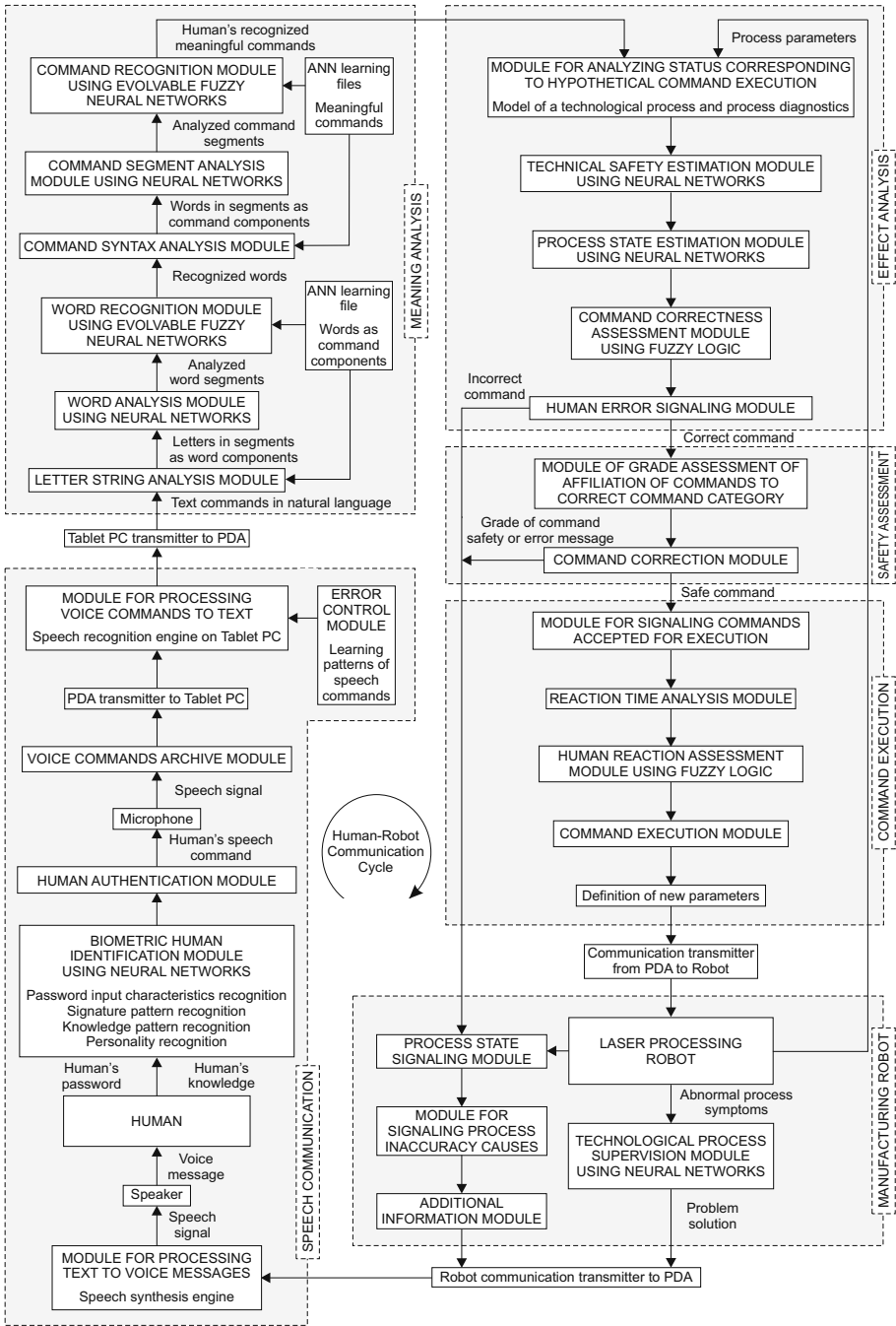
Fig. 1. Architecture of an intelligent human-robot speech communication system for mobile technology in a production process

is sent to the effect analysis subsystem which analyses the status corresponding to the hypothetical command execution, and consecutively estimates technical safety and the process state. The command is also sent to the safety assessment subsystem which assigns the command to the correct category and makes corrections. The command execution subsystem assesses the reactions of the human, and defines new parameters for the process. The manufacturing robot subsystem is composed of a laser processing robot, as well as modules for technological process supervision and signaling process state and causes of inaccuracy. The subsystem for speech communication produces voice messages to the human.

2   Automatic Command Recognition

In the automatic recognition process of commands in natural language, the speech signal is processed to text and numeric values with the module for processing voice commands to text on a Tablet PC.
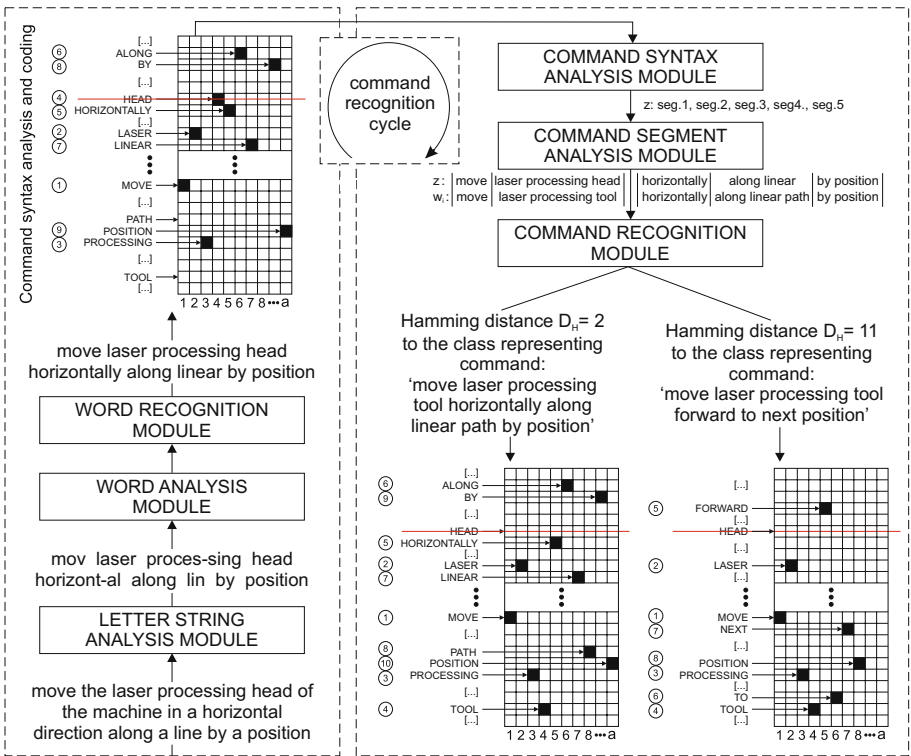
Fig. 2. Illustration of a command recognition cycle
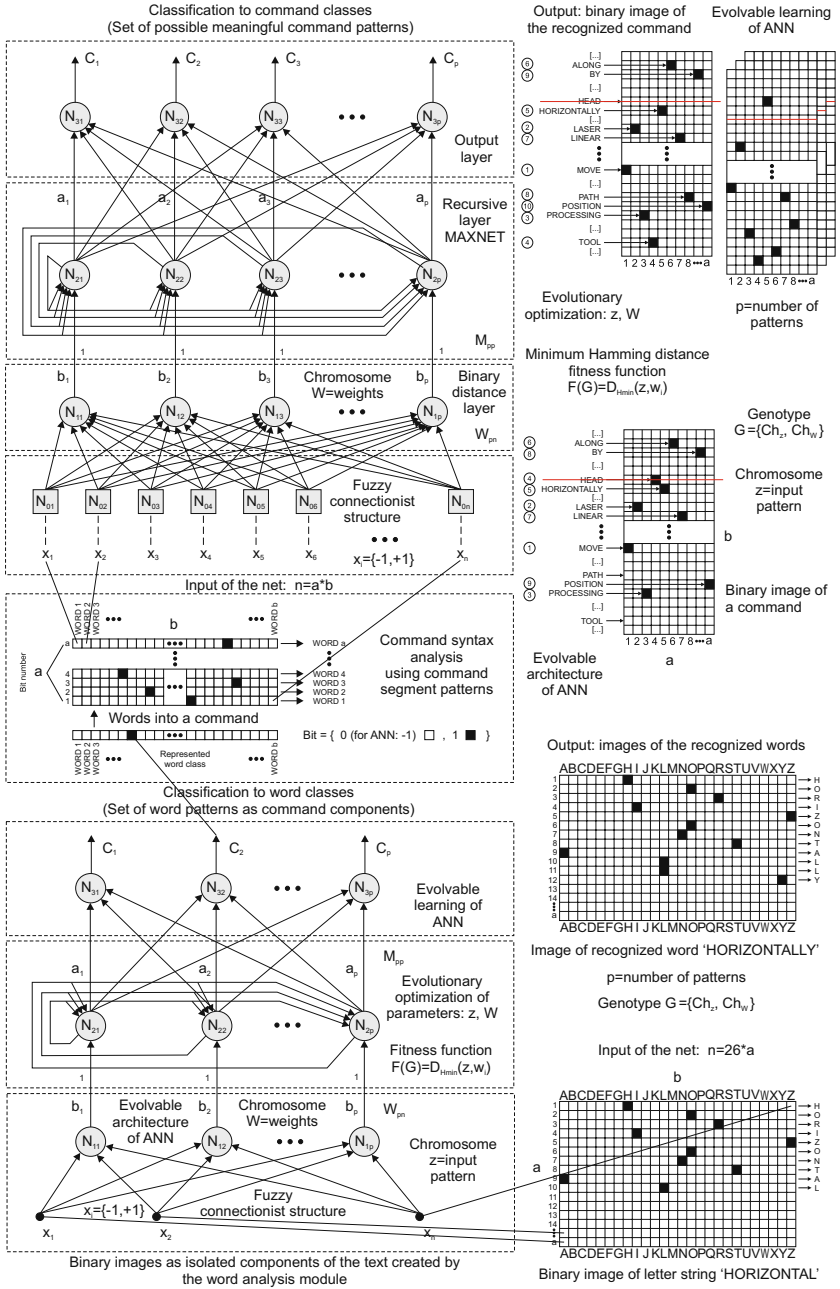
Fig. 3. Evolvable fuzzy neural networks for word and command recognition
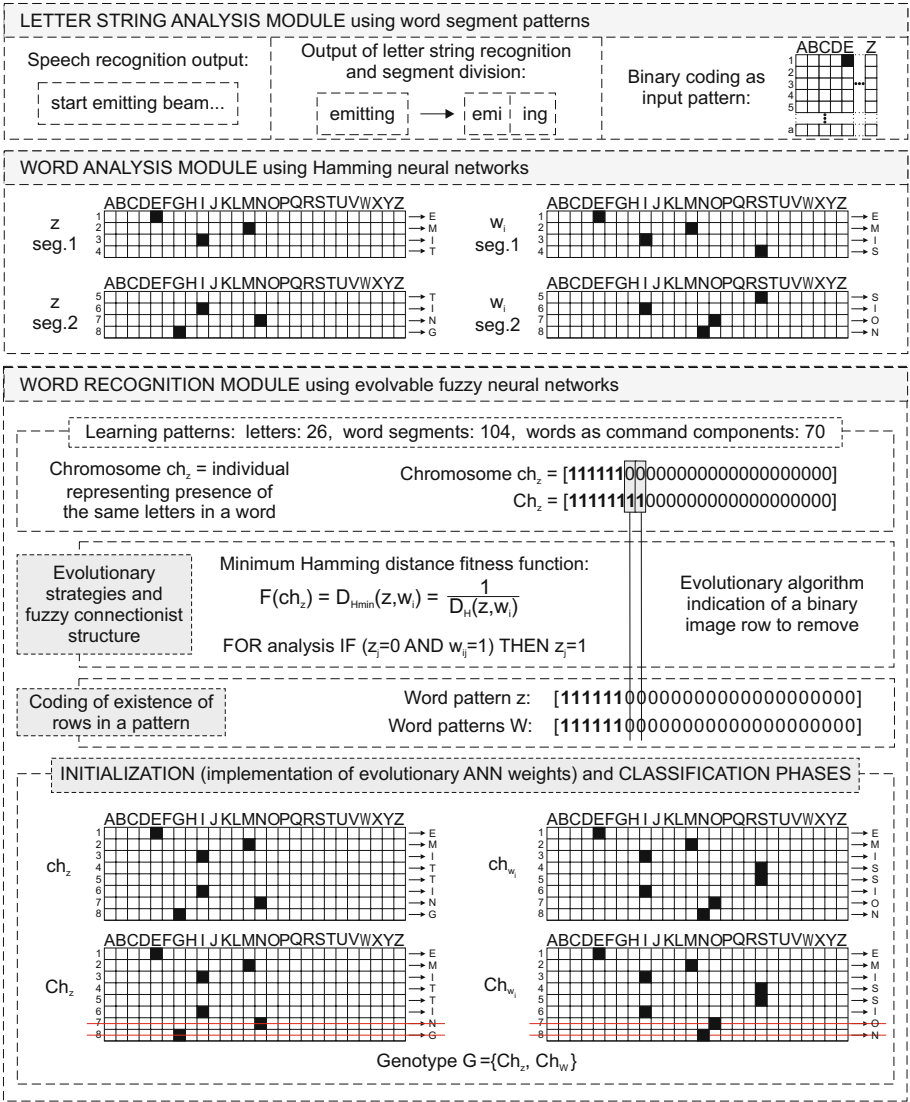
Fig. 4. Selected strategies for evolution of the neural network for word recognition

The speech recognition engine is a continuous density mixture Gaussian Hidden Markov Model system. The speech signal is transmitted from a PDA to the Tablet PC and after a successful utterance recognition, a text command in natural language is sent to the PDA. Individual words treated here as isolated components of the text are processed with the letter string analysis module.

Fig. 5. Selected strategies for neural network evolution for command recognition

The letters grouped in segments are then processed by the word analysis module. The analyzed word segments are inputs of the evolvable fuzzy neural network for recognizing words. The network uses a training file containing also words and is trained to recognize words as command components, with words represented by output neurons. In the next stage, the recognized words are transferred to the command syntax analysis module which uses command segment patterns. It analyses and divides commands into segments with regards to meaning, and codes commands as vectors. They are sent to the command segment analysis module using Hamming neural networks equipped with command segment patterns. The commands become inputs of the command recognition module. The module uses a 3-layer evolvable fuzzy Hamming neural network either to recognize the command and find its meaning or else it fails to recognize it (fig. 2). The neural network of this module uses a training file containing patterns of possible meaningful commands. The word and command recognition modules contain Hamming neural networks (fig. 3) which feature evolvable architectures and learning. Selected strategies for evolution of the matrix of the neural network weights 'W' and the input pattern 'z' as shown in abbreviated form on Fig. 4 and 5, increase the computational effectiveness of the recognition problems.

3   Command Effect Analysis and Safety Assessment

The effect analysis and safety assessment subsystems (fig. 6) make analysis of the recognized human command. The technical safety of the manufacturing robot is checked by analyzing the state of execution of the commands required to have been done as well as the commands to execute in subsequent decisions. The process parameters to be modified by executing the command are checked and the allowable changes of the parameter values are determined. They are both the input signals of the neural network of the technical safety and process state estimation modules. The neural network classifiers, which are trained with a model of the manufacturing robot work, model of the technological process, model of process diagnostics, allow for estimation of the level of safety of the recognized command as well as analysis of the status corresponding to hypothetical command execution. New values of the process parameters are the input signals of the neural networks.

An algorithm was created for assessing the technological safety of commands. The lines represent power dependence on the laser processing parameters for particular types of material. Based on the specified criteria, the laser processing power limit is determined for each type of material. Based on the laser processing power limit, the laser processing speed limit is assigned. According to the human's command, if the increase in speed makes the speed of the laser processing smaller than the smallest speed determined with the power limit, then the command is safe to be executed.
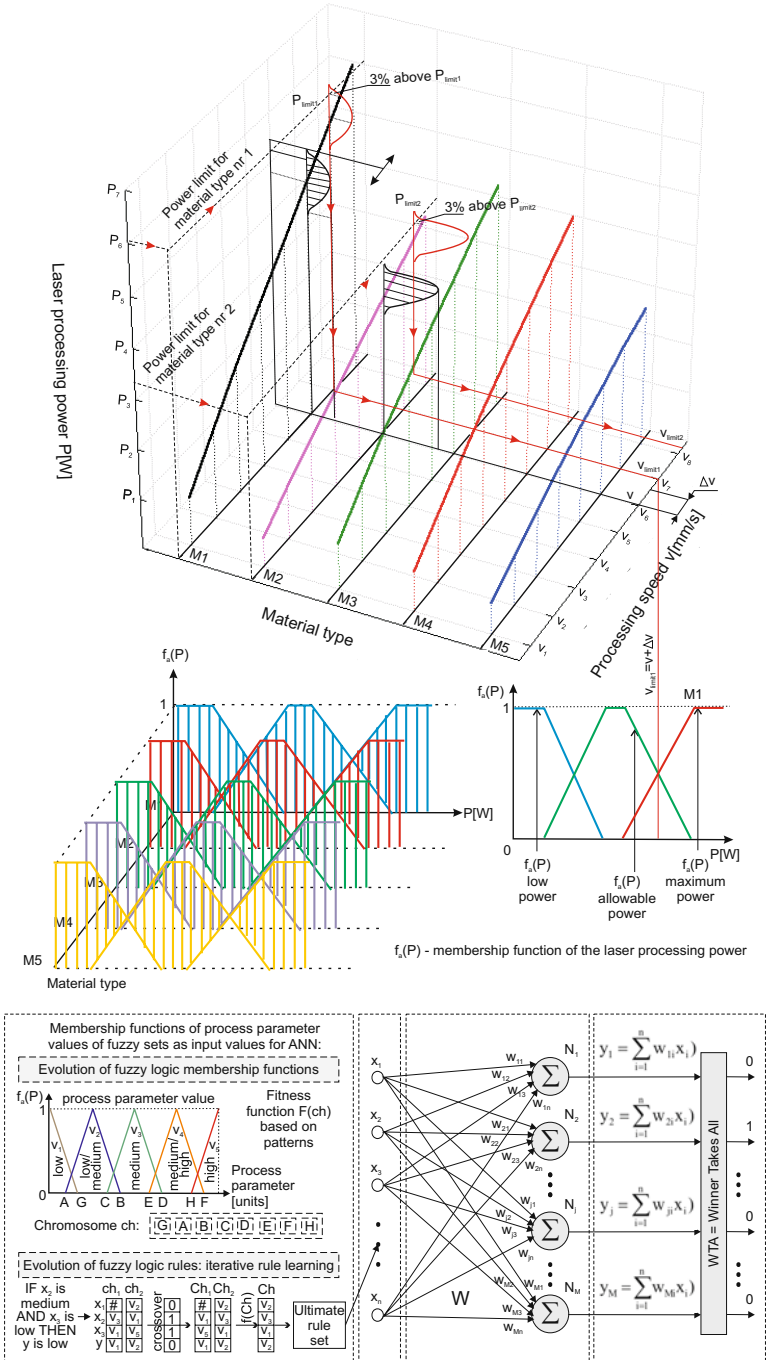
Fig. 6. Command effect analysis and safety assessment

A)
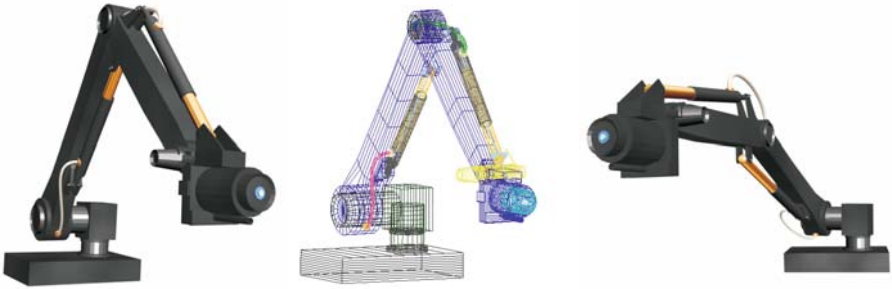
B)
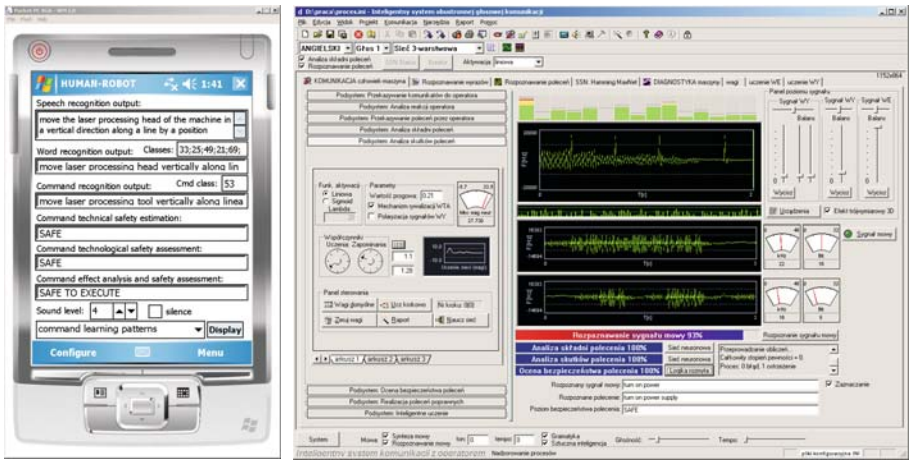
Fig. 7. Implementation of the intelligent human-robot speech interface for mobile technology: A) designed 3D manufacturing robot model executing tasks, B) written applications for a PDA and Tablet PC in C++ programming language
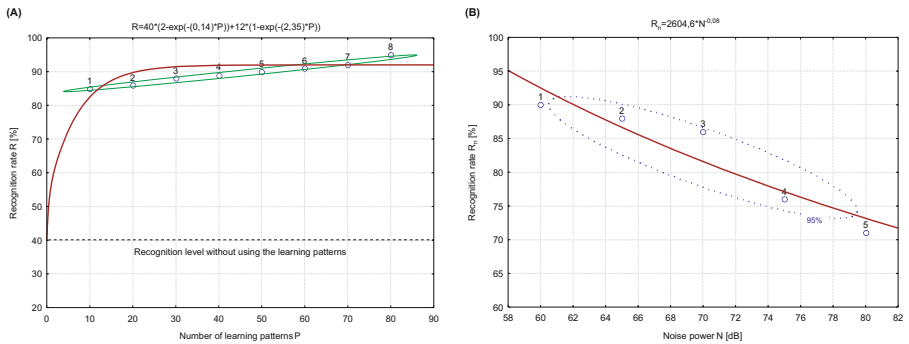
Fig. 8. Experimental results: A) Speech recognition rate; B) Speech recognition rate at different noise power

4   Experimental Results

The experimental dataset consisted of the learning files containing 70 training words and 108 meaningful training commands. The developed system (fig. 7) was executed on Windows Mobile^{TM} on a PDA (Pocket PC) and Tablet PC using Windows. The first test measured the performance of the speaker independent speech recognition with the implemented module on a PC. As shown in Fig. 8A, the speech recognition module recognizes 85-90% of the speaker's words correctly. As more training of the module is done, accuracy rises to around 95%.

For the research on command recognition at different noise power, the microphone used by the speaker was the PDA internal microphone. As shown in Fig. 8B, the recognition performance is sensitive to background noise. The recognition rate is about 86% at 70 dB and 71% at 80 dB. Therefore, background noise must be limited while giving the commands.

5   Conclusions and Perspectives

Application of evolvable fuzzy neural networks allows for recognition of commands in natural language with similar meanings but different lexico-grammatical patterns, which will undoubtedly be the most important way of communication between humans and robots. The developed flexible method can be easily extended for applications. The experimental results of the presented method of commands recognition show its excellent and promising performance, and can be used for further development and experiments.

The condition of the effectiveness of the presented system is to equip it with mechanisms of command meaning and effect analysis, as well as safety assessment. In the automated processes of production, the condition for safe communication between humans and robots is analyzing the state of the manufacturing robot and the process before the command is given and using artificial intelligence for technological effect analysis and safety assessment of the command.

References

1. Bosch, L., Kirchhoff, K.: Bridging the gap between human and automatic speech recognition. Speech Communication 49(5), 331–335 (2007)
2. O'Shaughnessy, D.: Speech Communications: Human and Machine. IEEE Computer Society Press, New York (2000)

A Robot Visual/Inertial Servoing to an Object with Inertial Sensors

Ho Quoc Phuong Nguyen[1,*], Hee-Jun Kang[2], Young-Soo Suh[2],
and Young-Shick Ro[2]

[1] Graduate School of Electrical Engineering, University of Ulsan,
680-749, Ulsan, South Korea
qtquocphuong@mail.ulsan.ac.kr
[2] School of Electrical Engineering, University of Ulsan, 680-749, Ulsan, South Korea
{hjkang,yssuh,ysro}@ulsan.ac.kr

Abstract. The paper introduces a robot visual/inertial servoing algorithm when the robot needs to track an object with inertial sensors inside. In this situation, first, inertial Jacobian is newly defined to show the relationship between an angle set velocity vector and angular velocity vector of the robot tip. That is combined with the conventional image Jacobian for the proposed robot servoing algorithm. While four landmarks have been used in the conventional visual servoing algorithm, the proposed algorithm requires only two landmarks with help of the IMU to track a moving object. Simulation and Implementation have been done to verify the feasibility of the proposed methodology.

Keywords: Visual Servoing, Visual/Inertial combined Jacobian, Attitude Estimation.

1 Introduction

Visual and inertial perception is an important combination in nature. Many tasks of human such as orientation, equilibrium, body posture control, tools control … need the combination of the two sensory systems. Some researches try to closely imitate that integration in which the inertial sensors are attached with vision sensors [1],[10],[11]. Inertial sensors were used to detect the vertical line in images captured by cameras [1] and to help reconstruct the clear images from the blur due to motion [11].

While most common researches focused on the configuration that IMU was attached to vision sensors, this paper describes a different configuration of inertial sensing: the Inertial Measurement Unit is put inside the object. A robot manipulator with stereo camera in its hand is set to track the moving object with help of the IMU information of the object. This configuration might be used to the possible application in the future such as a very quick object tracking which the vision can not follow, platform stabilization, HMD tracking ⋯ .

---

* Corresponding author.

For IMU information of the object, the attitude estimation method developed in our previous papers [8],[9] is used in this paper. However, the external acceleration is assumed to be negligible in this paper. We will takes into account the effect of external acceleration in the future work.

In this paper, inertial Jacobian is newly defined to show the relationship between an angle set velocity vector and angular velocity vector of the robot tip. The reason why we call it the inertial Jacobian is that the angle set for the robot tip will be compared with the attitude from the IMU to generate the attitude error. Then, the inertial Jacobian is combined with the conventional visual Jacobian to generate the required Cartesian velocity command to the robot.

In the field of visual servoing, conventional number of landmarks must be equal or more than fours due to the sufficient condition that is described in [13]. So, four landmarks have been widely used in the conventional visual servoing algorithm. But, the proposed algorithm requires only two landmarks with help of the IMU to track an moving object. Four landmark condition might not be easy to satisfy in natural world. It might be more comfortable if only two landmarks can be used like two rear lamps of an automobile, two eyes of an animal.

About control method for Visual Servoing, there are various ways such as an optimal control[6], LQG and pole assignment, robust control and general predictive control. Among them, one of the most effective control approach is the task function approach by Chaummette [2],[3],[4]. In this paper, we develop a robot visual/inertial servoing algorithm under the prescribed robot configuration, based on the task function approach. Simulation and real implementation have been done to verify the feasibility of the proposed methodology.

2   Nomenclatures

The key notations are listed below for convenience:

$^I\mathbf{f} \equiv \begin{bmatrix} ^Iu & ^Iv \end{bmatrix}^T$: is the feature point in image frame

$\mathbf{J}_v \in \mathbb{R}^{2k \times 6}$: Visual Jacobian, is the interaction matrix between the moving velocity of camera with respect to object and the velocity of the feature points

$\mathbf{J}_I(\mathbf{q}) \in \mathbb{R}^{2 \times 6}$: Inertial Jacobian, connects from Cartesian velocity to Euler angle given by inertial information.

$^c_b\mathbf{R}$: Rotation matrix from camera frame to body fixed frame [7]

$^b w_{0b}$: Gyro indicated value (Angular velocity of body fixed coordinate {b} with respect to base coordinate, considering in body fixed coordinate)

$\begin{bmatrix} ^b w_{0b} \times \end{bmatrix}$: Screw matrix of $^b w_{0b}$ components.

3   Visual – Inertial Jacobian

In this section, inertial Jacobian is newly defined to show the relationship between an angle set velocity vector and angular velocity vector of the robot tip. The reason why we call it the inertial Jacobian is that the angle set for the robot tip will be compared

with the attitude from the IMU to generate the attitude error. Then, the inertial Jacobian is combined with the conventional visual Jacobian to generate the required Cartesian velocity command to the robot. This process will be explained as follows.

First, we assume that the considered object is fixed and the robot moves to match $^{I}\mathbf{f}(t) \rightarrow {}^{I}\mathbf{f}^{*}$ in the image plane. The error in the image plane is defined as

$$\mathbf{e}_V(t) = {}^{I}\mathbf{f}(t) - {}^{I}\mathbf{f}^{*}. \tag{1}$$

The interaction between velocity of image features and Catersian velocity of camera is given as following:

$$\dot{\mathbf{e}}_V(t) = \mathbf{J}_V \begin{bmatrix} {}^{c}v_c \\ {}^{c}\omega_c \end{bmatrix} = \mathbf{J}_V{}^{c}\mathbf{v}_c. \tag{2}$$

In Eq. (2), $\mathbf{J}_V$ is well known stereo image visual Jacobian defined by Hiroshi Kase et. [5]

Here, the conventional exponential controller satisfying that $\dot{\mathbf{e}}_V = -\lambda\,\mathbf{e}_V$ leads to generate the input to the robot controller. Its pseudo-inversed solution is given by

$$^{c}\mathbf{v}_c = -\lambda\hat{\mathbf{J}}_V{}^{+}\mathbf{e}_V. \tag{3}$$

For six degree of freedom visual tracking, at least four landmarks are necessary for avoiding local minimum problem [13]. We'd like to develop the 6 dof visual-inertial servoing system which uses only two landmarks and the IMU information.
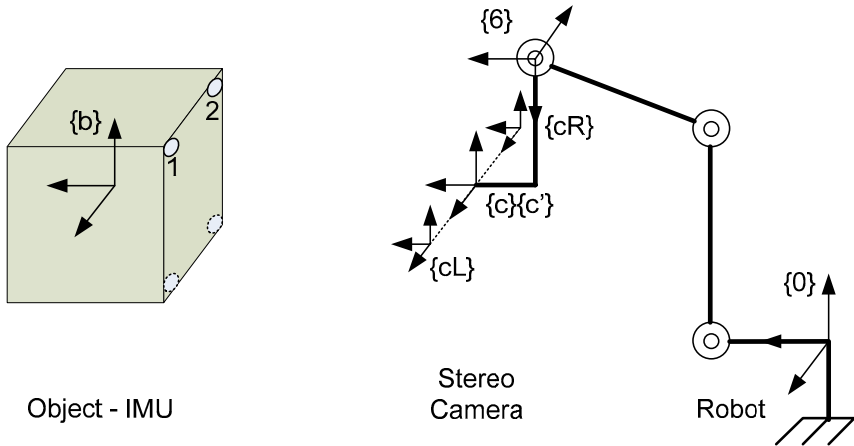
Fig. 1. Coordinate assignment of the considered system. {b} and {c} is body fixed object coordinate and camera coordinate frame, respectively. Intermediate camera coordinate frame {c'} is defined that it has same origin as {c}, but its initial orientation is the same as robot base coordinate {0}, instantly. {cL},{cR} is left and right camera coordinate, respectively.

In order to show the process, we start with attitude angle set velocity which has the relationship with the angular velocity of the camera with respect to the {c}. It is derived from Euler representation [8] such as

$$
\begin{bmatrix} \dot{\phi} \\ \dot{\theta} \end{bmatrix} = \begin{bmatrix} 1 & \sin\phi\tan\theta & \cos\phi\tan\theta \\ 0 & \cos\theta & -\sin\theta \end{bmatrix} {}^{c}\begin{bmatrix} \omega_x \\ \omega_y \\ \omega_z \end{bmatrix}_c .
\tag{4}
$$

Where ϕ and θ are the Euler angles which follow rotation order about yy' and then xx' to reach the camera frame {c'} from base frame. It is not for {c} frame.

The measured attitude angles(ϕ^* and θ^*) of the target object from the IMU would be set for the desired orientation of the manipulator. The current attitude angles is computed from manipulator forward kinematics. The difference of them gives

$$
\mathbf{e}_I = \begin{bmatrix} \phi - \phi^* \\ \theta - \theta^* \end{bmatrix} .
\tag{5}
$$

Here, the prescribed inertial Jacobian is defined as

$$
\dot{\mathbf{e}}_I = \begin{bmatrix} \dot{\phi} \\ \dot{\theta} \end{bmatrix} = \mathbf{J}_I {}^{c}\mathbf{v}_c .
\tag{6}
$$

Where the inertial Jacobian matrix is obtained from Eq. (6) as

$$
\mathbf{J}_I = \begin{bmatrix} [0_{2\times3}] & \begin{bmatrix} \cos\phi\tan\theta & \sin\phi\tan\theta & 1 \\ -\sin\theta & \cos\theta & 0 \end{bmatrix} \end{bmatrix} .
\tag{7}
$$

The required computation for those angles are as follows

$$
{}^{0}_{c'}R = {}^{0}_{6}R\,{}^{6}_{c}R\,{}^{c}_{c'}R ,
\tag{8}
$$

$$
\theta = \sin^{-1}\left({}^{0}_{c'}R_{3,1} \right) ,
\tag{9}
$$

$$
\phi = \tan^{-1}\left(\frac{{}^{0}_{c'}R_{3,2}}{{}^{0}_{c'}R_{3,3}} \right) .
\tag{10}
$$

$\mathbf{J}_I$ is a 2x6 matrix. When only two landmarks is used, $\mathbf{J}_V$ is 8x6 matrix but it doesn't contain enough information for the orientation of the object or $rank(\mathbf{J}_V) < 6$. In order to supply enough information for the movement of the object, the combined Jacobian with $\mathbf{J}_I$ and $\mathbf{J}_V$ is set up like

$$
\mathbf{J}_{VI} = \begin{bmatrix} \mathbf{J}_V \\ \mathbf{J}_I \end{bmatrix} .
\tag{11}
$$

Following the exponential control method, the required control command is expressed as

$$^c\mathbf{v}_c = -\lambda\hat{\mathbf{J}}_{VI}^+ \begin{bmatrix} \mathbf{e}_V \\ \mathbf{e}_I \end{bmatrix}. \tag{12}$$

The stability of the exponential control with the proposed combined Jacobian in Eq. (11) could be directly proved by Lyapunov Method as follows. the Lyapunov function is defined by the squared error norm:

$$L = \frac{1}{2}\|\mathbf{e}(t)\|^2, \tag{13}$$

$$\dot{L} = \mathbf{e}^T\dot{\mathbf{e}} < 0, \tag{14}$$

$$\mathbf{e}^T\dot{\mathbf{e}} = \mathbf{e}^T\mathbf{J}_{VI}{}^c\mathbf{v}_c \tag{15}$$

$$= -\lambda\mathbf{e}^T\mathbf{J}_{VI}\hat{\mathbf{J}}_{VI}^+ \mathbf{e}. \tag{16}$$

The Lyapunov condition for asymptotically stable system is satisfied if $\mathbf{J}_{VI}\hat{\mathbf{J}}_{VI}^+ \succ 0$.

4   Attitude Estimation in IMU

Attitude computation in the IMU will be briefly explained in this section. It might be helpful to refer to our previous work[9] for the details.

DCM is written in term of rotation matrix that describes the orientation of coordinates frames "b" with respect to navigation frame "n". Rotation order is about zz', yy' and then xx' corresponding to Euler angles: yaw (ψ), pitch (θ), roll (ϕ). Rotation matrix $_b^n\mathbf{C}$ can be expressed as [9]:

$$_b^n\mathbf{C} = \begin{bmatrix} \theta_c\psi_c & -\phi_c\psi_s + \phi_s\theta_s\psi_c & \phi_s\psi_s + \phi_c\theta_s\psi_c \\ \theta_c\psi_s & \phi_c\psi_c + \phi_s\theta_s\psi_s & -\phi_s\psi_c + \phi_c\theta_s\psi_s \\ -\theta_s & \phi_s\theta_c & \phi_c\theta_c \end{bmatrix}, \tag{17}$$

The notation "s" refers to sine and "c" refers to cosine. The transformation matrix $_b^n\mathbf{C}$ can be obtained with the following integration [9]:

$$_b^n\dot{\mathbf{C}} = {}_b^n\mathbf{C}\,[^b\omega_{nb}\times], \tag{18}$$

$$[^b\omega_{nb}\times] = \begin{bmatrix} 0 & -{}^b\omega_z & {}^b\omega_y \\ {}^b\omega_z & 0 & -{}^b\omega_x \\ -{}^b\omega_y & {}^b\omega_x & 0 \end{bmatrix}. \tag{19}$$

Normally, the whole DCM could be updated. However, only 3 components of DCM will be selected and updated as below:

$$
\begin{bmatrix} {}^n_b \dot{C}_{31} \\ {}^n_b \dot{C}_{32} \\ {}^n_b \dot{C}_{33} \end{bmatrix} = \begin{bmatrix} 0 & -{}^n_b C_{33} & {}^n_b C_{32} \\ {}^n_b C_{33} & 0 & -{}^n_b C_{31} \\ -{}^n_b C_{32} & {}^n_b C_{31} & 0 \end{bmatrix} \begin{bmatrix} {}^b \omega_x \\ {}^b \omega_y \\ {}^b \omega_z \end{bmatrix}.
\tag{20}
$$

Where, ${}^b \omega_x, {}^b \omega_y, {}^b \omega_z$ are approximated angular rates that measured by gyrometers when earth rotation velocity is neglected. Measurements model using accelerometers is constructed as:

$$
\begin{bmatrix} {}^b f_x \\ {}^b f_y \\ {}^b f_z \end{bmatrix} = \begin{bmatrix} C_{31} & C_{32} & C_{33} \end{bmatrix} \begin{bmatrix} 0 \\ 0 \\ g \end{bmatrix} + \alpha \, {}^c_0 R \begin{bmatrix} {}^0 a_x \\ {}^0 a_y \\ {}^0 a_z \end{bmatrix}.
\tag{21}
$$

In this equation, ${}^b f_x, {}^b f_y, {}^b f_z$ are measured by accelerometers, g is the gravitational acceleration. Parameter α is adjusted $\alpha \leq 1$ in general, $\alpha = 1$ when the tracking error is small enough. Normal Kalman filter or simply, the complement filter is used for the attitude estimation problem.

The Kalman filter with Eq. (20) of process model and Eq. (21) of measurement model estimates the attitude angles(ϕ^* and θ^*) of the target object.

5   Simulation Results

Simulation on Matlab Simulink has been performed to verify the validity of the algorithm and compare with conventional visual servoing algorithm before real implementation.

5.1   Object is Fixed

In this simulation, object is fixed with initial Euler angles are set at $\begin{bmatrix} \pi/4 & \pi/10 \end{bmatrix}$ while destination angle are $\begin{bmatrix} 0 & 0 \end{bmatrix}$ and center of destination object is put at $\begin{bmatrix} 0 & 0.7m & 0.4m \end{bmatrix}$ away from the initial point. Four landmarks are set up in the same plane. Fig 2. and Fig. 3 show the resultant trajectories of 4 landmarks in the image plane for both cases of visual inertial servoing and the conventional visual servoing. The visual inertial servoing used upper 2 landmarks and attitude information from IMU, while the visusal servoing used the 4 landmarks. In Fig. 2, the bigger landmarks are the used 2 upper landmaks, while smaller landmark trajectories are also shown for figuring out the overall trajectories of the object.
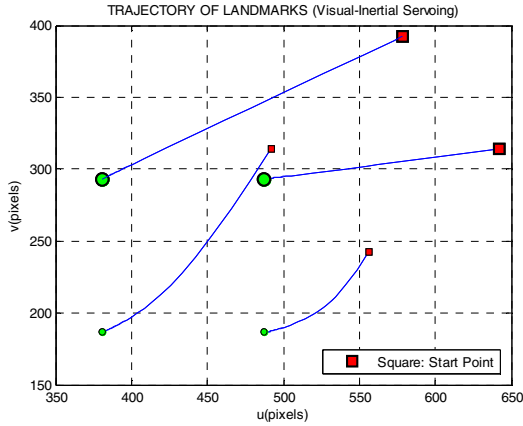
Fig. 2. Resultant trajectories of landmarks in visual inertial servoing simulation
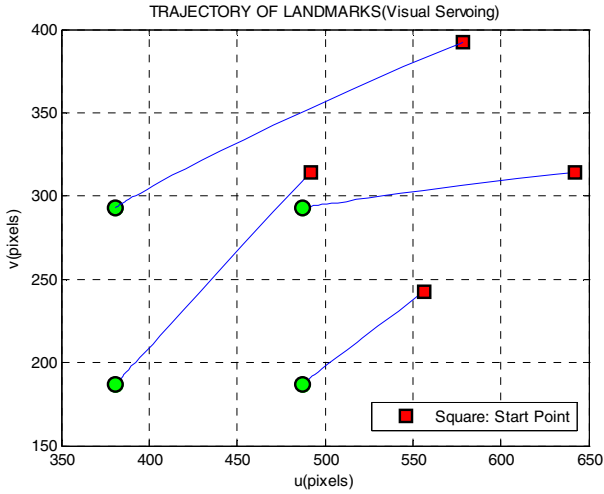
Fig. 3. Trajectories of 4 landmarks in the convetional visual servoing simulation

In visual inertial servoing, the command velocity for robot moving is not only optimized the visual error but also the Euler angle errors. Converging speed of each kind of optimized is different so that the trajectory on the image plan is less straight compared with visual servoing. This phenomenon is similar to partitioned approach to image-based visual servo control [12]. Investigate the error in detail are described in Fig. 4 and Fig.5

The Fig. 4 shows a good converging shape of Roll and Pitch angle error. It helps prove the control quality of visual inertial servoing algorithm. And visual errors are converged in the similar way of visual servoing:
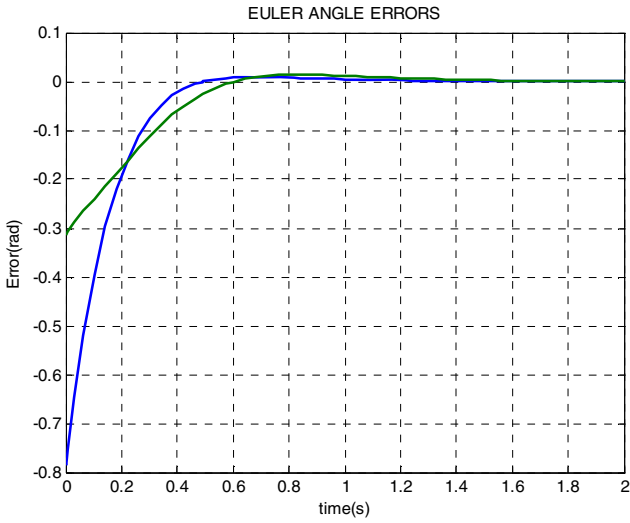
Fig. 4. Convergence of Euler angle errors in visual inertial servoing
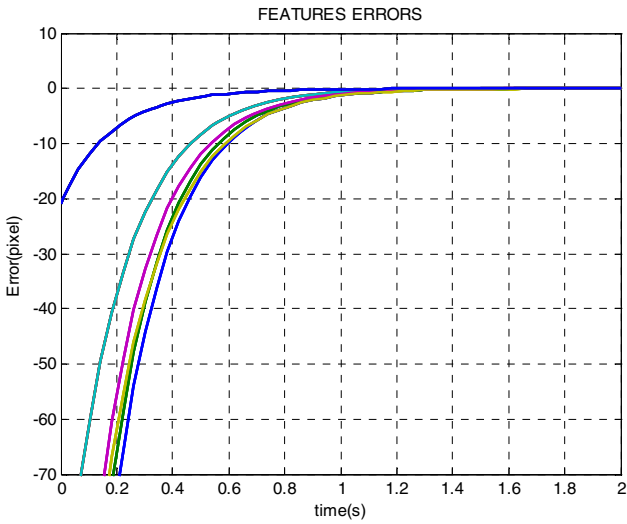
Fig. 5. Convergence of visual feature errors in visual inertial servoing

5.2   Object Rotates

The phenomenon happens in similar way, except the tracking errors is observed if the rotate speed is changed.

6 Real Implementation

Fig. 6 shows the configuration for experiment. This experiment illustrates the feedback visual inertial servoing. Control loop responses with frequency of about 15Hz. Robot used here is Samsung Fara AT2 which is controlled by MMC card and realtime API library on Window. Velocity control method is used in which robot executes certain velocity command while the computer is still processing the visual, inertial data and generate next robot command. For vision sensing, BumbleBee stereo camera of PGR is used. And Micro Infinitive AHRS system is used as inertial measurement unit.

Fig. 6. Real-time experiment system configuration. Inside the object's box with landmarks on its face is the IMU. Object can move freely on four hanging threads.

Similar to simulation configuration, object is fixed with initial Euler angles are set at $\begin{bmatrix} \pi/10 & 0 \end{bmatrix}$ while destination angle are $\begin{bmatrix} 0 & 0 \end{bmatrix}$ and center of destination object is put at about $\begin{bmatrix} 0 & 5cm & 2cm \end{bmatrix}$ far from the initial point. Let's investigate the trajectory of visual inertial servoing system in Fig. 7 and Fig.8

Result in Fig. 7 shows that visual inertial servoing control robot to reach the destination points in smooth curve shapes. Both real points and virtual point reach to its destination. The details features error is shown in Fig. 8. The implement results follow the curves that more complicate than what addressed by simulation because of many practical reasons such as the real disturbance and process speed. The shape of these curves change over different running times. But errors are always converged to zero.
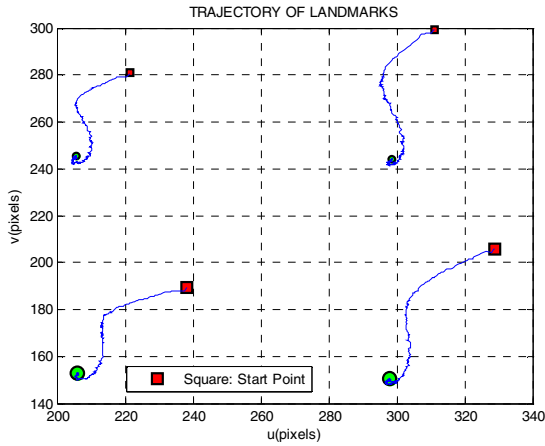
Fig. 7. Trajectory of landmarks in visual inertial servoing. The big line-end's mark face belong to real points. The small ones belong to the pseudo one(not real).
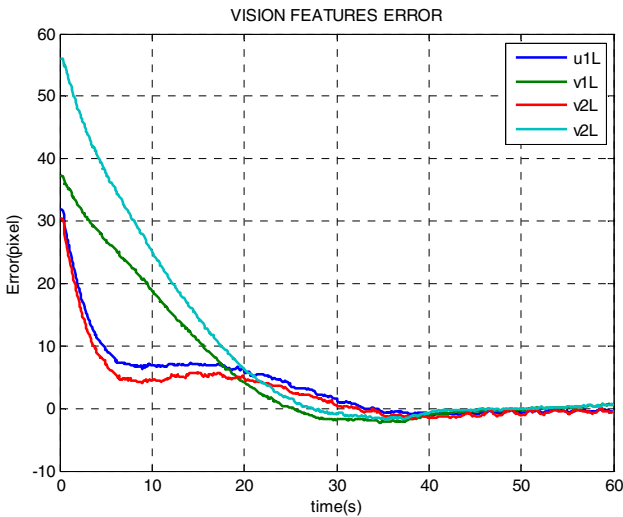
Fig. 8. Convergence of visual feature errors in visual inertial servoing

7   Conclusion

The paper proposes a new method of integration between vision and inertial in field of robot control, Visual Inertial Servoing. The most important advantage of the system is that its need only two landmarks for vision task. This simplicity made application of vision to industrial and human everyday world becomes more practical and reasonable. Many applications could be developed using the given algorithm namely automobile tracking, submarine and underwater vehicle tracking, virtual reality, cinematography and others.

Furthermore, inside the visual/inertial servoing system, the cooperation between the two sensors could be very close, they compensate the weakness of each others to increase the reliability of the sensors system. While vision sensors are easy to lose their information due to fast moving of object or robot, inertial sensors are still performing well. So, visual/inertial servoing system could be better in case of tracking rapidly rotating objects or shaking. On the other hand, inertial sensors have unwillingly bias which should to be calibrated over time. Vision sensors are long time stable, therefore, it could be use as calibration measurement. For this reason, an visual/inertial servoing could be an self calibrated system from the view of inertial application.

Acknowledgements

References

1. Peter, C., Jorge, L., Jorge, D.: An Introduction to Inertial and Visual Sensing. The International Journal of Robotics Research 26, 519 (2007)
2. Chaumette, F., Seth, H.: Visual Servo Control. Part I: Basic Approaches. IEEE Robotics & Automation Magazine (December 2006)
3. Espiau, B., Chaumette, F., Rives, P.: A new approach to visual servoing in robotics. IEEE Trans. Robot. Automat. 8(3), 313–326 (1992)
4. Hutchinson, S., Hager, G., Corke, P.: A tutorial on visual servo control. IEEE Trans. Robot. Automat. 12, 651–670 (1996)
5. Hiroshi, K., Noriaki, M., Atsushi, N., Shinya, Y., Fumio, M.: Visual Servoing of the Manipulator using the Stereo Vision. In: Proc. 1993 IEEE Int. Conf. on Industrial Electronics, Control and Instrumentation (IECON 1993), Hawaii, USA, pp. 1791–1796 (1993)
6. Hashimoto, K., Kimura, H.: LQ optimal and nonlinear approaches to visual servoing. In: Robotics and Automated Systems, vol. 7, pp. 165–198. World Scientific, Singapore (2006)
7. Craig, J.J.: Introduction to robotics mechanics and control, 3rd edn.
8. Young-Soo, S., Sang-Kyeong, P., Hee-Jun, K., Young-Shick, R.: Attitude Estimation Adaptively Compensating External Acceleration. JSME International Journal, Series C 49(1) (2006)
9. Phuong, N.H.Q., Kang, H.J., Suh, Y.S., Ro, Y.S., Lee, K.C.: A GPS/INS Integration System with New Orientation Measurement Scheme for Land Vehicle Application. In: IECON 2006 - 32nd Annual Conference on IEEE Industrial Electronics (2006)
10. Hespanha, J., Yakimenko, O., Kaminer, I., Pascoal, C.: Linear Parametrically Varying Systems with Brief Instabilities: An Application to Integrated Vision/IMU Navigation. IEEE Trans. Aerospace and Electronic Syst. 40(3), 889–902 (2004)
11. Qian, G., Chellappa, R., Zheng, Q.F.: Robust structure from motion estimation using inertial data. J. Opt. Soc. Am. A 18(12) (December 2001)
12. Corke, P.I., Hutchinson, S.A.: A New Partitioned Approach to Image-Based Visual Servo Control. IEEE Transaction on Robotics and Automation 17(4) (August 2001)
13. Malis, E., Chaumette, F.C., Boudet, S.: 2-1/2-D Visual Servoing. IEEE Transaction on Robotics and Automation 15(2) (April 1999)

A Service Framework of Humanoid in Daily Life

KangGeon Kim, Ji-Yong Lee, Seungsu Kim, Joongjae Lee, Mun-Ho Jeong,
ChangHwan Kim, and Bum-Jae You

Center for Cognitive Robotics Research,
Korea Institute of Science and Technology, Seoul, Korea
{danny,wisrage,sskim,arbitlee,mhjeong,ckim,ybj}@kist.re.kr

Abstract. This paper presents a service framework of a humanoid robot for the coordinated task execution. To execute given tasks, various sub-systems of the robot need to be coordinated effectively. The goal of our paper is to develop the service framework which makes it possible to execute various tasks in daily life environments. A script is used as a tool for describing tasks to easily regulate actions of the sub-systems while the robot is performing the task. The performance of the presented framework is experimentally demonstrated as follows: A humanoid robot, as the platform of the task execution, recognizes the designated object. The object pose is calculated by performing model-based object tracking using a particle filter with back projection-based sampling. An approach proposed by Kim et al. [1] is used to solve a human-like arm inverse kinematics and then the control system generates smooth trajectories for each joint of the humanoid robot. The results of our implementation show the robot can execute the task efficiently in human workspaces, such as an office or home.

Keywords: Task execution, Particle filter, Visual servoing, Human-like motion.

1 Introduction

Recently, there has been a great deal of interest in the Humanoid Robot. Human Robot Interaction (HRI) is the main study of the Intelligence Service Robot. To coexist with human beings, the robots have to adapt to more intelligent and varied environments [2, 3].

This paper deals with a service framework for the coordinated task execution by a humanoid robot. The humanoid robot consists of the following sub-systems: a perception system, a mobile platform, and a robotic manipulator mounted on the mobile platform [4]. The goal of our research is to develop a coordinated robot system that can execute various tasks in daily life environments. To execute diverse tasks in an indoor environment, cooperation between sub-systems is necessary. Sub-systems also need to be scheduled well according to the task sequence. In order to perform the whole task by the systematic arrangement of various modules, the service framework for scheduling, coordinating, and executing organically is required. Coordination among multiple sub-systems is necessary to achieve efficiency in humanoid robot task execution.

The task planner schedules the actions for the diverse sub-systems of the robot according to a script. We use an Extensible Markup Language (XML) script [5] for describing tasks. It makes task planning easier and has good extensibility. A user can easily arrange the tasks and actions. The task planner supplies the robot with information such as object position, path movement toward the goal, and other parameters for executing the task.

To perform the visual servoing in real time, the robots must recognize the selected object. For the recognition of the object, on the basis of the object information which is set forth, we extract the information of the object pose from the input image using particle filtering with back projection-based sampling. A back projection-based sampling reduces the search space using a depth map which can run in real time. The acquired information of the object pose is sent to the control system. The control system solves a human-like arm inverse kinematics using the method of Kim et al. [1]. The human-like arm posture is important when the robot grasps objects of various shapes without failure. The control system then generates smooth trajectories for each joint of the humanoid robot. The smooth trajectories are more stable and familiar to the person.

The remainder of this paper is organized as follows: the framework for a humanoid robot is introduced in Section 2; the task description method to execute given tasks is explained in Section 3; we describe 3D object pose estimation in Section 4.1; Section 4.2 describes human-like motion for the interaction between a human and the robot; and, in Section 5, we present the experiment of coordinated task execution.

2   Framework for a Humanoid Robot

The block diagram of the coordinated task execution framework is shown in Fig. 1. From the user interface, the information, such as a target object and a designated person, is obtained for the purpose of the task execution. Using a task script, we can schedule scenarios for the following parts: a perception system, a mobile platform, and a robotic manipulator. Each system can be controlled by a task coordinator which is essential to arrange and manage the system effectively in various situations. With our framework, diverse tasks are performed, such as moving to the object, picking up the object or handing the object, by controlling the diverse sub-systems. In the task execution process, the robot sub-system is commanded to perform autonomously and sends feedback to the task coordinator.

- The task planning step describes a task for the various sub-systems of the robot. A task could be derived from the task script autonomously or interactively by a human operator and it contains the entire information necessary to achieve the given task goal. We will describe the task planning and description method in Section 3.
- The task coordination step generates sequential actions for the execution step in order to achieve the given task goal. The task is provided by the task planning step. It is also possible to get feedback from the robot to interact with its environment. The task coordinator gets the real-time position of an object or a person, so it informs the robot the manner in which to move and how to manipulate.
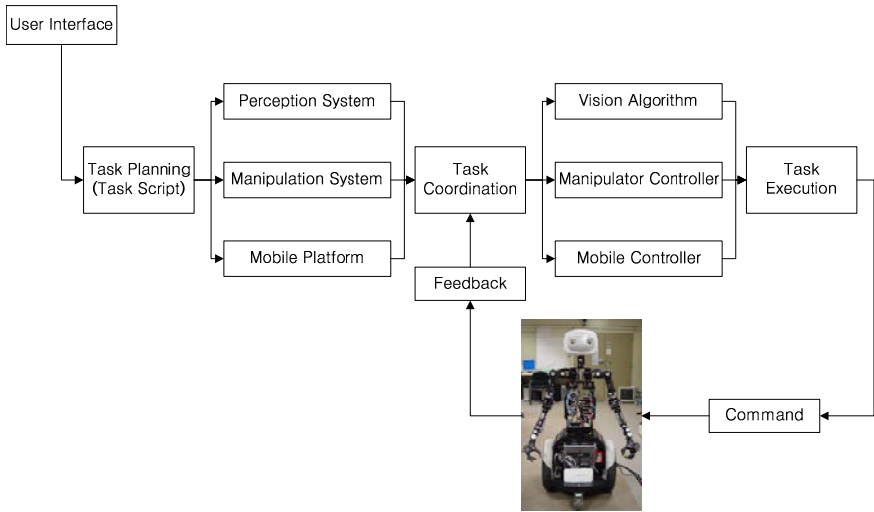
Fig. 1. Framework for coordinated task execution

- The task execution step is characterized by a control method to execute specified commands. We constructed specific modules such as object pose estimation, and a human-like motion generation module, which will be introduced in the following sections.

3 Task Description Method

The execution of a coordinated task requests a technique for an arrangement of the robot actions. To make this possible in various environments and situations, we used an XML script to describe and plan the tasks efficiently. The unit task makes up the operation, and the task script enables this task to be performed with a simple description. There are advantages to this task scrip. It makes it possible to construct the actions of the desired task for the users and it is changed by a simple description. The information obtained by the robot, not a static value, is applied in real time. Moreover, the execution sequence is obviously determined by using the condition of each task. The main elements of the task script are task, action, and condition, which are defined as follows:

Task   Task is a unit to execute a purpose. Unit tasks make up the entire task, which makes it possible for the robot to perform flexibly in various. Each task involves Action and Condition.

Action   Action is the work that is performed when the task is run. Action type is divided into manipulation, mobility, and perception depending on the use of each sub-system. Action is executed according to the parameter, which is described for each sub-system.

Condition   After the unit task is executed, the task to be performed next is decided according to the condition. This plays the role of deciding the direction of flow. For example, the condition decides which task is executed when the former task has succeeded or, when it has failed, it decides whether to terminate the whole task or to execute the same task repeatedly.

The finite-state machine for the task, "Classify Cups and Cans" is illustrated in Fig. 2. It shows the flow of the task. According to the result of each task, the robot can carry out the operation repeatedly, can carry out the next operation, or can terminate the operation according to the condition. If the robot failed to detect the pose of the cup, the robot can execute the next task, "Pose estimation of the can".
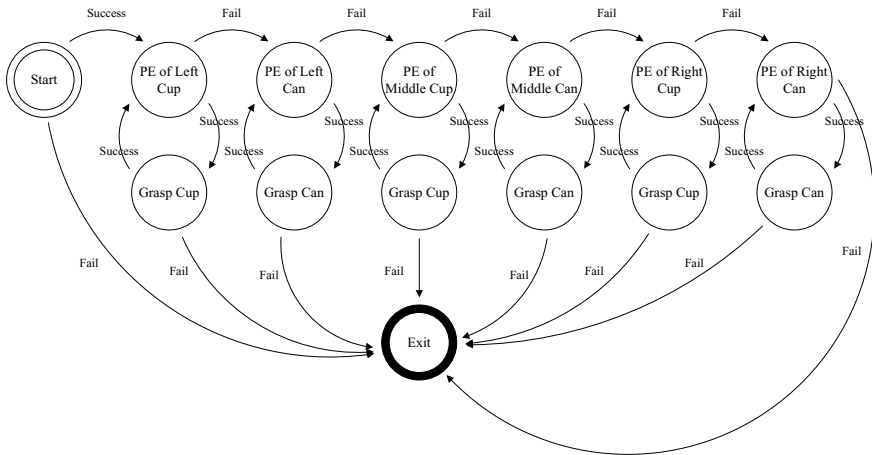
Fig. 2. Finite-State Machine for the task, "Classify Cups and Cans"

4   Main Task Units

In this section, we introduce the main task units which are the object pose estimation, human-like motion for human-robot interaction.

4.1   Pose Estimation for Handling the Object

Detection and tracking of three dimensional objects are important to intelligent service robots in that many of intelligent services are related to dealing with 3D objects. Particle filter has been used successfully for detection and tracking of two dimensional objects [6-9]. In the case that it is applied to 3D objects, the problem of high dimensional search space should be solved. In huge space, search procedure is tedious and it takes a long computational time to find the solution. So, we use a back projection-based sampling where parameters to be estimated are not directly sampled from a parameter distribution, but image points with depth information are sampled first and back-projected to the parameter space. This approach significantly reduces the search space.

The effect of back projection-based sampling is shown in Fig. 3. If the particles are generated directly into the 3D space, the search space forms a hexahedron; but using the back projection-based sampling, it forms surface patches. Various types of objects in daily life can be detected and tracked successfully using the back projection-based sampling (Fig. 4). The result of tracking concerning a cup is shown in Fig. 5.
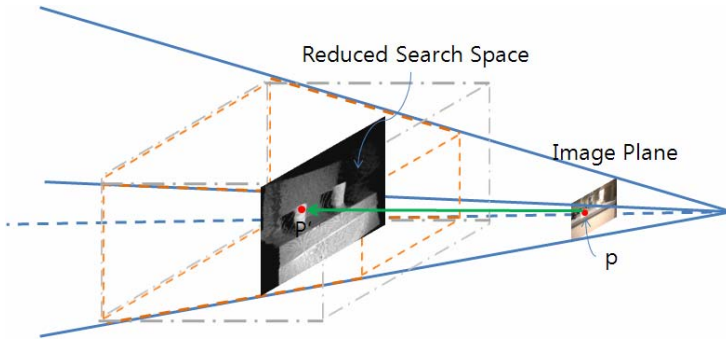
Fig. 3. Reduced Search Space The search space is reduced from the available 3D space to the surface patch by the back projection

(a)   (b)   (c)
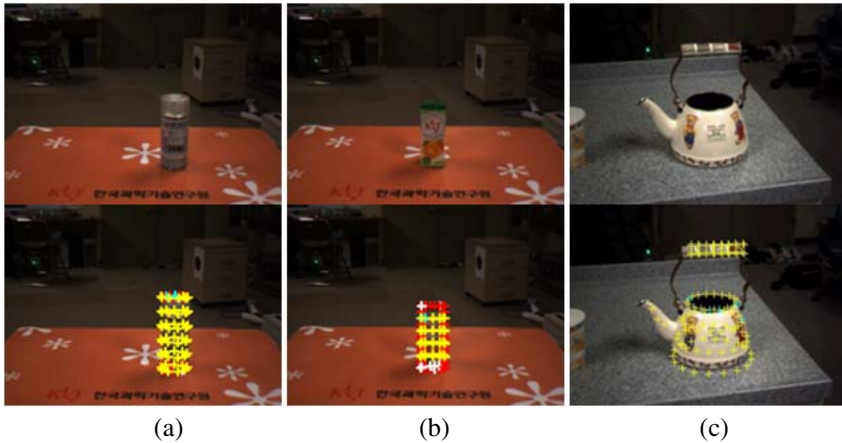
Fig. 4. Various Objects Tracking (a) can, (b) pack, and (c) kettle

4.2 Object Grasping

Our system consists of two systems: perception and control. The perception system recognizes an object, estimates the pose of the object and decides a hand position and palm direction angle to grasp the object; grasping information is then sent to the control system. The control system first solves human-like inverse kinematics problem to generate a grasping posture for a humanoid robot and then generates a smooth arm trajectory from current arm posture to the grasping posture. The smooth motion of the robot can make a human feel familiar and comfortable.
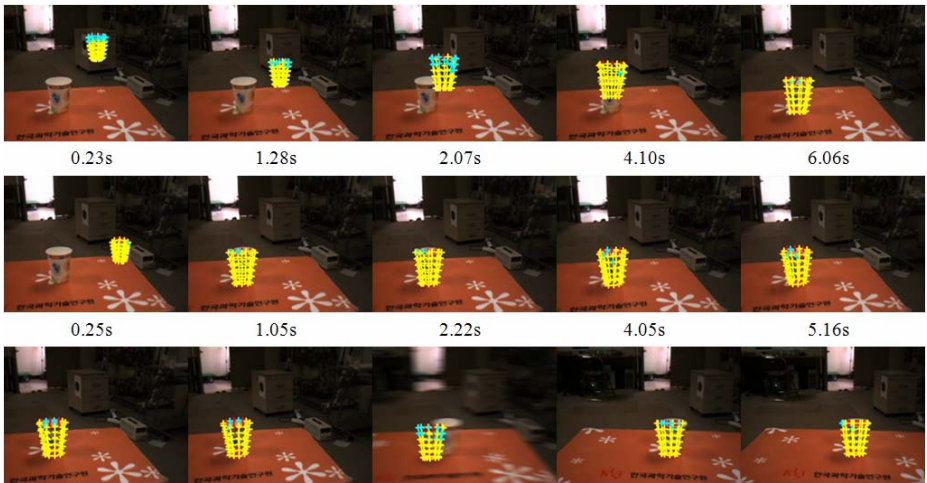
Fig. 5. Tracking of a cup In finding the initial pose, tracking using back projection-based sampling (middle row) is faster than one using pseudo random sampling (top row). And when a camera is moved rapidly, the pose of object is converged to the right pose again (bottom row).
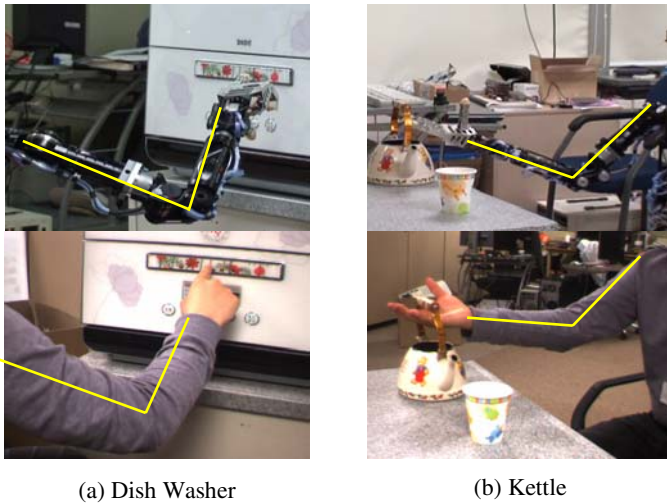
(a) Dish Washer
(b) Kettle

Fig. 6. Human-like arm postures of a robot and a human's arm posture with different types of objects

Human-like Arm Posture for Grasping the Object

When the robot grasps objects of diverse shapes, a human-like arm posture is essential. If the robot grasps each object, which has different grasping point, with the same arm posture, it will fail to grasp the object or be seen awkwardly. Therefore, the arm posture should be different from the grasping point to grasp the object adaptively. The elbow elevation angle (EEA) is employed to make an arm posture look like a human's arm

posture, which was developed by Kim et al. [1]. This method generates a human-like arm posture using a mathematical representation for characterizing human arm motions (Fig. 6). The human arm postures are characterized by EEAs which are determined using the position and orientation of human hands. That representation is mathematically obtained using an approximation tool, the Response Surface Method (RSM). The RSM can generate human-like arm postures for a humanoid robot in real time.

Smooth Motion Generation for User-Friendly Service

The robot is required to generate a smooth motion for user-friendly service. Smooth motion makes a human feel familiar and comfortable, whereas segmented motion is unfamiliar to a human. To generate an arm motion from a current posture to the destination posture, a trajectory generation method is required. We used a spring-damper dynamical system at joint angle space in order to generate the motion.

$$\ddot{\theta} + \frac{c}{m}\dot{\theta} + \frac{k}{m}\left(\theta^d - \theta\right) = 0 \tag{1}$$

where m, c, and k are mass, damper, and spring constant, respectively; $\theta \in \mathbb{R}^n$ is the arm joint angles vector; and, θ^d is the destination arm joint angles vector coming from human-like arm inverse kinematics for a destination posture.

To reach the destination joint angles smoothly without perturbation and overshooting, we designed the constants, the system to be a critical damped system.

$$\frac{c}{m} = 2\omega_n, \qquad \frac{k}{m} = \omega_n{}^2 \tag{2}$$

where ω_n is the natural frequency of the system. The speed to reach the destination joint angles is dependent on the ω_n.

Using this human-like inverse kinematics solver and smooth motion generation method, we were able to generate a human-like arm motion for a humanoid robot.

5   Experiment

We performed a demonstration of the coordinated task using a mobile humanoid robot. In our system, "MAHRU-M" is used as the mobile humanoid robot (Fig. 7). MAHRU-M was developed by the Korea Institute of Science Technology (KIST) and is equipped with 6DOF for each arm, 2DOF for the neck and 4DOF for each hand. A bumblebee stereo camera is mounted on the robot's head to obtain the depth image. All experiments were performed on two different types of PCs: Windows XP, Pentium(R) Core 2 Duo CPU T7700 2.4GHz, 1GB RAM for a perception system and Linux, Pentium(R) M CPU 1.73GHz, 1GB RAM for a control system.

Representative examples of the task such as "Deliver the Pack", "Pour Water from the Kettle" and "Classify Cups and Cans" is shown in Table 1. From among these tasks, we describe the specific sequence of the task, "Classify Cups and Cans", below.

The process of the task, "Classify Cups and Cans" is shown in Fig. 8. First of all, the robot detected the location and orientation of a cup on the left side (a) and the robot arm grasped the cup exactly in the correct position (b). The robot moved the cup

Perception system

Manipulation system
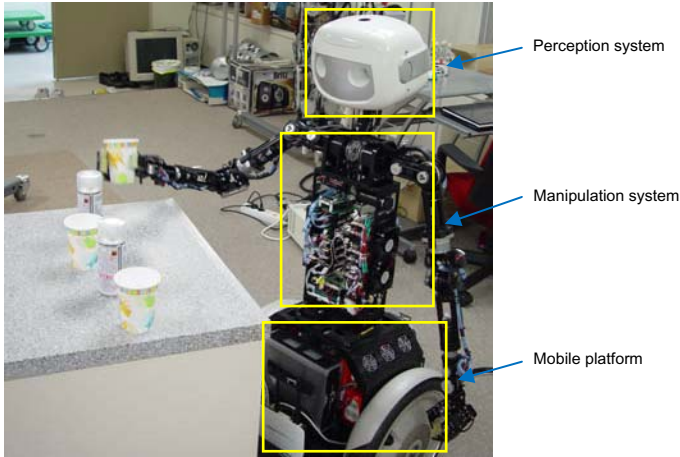
Mobile platform

Fig. 7. Mahru-M a mobile manipulation platform

Table 1. Representative examples of the task

Task	Unit Task	Brief Description
Classify Cups and Cans	1. Pose Estimation of Cup	pose estimation of a cup
	2. Grasp Cup	grasp the cup using a hand
	3. Move Cup	move the cup to the distribution box
	4. Release Cup	release the hand to drop the cup
	5. Pose Estimation of Can	pose estimation of a can
	6. Grasp Can	grasp the can using a hand
	7. Move Can	move the can to the distribution box
	8. Release Can	release the hand to drop the can
Deliver the Pack	1. Pose Estimation of Oven	pose estimation of an oven far away
	2. Move to Oven	move to the oven position closely
	3. Pose Estimation of Button	pose estimation of the oven button
	4. Push Button	push the button to open the oven door
	5. Pose Estimation of Pack	pose estimation of a pack in the oven
	6. Grasp Pack	grasp the pack using a hand
	7. Move to Person	move to person to hand the object
	8. Measure Distance	find person to measure the distance
	9. Turn Head	turn the robot's head to find the person
	10. Find Designated Person	find the designated person to hand over
	11. Hand Object	hand the object to the person
	12. Go to Original Position	go to the original position
Pour Water from the Kettle	1. Pose Estimation of Kettle	pose estimation of a kettle
	2. Grasp Kettle	grasp the kettle using a hand
	3. Pose Estimation of Cup	pose estimation of a cup to be poured
	4. Lift up Kettle	lift up the kettle to pour water
	5. Pour water	pour water from the kettle

(a) Pose Estimation of Cup   (b) Grasp Cup   (c) Move Cup   (d) Release Cup

(e) Pose Estimation of Cup   (f) Pose Estimation of Can   (g) Grasp Can   (h) Pose Estimation of Can

(i) Pose Estimation of Cup   (j) Pose Estimation of Can   (k) Pose Estimation of Cup   (l) Pose Estimation of Can

Fig. 8. The process of the task, "Classify Cups and Cans". The upper images are from the exterior camera and the lower images are from the interior camera.

to the distribution box (c) and released the hand to drop the cup (d). After detecting the cup, the robot tried doing it again (e), to check the possibility of the existence of the cup on the left side.

Using a stereo camera mounted on the robot's head, we precisely calculated 3D information which is used for grasping the object. The grasping position had been already set when modeling, and was obtained when assuming the pose estimation of the object. When we moved the hand of the robot to this position, the robot grasped the pack. When the robot did so, it was essential to control the grasp strength. We had set the strength of grasp experientially according to the object type.

If there were no cups on the left side, then the robot detected the pose of a can on the same side (f). In the same manner, the robot arm grasped the can exactly in the correct position (g) and moved the cup to the distribution box. It was also necessary to check the possibility of the existence of the can on the left side (h). From above procedure, the task for the left side was finished.

The task planner required the accurate position of objects and humanoid robots in a global frame of reference. Furthermore, in order to avoid collision with other objects, we made a series of sub-motions to guide the robot arm on a collision-free path toward the targeted object. We can see arm posture while approaching an object without any collision (b, g).

All procedures were performed in the same manner for the middle and right side (i-l) and we could confirm there was nothing else to do (l). All tasks of "Classify Cups and Cans" were performed successfully. These are actions which can be applied practically to our daily lives.

6 Conclusion

This paper has introduced a coordinated framework for a humanoid robot. Our humanoid robot consists of three sub-systems: a perception system, a mobile platform, and a robotic manipulator. The sub-system should be scheduled according to the task sequence. The task planner scheduled the actions for the diverse sub-systems and the XML script had been presented for describing tasks. We estimated the pose of the object by using the particle filter and therefore the real-time object tracking was possible. The humanoid robot could perform a human-like arm motion by using the arm inverse kinematics solver and smooth motion generation method. In the experiment, we have shown that the task execution by a humanoid robot was successfully performed.

As future work, we plan to extend this description system to a new system which makes possible the operation of applying only the user interface function.

References

1. Kim, S., Kim, C., Park, J.H.: Human-like Arm Motion Generation for Humanoid Robots Using Motion Capture Database. In: Proc. of IEEE/RSJ Int. Conf. on Intelligent Robots and Systems, pp. 3486–3491 (2006)
2. Morales, A., Asfour, T., Osswald, D., Schulz, S., Dillmann, R.: Towards an Anthropomorphic Manipulator for an Assistant Humanoid Robot. In: Robotics: Science and Systems-Workshop on Humanoid Manipulation. MIT, USA (2006)
3. Dillman, R., Becher, R., Steinhaus, P.: ARMAR II – A learning and Cooperative Multimodal Humanoid Robot System. Int. Jour. of Humanoid Robotics 1(1), 143–155 (2004)
4. Yamamoto, Y., Hiroshi, E., Xiaoping, Y.: Coordinated Task Execution of a Human and a Mobile Manipulator. In: Proc. of the IEEE Int. Conf. on Robotics and Automation, vol. 2, pp. 1006–1011 (1996)
5. Extensible Markup Language (XML), `http://www.w3.org/XML`
6. Isard, M., Blake, A.: Conditional Density Propagation for Visual Tracking. Int. Jour. of Computer Vision 29(1), 5–28 (1998)
7. Dubois, C., Davy, M.: Joint Detection and Tracking of Time-Varying Harmonic Components: a Flexible Bayesian Approach. IEEE Trans. on Audio, Speech and Language Processing 15(4), 1283–1295 (2007)
8. Pupilli, M., Calway, A.: Real-Time Camera Tracking Using a Particle Filter. In: Proc. of the British Machine Vision Conference, pp. 519–528 (2005)
9. Najim, K., Ikonen, E., Gomez-Ramirez, E.: Trajectory Tracking Control Based on a Genealogical Decision Tree Controller for Robot Manipulators. Int. Jour. of Innovative Computing and Control 4(1), 53–62 (2008)

Self-stabilizing Human-Like Motion Control Framework for Humanoids Using Neural Oscillators

Woosung Yang[1], Nak Young Chong[2], Syungkwon Ra[1], Ji-Hun Bae[1], and Bum Jae You[1]

[1] Center for Cognitive Robotics Research, Korea Institute of Science and Technology, Seoul, Korea
{wsyang,sykra,joseph,ybj}@kist.re.kr
[2] School of Information Science, Japan Advanced Institute of Science and Technology, Ishikawa, Japan
nakyoung@jaist.ac.jp

Abstract. We propose an efficient and powerful alternative for adaptation of human motions to humanoid robots keeping the bipedal stability. For achieving a stable and robust whole body motion of humanoid robots, we design a biologically inspired control framework based on neural oscillators. Entrainments of neural oscillators play a key role to adapt the nervous system to the natural frequency of the interacted environments, which show superior features when coupled with virtual components. The coupled system allows an unstable system to stably move according to environmental changes. Hence the feature of the coupled system can be exploited for sustaining the bipedal stability of humanoid robots. Also based on this, a marionette-type motion conversion method to adapt captured motions to a humanoid robot is developed owing that there are the differences in capabilities of dynamics and kinematics between a robot and a human. Then this paper discuss on how to stably show human motions with a humanoid robot. We verify that a real humanoid robot can successfully sustain the bipedal stability exhibiting captured whole body motions from various simulations and experiments.

Keywords: Humanoid robot, Neural oscillator, Biologically inspired robot control, Bipedal control, Human-like motion, Imitation.

1 Introduction

Researches on human-like motion have been paid attention to achievement of real or virtual human-like robots in the humanoid robot community for the last decade. Considerable efforts have been mainly devoted to how to solve a highly nonlinear nature of robot dynamics and disturbances from an uncertain environment. In particular, since humanoid robots have a large number of degrees-of-freedom and should maintain the bipedal stability, efficient motion generation and control which the stability problem is considered still remain challenging.

From a practical viewpoint, imitation for human-like motion generation of humanoid robots is considered as a powerful means of enhancing pattern generation competence. Many researchers have studied efficient imitation models to obtain reliable

motion data in noisy stochastic environments [1]-[5]. Especially, Inamura *et al.* devised the mimesis model based on HMM which can imitate the motion of others and abstract the time-series motion patterns as symbol representation [1]. Samejima *et al.* suggested a special framework MOSAIC, where plural dynamics and inverse dynamics are implemented to predict and control motions [4]-[5]. Also, for human motion generation, Yamane and Nakamura computed the interacting dynamics of structure-varying kinematic chains by incorporating an efficient collision/contact model [6]-[7]. Yang *et al*. showed stable bipedal motions of humanoid robots using the proposed SAA through imitation learning [8]-[10].

Although recent progress in imitation learning for human motion generation has yielded notable results in many cases, a unified approach to behavior generation of humanoid robots remains so far largely undetermined. Particularly, only a few approaches have been reported concerning imitation methods that requires effect reproduction through close interaction with the environment. On the other hand, relating to realizing the stable bipedal locomotion, these efforts include such approaches as the zero moment point criterion [11]-[12], the linear inverted pendulum mode [13], virtual model control [14], and biologically inspired approaches [15]-[16].

Several methods do not depend heavily on the predefined reference patterns. However, the motion data captured from humans may not be straightforwardly used owing that these mainly discuss some forms of pattern generation and tracking control for the locomotion of humanoid robots. Moreover, many existing methods require substantial constraints for sustaining the bipedal stability. This causes confliction between the stability problem and imitation motion generation of humanoid robots. Consequently, it is difficult to attain an appropriate human-like motion through imitation with humanoid robots.

This work addresses how to exhibit human-like movements stabilizing bipedal motions. For achieving both of them, the authors propose a marionette-type motion conversion method which helps properly generate human-like motions from captured motion data for humanoid robots. Due to using forward dynamics and virtual force constraint, our motion adaptor shows superior features in mathematical computation and acquisition of more smoothed marker data. In addition, the neural oscillator based self-stabilizer allows bipedal robots to easily embody an appropriate bipedal motion corresponding to various desired the ZMP patterns sustaining the bipedal stability. Also this requires that humanoid robots should maintain the stability even in an unknown disturbance without an additional controller owing to an entrainment property of the neural oscillator. It is verify through simulations and experiments that the proposed approach yields a robust and efficient control of human-like motions.

2   Neural Oscillator Based Control

Our work is motivated by studies and facts of biologically inspired locomotion control employing oscillators. Especially, the basic motor pattern generated by the Central Pattern Generator (CPG) of inner body of human or animal is usually modified by sensory signals from motor information to deal with environmental disturbances. Similar to the sensory system of human or animal, the neural oscillators are entrained with external stimuli at a sustained frequency. They show stability against
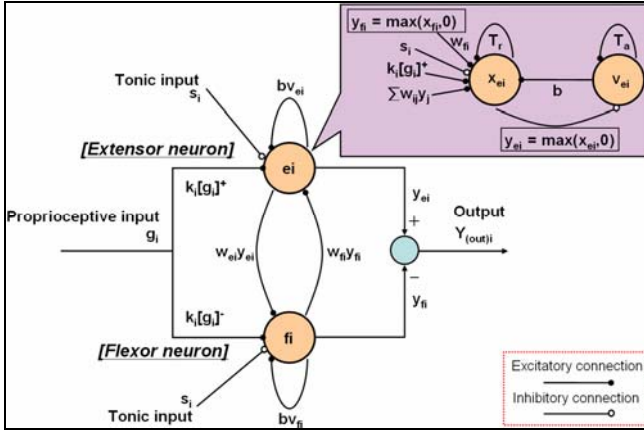
Fig. 1. Schematic diagram of Matsuoka's Neural Oscillator

perturbations through global entrainment between the neuro-musculo-skeletal systems and the ground [17]. Thus, neural oscillators have been applied to the CPG of humanoid robots with rhythmic motions [18]-[19].

$$T_r \dot{x}_{ei} + x_{ei} = -w_{fi} y_{fi} - \sum_{j=1}^{n} w_{ij} y_j - b v_{ei} - \sum k_i [g_i]^+ + s_i$$

$$T_a \dot{v}_{ei} + v_{ei} = y_{ei}$$

$$y_{ei} = [x_{ei}]^+ = \max(x_{ei}, \ 0)$$

$$T_r \dot{x}_{fi} + x_{fi} = -w_{ei} y_{ei} - \sum_{j=1}^{n} w_{ij} y_j - b v_{fi} - \sum k_i [g_i]^- + s_i \qquad (1)$$

$$T_a \dot{v}_{fi} + v_{fi} = y_{fi}$$

$$y_{fi} = [x_{fi}]^+ = \max(x_{fi}, \ 0), \ (i = 1, 2, \cdots, n)$$

where x_{ei} and x_{fi} indicate the inner state of the i-th neuron for $i=1\sim n$, which represents the firing rate. Here, the subscripts 'e' and 'f' denote the extensor and flexor neurons, respectively. $V_{e(f)i}$ represents the degree of adaptation and b is the adaptation constant or self-inhibition effect of the i-th neuron. The output of each neuron $y_{e(f)i}$ is taken as the positive part of x_i and the output of the oscillator is the difference in the output between the extensor and flexor neurons. W_{ij} is a connecting weight from the j-th neuron to the i-th neuron: w_{ij} are 0 for $i \neq j$ and 1 for $i=j$. $w_{ij}y_i$ represents the total input from the neurons arranged to excite one neuron and to inhibit the other, respectively. Those inputs are scaled by the gain k_i. T_r and T_a are the time constants of the inner state and the adaptation effect, respectively, and s_i is an external input with a constant rate. $W_{e(f)i}$ is a weight of the extensor neuron or the flexor neuron and g_i indicates a sensory input.
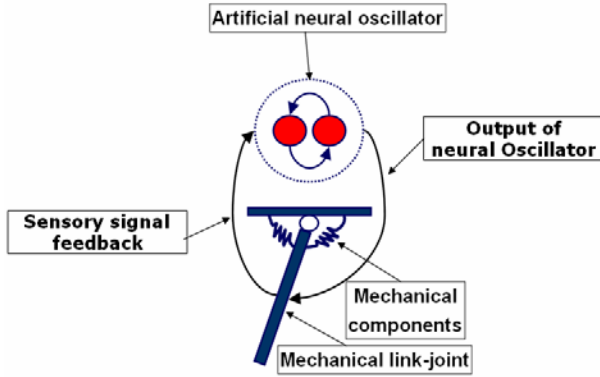
Fig. 2. Mechanical system coupled to the neural oscillator

For implementing the neural oscillator, the oscillator should be coupled to the dynamic system that closely interacts with environments. This enables a robot to adapt to changing environment conditions. For simplicity, we employ a general 2nd order mechanical system connected to the neural oscillator as seen in Fig. 2. The desired torque signal to the joint can be given by

$$\tau = p(\theta_{od} - \theta) - v\dot{\theta}, \tag{2}$$

where p is the position gain, v the velocity gain, θ the actual angle, and θ_{od} the desired angle of the joint, respectively. Specifically, θ_{od} is the output of the neural oscillator that produces rhythmic commands of the joint of the mechanical system. The neural oscillator follows the sensory signal from the joint, thus the output of the neural oscillator may change corresponding to the sensory input. This is what is called "entrainment" that can be considered as the tracking of sensory feedback signals so that the mechanical system can exhibit adaptive behavior interacting with the environment [20]. The key to implementing this method is how to incorporate the input signal's amplitude information as well as its phase information.

3   The Control Model for Bipedal Stability

It would be advantageous if humanoid robots can maintain its stability without using sophisticated controllers. The proposed control approach involves a new application of the interaction between neural oscillators and virtual components to the bipedal stability control of humanoids. This allows humanoid robots to adapt their motions through entrainment responding to unknown disturbances. Now we explain how to embody the stable single support or double phases of humanoid robots corresponding to the lower legged motion generation. In order to maintain stability, the neural oscillator plays an important role by controlling the trajectory of the COM in phase with the ZMP input.

Note that the single support phases bring about significant effects on the landing stability of swing legs that may cause an unexpected perturbation with imperfect

contact. To avoid this, we consider a virtual inverted pendulum model coupled to such a virtual mechanical component as a spring and damper and the neural oscillator, as seen in Fig. 7 for generating an appropriate rolling and pitching motion. The coupled model enables the inverted pendulum to stably move in a frontal and sagittal plane according to a desired ZMP trajectory sustaining the stability. For technically implementing this to a humanoid robot, we simplified the bipedal walking model as a well known linear inverted pendulum model.

Assuming that θ, the inclined angle between the vertical axis and the pendulum in Figs. 3 (a) and (b), is small enough and linearized near 0, the dynamic equations of the coupled inverted pendulum in the pitching and rolling direction are given by

$$\begin{bmatrix} \ddot{x} \\ \ddot{y} \end{bmatrix} = \frac{G}{l} \begin{bmatrix} x - u_x & 0 \\ 0 & y - u_y \end{bmatrix} + \begin{bmatrix} F_x \\ F_y \end{bmatrix}, \tag{3}$$

where x and y are the displacement of the pendulum in the pitching and rolling direction, respectively. l is the length of the pendulum, and u is the position of the massless cart of the pendulum. Note that the subscriptions, 'x' and 'y', indicate the pitching and rolling direction in the entire paper, respectively. G is the gravitational constant. F_x and F_y indicates the force that should be applied to the Center of Mass (COM) of the pendulum in the pitching and rolling direction, respectively.

If the desired ZMP trajectory, u_y, is given in Eq. (3), a stably periodic motion of the COM of the pendulum is generated in terms of the coupled neural oscillator with state feedback [16] (see Fig. 3(c)). If a mechanical system is coupled to the neural oscillator, the dynamic stability is improved [20]. Hence, a stable limit cycle behavior is induced and the periodic COM motion is achieved by the impedance control of the virtual components. Accordingly, F_x and F_y in Eq. (4) are given by

$$\begin{bmatrix} F_x \\ F_y \end{bmatrix} = \frac{1}{ml} \left\{ 2k_s \left(\begin{bmatrix} h_x & 0 \\ 0 & h_y \end{bmatrix} \begin{bmatrix} \theta_{o,x} \\ \theta_{o,y} \end{bmatrix} - \begin{bmatrix} P_x & 0 \\ 0 & P_y \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} \right) - \begin{bmatrix} D_x & 0 \\ 0 & D_y \end{bmatrix} \begin{bmatrix} \dot{x} \\ \dot{y} \end{bmatrix} \right.$$
$$\left. + \begin{bmatrix} i_{p,x} & 0 \\ 0 & i_{p,y} \end{bmatrix} \begin{bmatrix} x_d - x \\ y_d - y \end{bmatrix} - \begin{bmatrix} i_{v,x} & 0 \\ 0 & i_{v,y} \end{bmatrix} \begin{bmatrix} \dot{x} \\ \dot{y} \end{bmatrix} \right\}, \tag{4}$$

where k_s is the stiffness coefficient. h_x and h_y are the output gains of each neural oscillator in the pitching and rolling direction, respectively. P and D are the gains of state feedback, and $i_{p,x}$, $i_{p,y}$, $i_{v,x}$ and $i_{v,y}$ are the gains of the impedance controller. In the proposed controller, $\theta_{o,x}$ and $\theta_{o,y}$ denote the outputs of the neural oscillators coupled to the COM of the inverted pendulum corresponding to the individual direction as illustrated in Eq. (2). x_d and y_d denote the desired ZMP inputs.

The current COM position and velocity of the humanoid robot are obtained again by Eq. (3). For a stable pitching and rolling motion corresponding to an arbitrary ZMP input, F_x and F_y in Eq. (4) are transformed into joint torques using the Jacobian that needs to be applied to each joint of both legs of humanoid robots. For example, as illustrated in Fig. 3 (b), the humanoid robot exhibits stable rolling motion satisfying the desired ZMP. Also the stable pitching motion in Fig. 3 (a) can be attained in the same way.

(a)   (b)

(c)
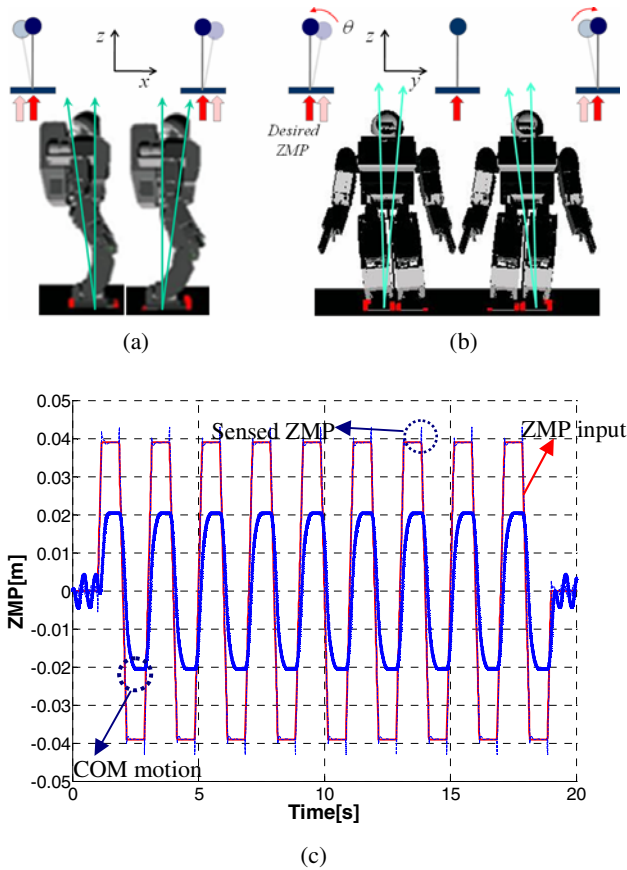
Fig. 3. Simulation result with respect to the stable pitching (a) and rolling (b) motion of the humanoid robot, (c) Plot of the COM motion of humanoid robot in the rolling direction according to the ZMP input

4 Implementation and Basic Tests of the Coupled Model

This section describes how to embody the bipedal motion of a real humanoid robot using the proposed control scheme to guarantee the stability on the single and double support phases. The humanoid robot in Fig. 4, developed by KIST, is used to test and evaluate the proposed controller. It has 35 degrees of freedom (DOF): 6 in each arm, 6 in each leg, 4 in each hand, 2 in neck, and 1 in waist. It's 150cm high and weighs about 67kg. The robot is equipped with a stereo camera, a microphone, four force/torque sensors, and a pose sensor.

We firstly performed numerical simulations to verify the proposed control method. In this simulation, we use the humanoid robot simulator, named *Simstudio*, developed in *SimLab* co. Prior to experiment on human-like whole body motion, we test a few motions during the double support and single support phases, since the analysis of those supporting phases is essential for bipedal walking or other movements. Also, we
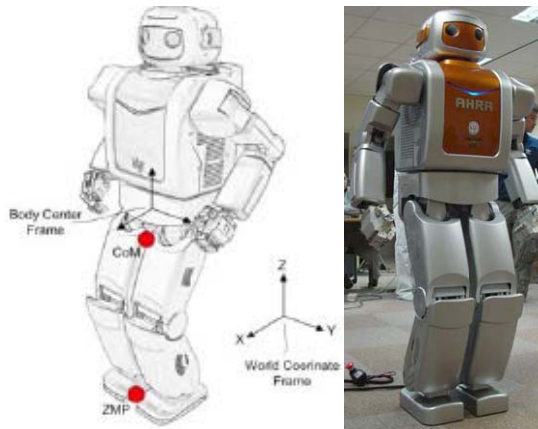
Fig. 4. KIST Humanoid robot

investigate through experiments of following section whether the output of the neural oscillator is adapted to the dynamic motion of the humanoid robot.

After implementing the control algorithm to the *Simstudio*, we acquired the simulation results of the hip positions during rolling and pitching as seen in Figs. 5 and 6. The hip positions are generated appropriately by an arbitrarily defined ZMP input seen in Figs. 5 and 6. This implies that the proposed control method in section 3 works properly. The dashed line in the figures indicates the input pattern of the desired ZMP. The solid lines of Figs. 5 and 6 show the COM of the humanoid robot with regard to the rolling and pitching motions, respectively. Remarkably, the dotted line in Fig. 5 and the dash-dotted line in Fig. 6 are the outputs of the neural oscillators, when the humanoid robot periodically moves in the lateral and sagittal planes. The COM motion of the humanoid robot fed again is considered as the sensory signal of the neural oscillator. Then the outputs of neural oscillator entrain that of the humanoid robot and drive the humanoid robot corresponding to the sensory input appropriately as a torque input. From these results, it can be observed that the neural oscillator causes the self-adapting motion to follow the sensory input. Consequently, we note that this leads the adaptive motion of humanoid robots to maintain the bipedal stability even under an unknown disturbance.

There is the difficulty on how to or when to switch the double support phase and the single support phase under the bipedal locomotion or various humanoid behaviors. To solve this problem, we propose the proper switching rule based on the COM position and the ZMP position. Basically the balancing motion is controlled considering the only COM position. If the projected COM position in the rolling direction moves within the size of the left foot, this indicates that the left leg only supports the whole body of the humanoid robot. On the contrary, in the right leg, the single support phase becomes diametrically opposed to that. In consequence, there is the double support phase when the projected COM position is placed at inner empty space of the size of both foots. Then both legs control the whole body of the humanoid robot.
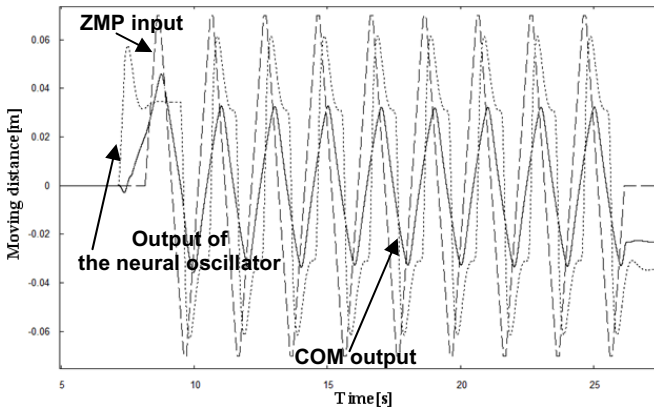
Fig. 5. Input ZMP profile (dashed line), the output of the COM position (solid line) and the output of the neural oscillator in the rolling motion (dotted line)
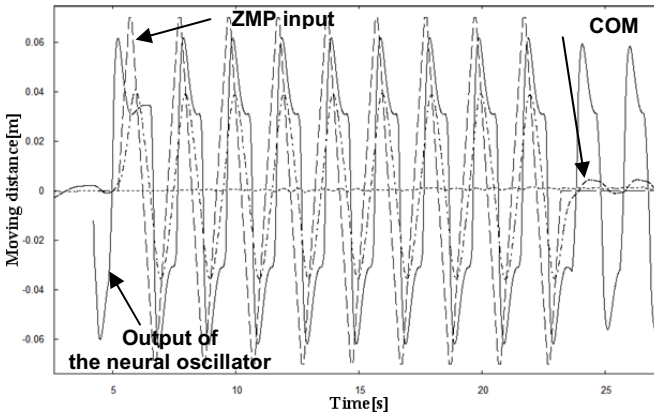
Fig. 6. Input ZMP profile (dashed line), the output of the COM position (solid line) and the output of the neural oscillator in the pitching motion (dash-dotted line)

The lifting motion of the leg is dominated in terms of the calculated ZMP position. If the ZMP position exists within the ZMP criterion of a foot, the corresponding leg can be used as the supporting leg in order to maintain the lifting motion of another leg. For instance, when the calculated ZMP position is inside the ZMP criterion of the right leg, the lifting motion of the left leg becomes possible. However, this criterion is changeable according to the moving velocity of the current ZMP position between the ZMP criterion of a foot and that of another foot. Here we should establish the condition to evaluate the performance of the bipedal walking control based on the neural oscillator coupled virtual model. The motion for keeping balance of the humanoid robot can be yielded properly in terms of the coupled model. In the simulation seen in Fig. 7, we verified the smooth and natural lifting motion regardless of the double or single support phase.
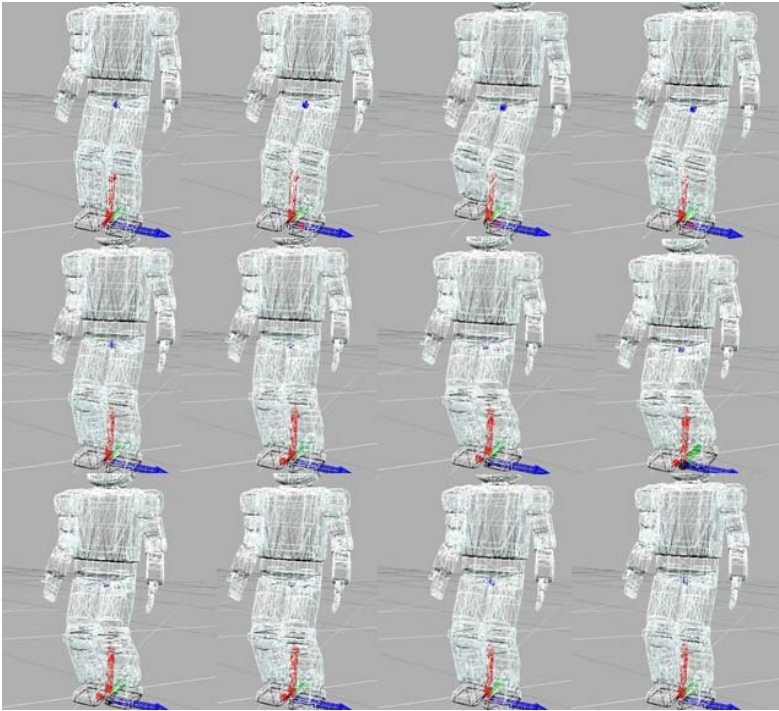
Fig. 7. Simulation result on the lifting leg motion with the transition velocity of the ZMP, 0.4cm/s

5   Experimental Verification on Human-Like Whole Body Motion

5.1   Human-Like Whole Body Motion Generation

Figure 8 describes the conceptual virtual force based motion adaptor for the motion generation of a humanoid robot's upper body. If there is a desired motion data corresponding to each joint of the robot, the data is properly changed exploiting the marionette-type motion conversion method instead of a marker position data obtained from a motion capture system. However, in experiment, we employed a marker position data extracted from our motion capture system owing that it is difficult to create unified motions such as human-like motions. The individual joint angles and the Cartesian COM position are obtained by solving the respective forward kinematics problems. And then the ZMP position is acquired by calculating the measured data in terms of the force plate. The self-stabilizing bipedal controller drives the humanoid robot at an appropriate equilibrium point corresponding to a time-varying ZMP input, COM input and other joint motion.

Although recorded human motions can be a good means for teaching various movements to robots, they may not be applicable directly to a robot because of such geometrical and systemic differences between a human and a robot as length, degrees of freedom, and kinematics and dynamics capabilities. A marionette is a type of
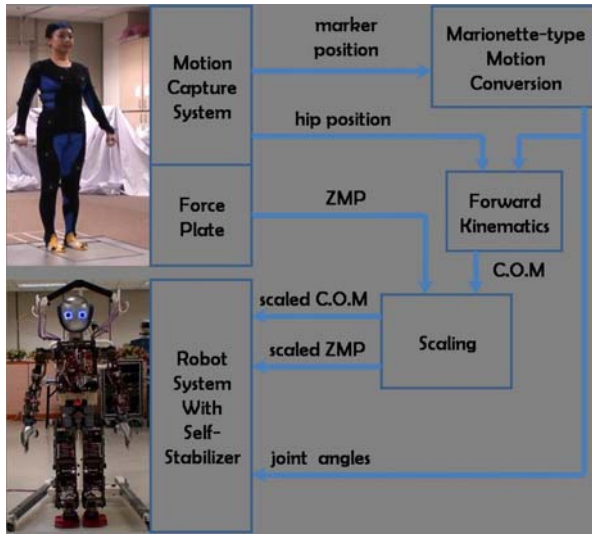
Fig. 8. Conceptual figure with respect to virtual force based motion adaptor

puppets with strings driven by a human as seen in Fig. 9. The feature of a marionette is that it is a system with passive joints and is moved by the force of strings. Applying this principle, given a robot with all passive joints, we virtually connect the position of a human hand to that of the end-effector of the robot with an elastic strip. Then the robot follows the movement of a human arm. The movement of the robot is calculated by following forward dynamics equation.

$$M(q)\ddot{q} + V(q,\dot{q}) + G(q) + J^{T}F_{v} = \tau \tag{5}$$

where q is the joint angle vector, M is the mass matrix, V is the Coriolis vector, G includes the gravity force, τ is the joint torque vector, and J is the Jacobian matrix. As the robot has only passive joints, the joint torque is zero. F_{v} is the virtual force caused in terms of the elastic strip and the external force of dynamics equation.

Fig. 9. A typical marionette controlled by strings

An elastic strip is modeled with a virtual spring and damper between the hand positions of a human and a robot. The virtual force created by elastic strip is calculated as follow:

$$F_v = k_p(x_h - x_r) + k_d(\dot{x}_h - \dot{x}_r) \tag{6}$$

where x_h, $\dot{x}_h$, x_r, and $\dot{x}_r$ are the positions and velocities of the human hand and the end-effector of a robot, respectively. They are the 6-dimensional vectors including the rotation information. k_p is the spring stiffness and k_d is the damping coefficient. We are able to tune the conversion phase by changing them.

One issue of the proposed method is the effect of gravity force. When the elastic strip draws a robot at the upper position, the robot moves less than at the lower position because of gravity force. So we introduce the gravity compensation term in the joint torque instead of zero vector as follow:

$$\tau = \hat{G}(q) \tag{7}$$

where $\hat{G}$ is the computed gravity force, which is ideally equal to G in Eq. (5) ideally.

Another issue is the joint constraint to prevent the damage of the robot. Hence, the proposed motion conversion method includes constraint conditions with respect to the individual joint angles. For this, the virtual force constraint is involved as seen in Fig. 10. If a joint has the limit q_{max} and q_{min}, the virtual force works as the joint angle go near its limit. The force increases exponentially when the joint angle approaches its limit.
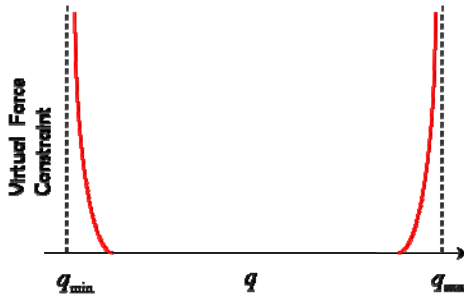
Fig. 10. Virtual force constraint

5.2 Experimental Verification

Figure 11 shows the experimental results on the motion conversion between a human and a robot. It can be observed from the results that four representative motions of a human are well converted into a robot motion. The motion conversion process is operated fast over 100 Hz in real time. Also the virtual force led by the virtual mechanical components of the elastic strip plays on roll in noise filtering of the measured marker data such as an impedance model. So the resultant motions show smooth human-like motions.

The authors implemented the proposed control method to the real humanoid robot to verify the validities of the proposed motion adaptor with the self-stabilizing bipedal
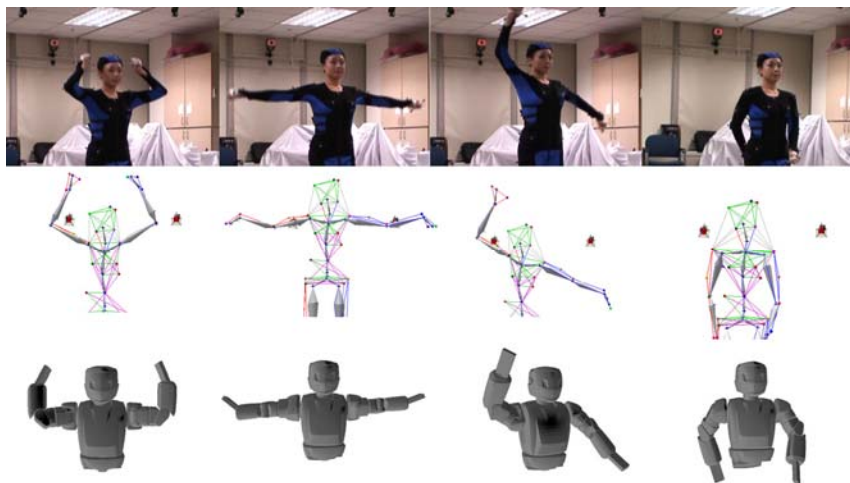
Fig. 11. The motion conversion procedure employing the marionette-type motion conversion method. The upper figures are human motion. The middle and lower figures are marker set and robot motion, respectively.

Fig. 12. Snap shots on the whole body motion and balancing test of humanoid robot under unknown disturbances, Dotted circles indicate the dumbbells employed as unknown disturbances

controller. The kinematic and dynamic properties of a robot such as the width of feet, the area of foot, the length and mass of leg are different from those of a human. So, considering the calculated COM, we adjust the ZMP measured from the human motion so that the bipedal stability of the humanoid robot is maintained within the stable ZMP range. It is done by just multiplication of suitable scale factor. And the modified ZMP is input to the self-stabilizing bipedal controller. Then the stable COM corresponding to the ZMP input is autonomously generated in terms of the self-stabilizing balance controller as illustrated in Fig. 3 (c). Under the dancing motion of the humanoid robot to employ the motion data obtained from the motion conversion method, unknown disturbances composed of two dumbbells of weigh 10kg are applied directly to the humanoid robot in the frontal and sagittal sides, respectively, as seen in Fig. 12. It is observed from the experiment that the humanoid robot is able to maintain the bipedal stability. Hence even though the external disturbances of a sinusoidal form effect to the humanoid robot, the humanoid robot coupled to the neural oscillator and virtual components can stably exhibit a novel adaptive motion corresponding to an unknown disturbance.

6 Conclusion

We have presented a new control framework for generating human-like motion of a humanoid robot. For achieving this, the end-effector of each limb of a humanoid robot is virtually connected to an elastic strip that allows a robot to exhibit human-like motion corresponding to motion captured from humans incorporating virtual forces. In addition, to attain a stable bipedal motion of humanoid robots, a new control method consisting of neural oscillators and virtual components was developed. The COM position was controlled to follow the time-varying desired ZMP input sustaining the bipedal stability, which enables a humanoid robot to maintain the bipedal stability regardless of the single or double phases even under unknown disturbances. Extensive simulations and experiments were carried out to verify the effectiveness of the proposed method. More experiments are currently under way for further evaluation of the proposed control method.

References

1. Inamura, T., Toshima, I., Tanie, H., Nakamura, Y.: Embodied Symbol Emergence Based on Mimesis Theory. The Int. Journal of Robotics Research 23, 363–377 (2004)
2. Ijspeert, A.J., Nakanishi, J., Schaal, S.: Movement Imitation with Nonlinear Dynamical Systems in Humanoid Robots. In: Proc. IEEE Int. Conf. on Robotics and Automation, pp. 1398–1403 (2002)
3. Schaal, S.: Is Imitation Learning the Way to Humanoid Robots? Trends in Cognitive Science 3, 233–242 (1999)
4. Samejima, K., Katagiri, K., Doya, K., Kawato, M.: Symbolization and Imitation Learning of Motion Sequence Using Competitive Modules. Trans. of the Institute of Electronics, Information, and Communication Engineers J85-D-II, 90–100 (2002)
5. Samejima, K., Doya, K., Kawato, M.: Inter-module Credit Assignment in Modular Reinforcement Learning. Neural Networks 16, 985–994 (2003)

6. Nakamura, Y., Yamane, K.: Dynamics Computation of Structure-varying Kinematic Chains and Its Application to Human Figures. IEEE Trans. on Robotics and Automation 16, 124–134 (2000)
7. Yamane, K., Nakamura, Y.: Dynamics Filter – Concept and Implementation of Online Motion Generator for Human Figures. IEEE Trans. on Robotics and Automation 19, 421–432 (2003)
8. Yang, W., Chong, N.Y.: Goal-directed Imitation with Self-adjusting Adaptor Based on a Neural Oscillator Network. In: Proc. Int. Conf. on Advanced Robotics, pp. 404–410 (2005)
9. Yang, W., Chong, N.Y., Kim, C., You, B.J.: Locomotion Imitation of Humanoid Using Goal-directed Self-adjusting Adaptor. In: Proc. IEEE/RSJ Int. Conf. on Intelligent Robots and Systems, pp. 5650–5656 (2006)
10. Yang, W., Chong, N.Y.: Imitation Learning of Humanoid Locomotion Using the Direction of Landing Foot. Int. Journal of Control, Automation and Systems (2009) (to be published)
11. Vukobratovic, M., Juricic, D.: Contribution to the Synthesis of Biped Gait. IEEE Trans. Bio-Med. Eng. BME-16, 1–6 (1969)
12. Vukobratovic, M.: How to Control Artificial Anthropomorphic Systems. IEEE Trans. Syst., Man, Cybern. SMC-3, 497–507 (1973)
13. Kajita, S., Matsumoto, O., Saigo, M.: Real-time 3D Walking Pattern Generation for a Biped Robot with Telescopic Legs. In: Proc. IEEE Int. Conf. on Robotics and Automation, pp. 2299–2306 (2001)
14. Pratt, J., Chew, C.-M., Torres, D.A.P., Pratt, G.: Virtual Model Control: An Intuitive Approach for Biped Locomotion. The Int. Journal of Robotics Research 20, 129–143 (2001)
15. Endo, G., Nakanishi, J., Morimoto, J., Cheng, G.: Experimental Studies of a Neural Oscillator for Biped Locomotion with QRIO. In: IEEE Int. Conf. on Robotics and Automation, pp. 598–604 (2005)
16. Yang, W., Chong, N.Y., Kim, C., You, B.J.: Self-adapting Humanoid Locomotion Using a Neural Oscillator Network. In: Proc. IEEE/RSJ Int. Conf. on Intelligent Robots and Systems, pp. 309–316 (2007)
17. Taga, G., Yamagushi, Y., Shimizu, H.: Self-organized Control of Bipedal Locomotion by Neural Oscillators in Unpredictable Environment. Biological Cybernetics 65, 147–159 (1991)
18. Taga, G.: A Model of the Neuro-Musculo-Skeletal System for Human Locomotion. I. Emergence of basic gait. Biological Cybernetics 73, 97–111 (1995)
19. Taga, G.: A Model of the Neuron-musculo-skeletal System for Human Locomotion. II. Real-time Adaptability under Various Constraints. Biological Cybernetics 73, 113–121 (1995)
20. Yang, W., Chong, N.Y., Kim, C., You, B.J.: Entrainment-enhanced Neural Oscillator for Rhythmic Motion Control. Journal of Intelligent Service Robotics 1, 303–311 (2008)

A New Low-Cost Eye Tracking and Blink Detection Approach: Extracting Eye Features with Blob Extraction

Ibrahim Furkan Ince[1] and Tae-Cheon Yang[2]

[1] Graduate School of Digital Design, Kyungsung University
Busan, Republic of Korea
[2] Department of Computer and Information Science, Kyungsung University
Busan, Republic of Korea
`{furkan,tcyang}@ks.ac.kr`

Abstract. The systems let user track their eye gaze information have been technologically possible for several decades. However, they are still very expensive. They have limited use of eye tracking and blink detection infra-structure. The purpose of this paper is to evaluate cost effects in the sector and explain our new approach in detail which reduces high costs of current systems apparently. This paper introduces an algorithm for fast and sub-pixel precise detection of eye blobs for extracting eye features. The algorithm is based on differential geometry and still exists in OpenCpV library as a class. Hence, blobs of arbitrary size that means eye size can be extracted by just adjusting the scale parameter in the class function. In addition, center point and boundary of an eye blob, also are extracted. These describe the specific eye location in the face boundary to run several algorithms to find the eye-ball location with its central coordinates. Several examples on real simple web-cam images illustrate the performance of the proposed algorithm and yield an efficient result on the idea of low-cost eye tracking, blink detection and drowsiness detection system.

Keywords: Eye Tracking, Blink Detection, Drowsiness Detection, Eye Blob Extraction.

1 Introduction

There have been so many computer vision research dedicated to the implementation of eye tracking systems which were designed to detect eye movements and gestures so far. Such systems have been developed to provide a way of user interface controlling for people with disabilities or limited motor skills especially. However, most of them were expensive to use and not practical. Therefore, this paper aims to present an inexpensive, low-cost and robust eye tracking methodology for disabled people to interact with simple computer applications in a meaningful way that requires minimal effort.

A variety of eye trackers based on image processing have been described in the literature. Deng et al. [1] presented a region-based deformable template method for locating the eye and extracting eye features. A system based on a dual state model for tracking eye features is proposed in [2]. Both of these approaches require manual

initialization of eye location. A blink detection algorithm for human computer interaction has been proposed by Morris [3], in which the initialization step requires motion detection to locate the eyelids. Any other significant movement on the face, such as that induced by speaking, may cause the system to fail in detecting eyelids. De la Torre et al. [4] describes a similar system for driver warning based on principal component analysis (PCA) and active appearance models (AAM), which requires a training period for each user. Although this method is robust to translation and rotation, it may fail if the face view is not fully frontal. Recently, a system based on feature detection using one camera has been suggested by Smith [5]. Eye and gaze tracking based on pupil detection using infra-red (IR) illumination with special hardware have been proposed by several authors [6], [7]. These methods are robust in indoor lighting conditions while the performance may be degraded in direct sunlight. There still have not been many blink detection related systems designed to work with inexpensive USB webcams [8, 9]. There have, however, been a number of other feature detection systems that use more expensive and less portable alternatives, such as digital and IR cameras for video input [10, 11, 12, and 13]. Aside from the portability concerns, these systems are also typically unable to achieve the desirable higher frame rates of approximately 30 fps that are common with USB cameras. The other part of the issue is that most of the eyes tracking systems are using template matching methods. However, it costs very high in software engineering paradigm.

In this paper, we propose a new method for locating the eye and eye feature extraction in order to overcome some of the shortcomings of earlier methods. We describe a novel eye locating algorithm based on face geometry and number of blobs detected in facial area that is robust to scaling and translation. In our project, we used blob extraction algorithm supplied from OpenCV library. As an input device, we used a simple 2 megapixel resolution webcam without Infra-Red and Processing as a SDK. No other stuff was used and it yielded very good results. In subsequent sections, we describe the basic methodology of our eye tracking approach that is blob extraction, suggested algorithm and experimental results.

2   Blob Detection and Extraction

Blob detection is a fast and sub pixel precise detection of small, compact image primitives called as "blobs". The algorithm is based on differential geometry and incorporates a complete scale-space description. Hence, blobs of arbitrary size can be extracted by just adjusting the scale parameter. In addition to center point and boundary of a blob, also a number of attributes are extracted. Blob extraction is an image segmentation technique that categorizes the pixels in an image as belonging to one of many discrete regions. Blob extraction is generally performed on the resulting binary image from a thresholding step. Blobs may be counted, filtered, and tracked [14].

OpenCV blob extraction library finds many blobs; however, purpose of the system and proposed algorithm is to get only eye blob among all the blobs. Consequently, eye blob area is the certain location of the eye and yields eye features as well.

Fig. 1. Numerous blobs were extracted from the source image by using OpenCV Blob Extraction library

3   Suggested Algorithm

After mentioning about the preliminaries and fundamentals of image processing issues, we can step into the algorithm that we have applied. It can be shown step by step as follows:

1. Get the camera image at 320 X 240 resolutions and convert the image into black and white colors with a given threshold value. The default value is 60 and it can vary with respect to light in the environment manually by the user. After this operation, the image is in white and black pixel colors. This operation is required for edge detection algorithm that will be mentioned in further steps.
2. Detect the face with OpenCV Haar_Do_Canny_Pruning Face Detection library.
3. Get the detected faces into one-dimensional array. Each face has x-position, y-position, width and height property. Then, draw the face rectangle onto the image. Because eye ball recognition will be performed by Circular Hough Transform algorithm and it requires a radius size as a parameter, a varying radius size has to be set with respect to the face size. For instance, if the user keeps away from the screen, radius must be smaller. On the contrary, if the user gets closer to the camera, radius has to be larger than before. In order to fix this point, radius has to be a varying variable depending on the face size because face also becomes larger when the user gets closer to the camera and smaller gets apart from the camera. Hence, radius is proportional with face size.
4. Give a threshold value for high-pass filtering. Because eye ball will be recognized and eye ball is a perfect circle, an edge detection algorithm must be run for detecting the circles in the certain image area. The certain image area will be the eye-blob area that will be detected in further steps. Thresholding and its value is given for edge detection. A Laplacian edge detection algorithm is used with 3 X

3 convolution matrix or kernel. Now, the image is filtered by high-pass and it shows the clear edges of the image.

5. OpenCV Blob Extraction algorithm is run. OpenCV function returns one-dimensional blobs array and it has the x-position, y-position, width and height properties of the blob rectangles. In order to run the blob detection function, certain parameters should be entered to the function. First parameter is the smallest area in the image, second parameter is the largest area in the image, and third parameter is the number of blobs. The last two parameters are optional parameters. In addition; first parameter and second parameter are also dependent on radius of the eye because eye blob area is also changing when user get closer to the camera or the reverse. In our algorithm, we put the number of blobs as 100000 because we need maximum amount of blobs in a certain area. This algorithm is not running in the entire image. It runs in a certain area that is determined by us with our additional algorithm inside the class. If the blob rectangle is at the same location of right eye, then return the blob into the array. In this step, the location of the right eye is given roughly with respect to universal biological properties. Hence, blob detection algorithm returns only one blob location and size in each loop in the whole algorithm. That blob is called eye blob and it is the estimated location of the right-eye.

6. Circular Hough Transform algorithm is run at this step. The area where this algorithm will be run, the radius and quantity of the circles to be found are certain at this moment. First, it must be declared that the searching of circles should be in the area of eye blobs. The number of circles is only 1 because one eye is enough for getting eye recognition and eye gaze estimation. We preferred right eye for this purpose.

7. After Circular Hough Transform function started to search for eye circles in the eye blob area, eye balls are detected and eye ball circle is drawn onto the image. In order to recognize eye blink detection and drowsiness detection, all the pixels in this circle are filled with a color. We chose blue color for this purpose. And the vertical number of blue pixels are counted in the loop.

8. If ratio of the vertical number of blue pixels to the eye size (2 X Radius) is less than 0.3, then it means that this is an eye blink. If ratio of the vertical number of blue pixels to the eye size (2 X Radius) is less than 0.3 and it insists at least 5 frames (1/6 second at 30 fps), then it means that this is a drowsy eye.

4 Experimental Results

The system was primarily developed and tested on a Windows XP PC with an Intel Pentium IV 2.8 GHz processor and 1 GB RAM. Video was captured with a Logitech Quickcam Pro 4000 webcam at 30 frames per second. All video was processed as binary images of 320 X 240 pixels using various utilities from the Intel OpenCV library [15]. We made experiments using 8 users to evaluate the performance of the proposed system. 6 of them were Korean and 2 of them were Turkish. 3 of the users were using eye glasses and experiment was held with their eye glasses as well. Experiments showed that our proposed system detects eye ball, blink and drowsiness even though users were wearing eye glasses as shown in the Figure 3 and 4. The

highest success rate of the proposed system to detect the eye ball, blink and drowsiness was 96% by moving the head every 20 pixels to the right and 20 pixels up and the success rate was 80% even for the worst moving. Figure 2,3, and 4 show examples of the images for which the proposed system gave successful results. The error occurred when the captured image was far away from the entire head image. If the images were taken far, the captured image tends to be small and difficult to detect. However, proposed system worked accurately with the distance up to 1.5m as seen in Figure 4.
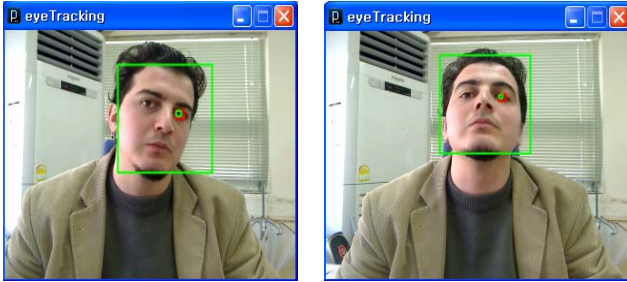
Fig. 2. Moving the head towards right and up hand side respectively. Eye tracking has been achieved successfully again. (Red rectangle is the eye blob).
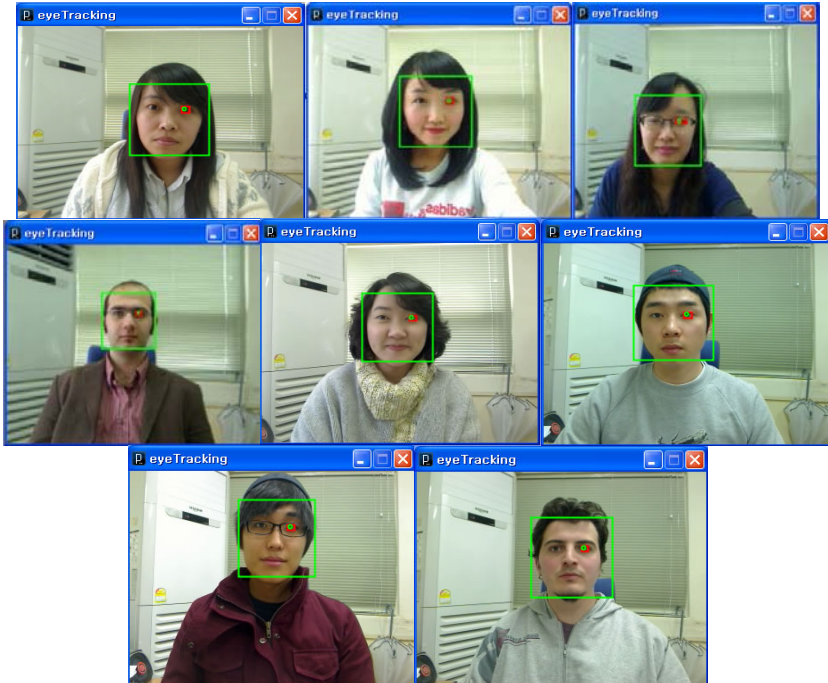
Fig. 3. Sample frames from sessions for each of the eight test subjects

Table 1. Test of the system in different conditions

Shifted Head View	Far View (1.5m)	Bright Environment	Dark Environment

Fig. 4. Sample frames from sessions testing varying head positions, distance to camera and lighting conditions. The system still works accurately with shifted head position, test subject distance up to 1.5m, and exceedingly bright and dark environments.

5   Discussion and Conclusions

In this paper, the factors contributing to the high cost of eye-tracking systems were reported and proposed ideas that may help to reduce the cost of these systems were presented. Using a very simple web-cam, OpenCV and Processing is the only cost of the proposed system. Because OpenCV and Processing are open-source library and

SDK, the only cost is the simple web-cam. In this research the eye location can be detected but some constraint had been exposed. The separability filter tends to detect the possible coordinates of the eye blob. Then, using the CHT, the obtained coordinates that are not considered as circle will be eliminated. The main constraint is the size of the circle. A certain circle radius is used to assign the radius of the eye largeness. We made experiments using 20 images to evaluate the performance of the proposed system. As the results of the experiments, the highest success rate of the proposed system to detect eye was 96% by moving the head every 20 pixels to the right and 20 pixels up and the success rate was 80% even for the worst moving. The problem arises if the user image is captured far away and the image tends to be small. Beside that, some of the eye blobs are in a bunch and overlaps between each other. These make the detection method difficult in carrying out the task. One such consideration is the performance of the system under different lighting conditions (see Figure 4). The experiments indicate that the system performs equally well in extreme lighting conditions (i.e with all lights turned off, leaving the computer monitor as the only light source, and with a lamp aimed directly at the video camera). The accuracy percentages in these cases were approximately the same as those that were retrieved in normal lighting conditions. However, as lightining conditions change, the thresholding value also has to be modified throughout the system. We adjusted threshold value manually in the experiments. It is possible to run an automatic thresholding value with respect to the lighting conditions, however; we plan to do it as a future work. Another important consideration is the orientation of head in the captured image. Images taken and tested by shifting the head of users around 45 degrees and results show that system again detects the eyeball accurately (see Figure 4). And also images taken far from the camera were tested. Proposed system worked fine and accurately in terms of detecting the eye ball and blinks up to 1.5m away from the camera (see Figure 4). Some tests were also conducted with users wearing glasses (see Figure 3 and 4), which indicated that glasses are not a limitation of the proposed system. Because the system works with a binary image in a certain threshold value, the eye glasses do not appear in the black and white image. Experiments showed that system works accurately with the users who wear glasses as well (see Figure 3 and 4). In this research, automatic eye recognition can be used to assist the eye blink and drowsiness detection process. Future work in progress includes detection of several eye blobs in a bunch and overlapping between each other and automatic thresholding.

References

1. Deng, J., Lai, F.: Region Based Template Deformation and Masking for Eye-feature Extraction and Description. Pattern Recognition 30(3), 403–419 (1997)
2. Tian, Y., Kanade, T., Cohn, J.: Dual-state Parametric Eye Tracking. In: Proc. of Conf. on Automatic Face and Gesture Recognition, pp. 110–115 (2000)
3. Morris, T., Zaidi, F., Blenkhorn, P.: Blink Detection for Real-time Eye Tracking. J. Networking and Computer Applications 25(2), 129–143 (2002)
4. De la Torre, F., Garcia Rubio, C., Martinez, E.: Subspace Eyetracking for Driving Warning. In: Proc. Of ICIP, vol. 3, pp. 329–332 (2003)
5. Smith, P., Shah, M., Da Vitoria Lobo, N.: Determining Driver Visual Attention with One Camera. Trans. on Intelligent Transportation Systems 4(4), 205–218 (2003)

6. Ji, Q.: Face Pose Estimation and Tracking from a Monocular Camera. Image and Vision Computing 20(7), 499–511 (2002)
7. Ruddarraju, R., Haro, A., Essa, I.: Fast Multiple Camera-head Pose Tracking. In: Proceedings VI (2003)
8. Gorodnichy, D.O.: On Importance of Nose for Face Tracking. In: Proceedings of the IEEE International Conference on Automatic Face and Gesture Recognition (FG 2002), pp. 188–196, 20–21 (2002)
9. Gorodnichy, D.O.: Second Order Change Detection, and its Application to Blink-controlled Perceptual Interfaces. In: Proceedings of the IASTED Conference on Visualization, Imaging and Image Processing (VIIP 2003), pp. 140–145, 8–10 (2003)
10. Bhaskar, T.N., Keat, F.T., Ranganath, S., Venkatesh, Y.V.: Blink Detection and Eye Tracking for Eye Localization. In: Proceedings of the Conference on Convergent Technologies for Asia-Pasific Region (TEN-CON 2003), pp. 821–824 (2003)
11. Moriyama, T., Kanade, T., Cohn, J.F., Xiao, J., Ambadar, Z., Gao, J., Imamura, H.: Automatic Recognition of Eye Blinking in Spantaneously Occuring Behavior. In: Proceedings of the International Conference on Pattern Recognition (ICPR 2002), vol. IV, pp. 78–81 (2002)
12. Rurainsky, J., Eisert, P.: Eye Center Localization Using Adaptive Templates. In: Proceedings of the CVPR Workshop on Face Processing in Video, FPIV 2004 (2004)
13. Wei, X., Zhu, Z., Yin, L., Ji, Q.: A Real-time Face Tracking and Animation Ssystem. In: Proceedings of the CVPR Workshop on Face Processing in Video, FPIV 2004 (2004)
14. Horn, B.K.P.: Robot Vision, pp. 69–71. MIT Press, Cambridge (1986)
15. OpenCV Library, `http://sourceforge.net/projects/opencvlibrary`

Vehicle Detection Algorithm Using Hypothesis Generation and Verification

Quoc Bao Truong[1] and Byung Ryong Lee[2]

[1] Department of Mechanical and Automotive Engineering, University of Ulsan, Ulsan Korea
baotruong@mail.ulsan.ac.kr
[2] School of Mechanical and Automotive Engineering, University of Ulsan, Ulsan Korea
brlee@ulsan.ac.kr

Abstract. In this paper, we present a two-stage vision-based approach to detect front and rear vehicle views in road scene images using eigenspace and a support vector machine for classification. The first stage is hypothesis generation (HG), in which potential vehicles are hypothesized. During the hypothesis generation step, we use a vertical, horizontal edge map to create potential regions where vehicles may be present. In the second stage verification (HV) step, all hypotheses are verified by using a Principle Component Analysis (PCA) for feature extraction and a Support Vector Machine (SVM) for classification, which is robust for both front and rear vehicle view detection problems. Our methods have been tested on different real road images and show very good performance.

Keywords: Autonomous Vehicle, Vision-based, Hypothesis Generation (HG), Hypothesis Verification (HV), Principle Component Analysis (PCA), Support Vector Machine (SVM).

1   Introduction

Traffic accident statistics show that the main threat facing drivers is other vehicles. It is one of most importance to detect preceding vehicles in real time in order to maintain a safe distance and avoid collision.

A vehicle or detection algorithm must also distinguish vehicles or cars from other visual patterns, such as similar looking rectangular objects. In our case, we focus on detection via the views of the front and rear of a moving vehicle which precedes our vehicle on the road.

A common approach for the hypothesis generation step in monocular vision is shifting the search window over an input image and categorizing the vehicle in the window with a classifier [1]. Each vehicle is designated on an image window with a specific size. To speed up the system without losing classification performance, we need to perform the HG step automatically and as quickly as possible. Based on this requirement, we use the two-stage scheme approach as presented in [2] for our vehicle detection problems. First, all regions in the image that potentially contain target vehicles are identified. This we call the *"Focus attention mechanism"*. Next, the selected regions are verified by a classifier as illustrated in Figure 1.
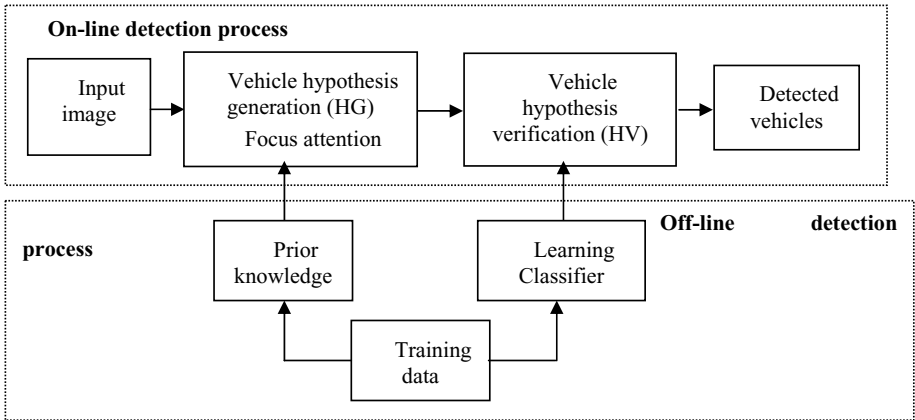
Fig. 1. Two-stages approach for vehicle detection

The rest of paper is organized as follows: Section 2 presents the hypothesis generation step using a long vertical and horizontal edges map and the repair long horizontal method. Section 3 details the hypothesis verification with PCA features and the SVM classifier. Section 4 describes some simulation results. Finally, conclusions and future work are shown in section 5.

2   Hypothesis Generation Using Long Vertical, Horizontal Edges and Repair Long Horizontal Edges

To hypothesize potential vehicle locations in an image, prior knowledge about the appearance of the front and rear of the vehicles may be used. For example, both a vehicles front and rear contain several horizontal and vertical structures, such as the rear-window, front-window, fascia, and bumpers. Based on this observation, the following procedure could be applied to hypothesize candidate vehicle locations. First, horizontal and vertical structures of interest could be identified by applying horizontal and vertical Sobel edge detectors to create the vertical and horizontal edge maps (Fig. 2).
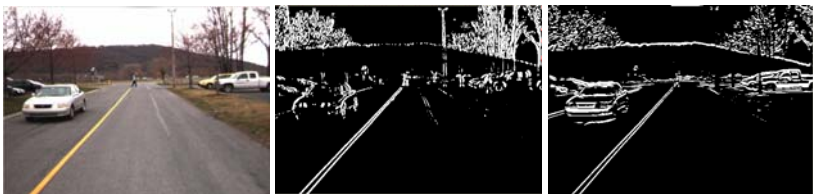
Fig. 2. The original, vertical, and horizontal edges images (from left to right)

To select the most promising horizontal and vertical structures, further analysis would be required, such as performing an analysis to get long vertical and horizontal edges from vertical and horizontal edges map (Fig. 3).

Fig. 3. Vertical and horizontal edges images (before and after repaired)

In the vehicle region, there are many horizontal edges in addition, horizontal edges that are interrupted or gather with each other. Therefore, if we use this particular edge map information (left image in Figure 3) to detect the vehicle candidate regions it will require a large amount of time for processing and the final result will be inaccurate as shown in left image in Figure 4.

Fig. 4. Incurracy vehicle candidate regions(left) and right candidate regions (right) are detected

To overcome above problem, we propose a simple method that we call *"Repaired long horizontal edges"*. The neighborhood long horizontal edges in one region will be connected to each other into one long horizontal edge and the horizontal edges which are used are deleted.

We call:

> Height - height of image,
> Hor_List - list of horizontal edges in image.

The pseudo code of this purpose is presented as follows:

```
i = Height;
    While (i > 0) do begin   // Scan image belonging to
the vertical direction
       HE = Hor_List [i];
       For j from i - 1 to i - 10 do begin
          Find  horizontal  edge  HE1  belonging  to  this
area and with an x-coordinate that intersects with an x-
coordinate of HE;
    Update x-coordinate of the HE to connect HE and HE1
into one horizontal edge;
    Delete HE form horizontal edges list;
  End for;
  i = i -1;
 End while;
```

After applying the above procedure to repair the horizontal edges of the image, we have a result to that shown in the right image in Figure 3 and the number of long horizontal edges reduces (from 521 to 136 in our tested case). The sizes of the horizontal edges increases and some edges cover vehicle's regions. Next, we extract these potential regions that contain long horizontal edges after repair; this is the bottom line in which a vehicle may be covered. For each of the long horizontal edges after repair, we use following procedure to extract the potential vehicle's regions:

```
For i = 1 to number Horizontal edge in Hor_List
Begin
    HE = Hor_List [i];
    Check if the region that contains the bottom line is
a HE and has 2 vertical edges
    Then if the region has 2 vertical edges, we check if
region has more than 2 horizontal edges?
    If two above conditions are satisfied, it becomes a
vehicle candidate region. Otherwise, it is rejected from
the candidate regions list.
End for;
```

The proposed approach presented above is robust in reducing the sensitivity to noise and environmental conditions in the hypothesis generation step by using connecting neighborhood long horizontal edges in the procedure to repair long horizontal edges. Forming the first hypotheses at the lowest level of detail is very useful since this level contains only the most salient structural features. Besides improving robustness, the proposed scheme also speeds-up the whole process since it utilizes very few regions compared with windowing and some other methods.

Moreover, this approach does not need to know information about the distance from the ego to another vehicle and camera calibration and can be applied to generate candidate regions for many different kinds of vehicles. Several examples are addressed in the simulation results section which will confirm our statement.

3   Hypothesis Verification Using PCA Feature and SVM Classifier

Verifying a hypothesis or vehicle recognition is essentially a two-class pattern classification problem (i.e., vehicle versus non-vehicle). Building a pattern classification system involves two main steps: (1) extracting a number of features and (2) training a classifier using the extracted features to distinguish among different class instances. We follow the method that describes in [3] to extract and build feature vector of vehicles that we call *"Eigenspace of vehicle"*.

3.1   Principle Component Analysis (PCA)

As described in general face recognition applications [3], principal components analysis (PCA) is used with two main purposes. Firstly, it reduces the dimensions of data to a computationally feasible size. Secondly, it extracts the most representative features out of the input data so that although the size is reduced, the main features remain, and are still able to represent the original data. A set of M vehicle images were

collected as the training set for our classifier. Each data image I with size NxN can be viewed as a vector I with size N^2. The whole set of input data can be considered as a matrix with M rows and N^2 columns. We will first try to obtain the covariance matrix from this set of image vectors, as understood as one of the algorithmic transformations of those data. Then, the eigenvectors of this covariance transformation will be obtained. Eigenvectors are those that are invariant in direction during a transformation, which can be used as a representation set of the whole big dataset. Those components are called eigenfaces by the authors who created a face detection application [3]. The covariance matrix of the input data is calculated starting from the algorithmic mean of all vectors $I_1, I_2, \cdots, I_M$

$$\psi = \frac{1}{M} \sum_{i=1}^{M} \Gamma_i \qquad (1)$$

The difference image vector I_i and mean ψ is called Φ with:

$$\Phi_i = I_i - \psi \qquad (2)$$

The theoretical covariance matrix C of all Φ_i is:

$$C = \frac{1}{M} \sum_{i=1}^{M} \Phi_i \Phi_i^T = AA^T \qquad (3)$$

where, $A = [\Phi_1, \Phi_2, \cdots, \Phi_M]$. The eigenspace can then be defined by computing the eigenvectors μ_i of C. Since C is very large (N^2), computing its eigenvector will be very expensive. Instead, we can compute v_i, the eigenvectors of $A^T A$, an MxM matrix. Then μ_i can be computed from v_i as follows:

$$\mu_i = \sum_{j=1}^{M} v_{ij} \Phi_j, \quad i = 1 \cdots M \qquad (4)$$

The collection of M eigenvectors v_i can be seen as the reduced dimension representation of the original input data (with size N^2) when $M << N^2$ This set of M eigenvectors will have a corresponding eigenvalue associated with it, which indicates the distribution of this eigenvector in representing the whole dataset. Many studies have shown that only a small set of the eigenvectors with top eigenvalues is enough to generate the entire image characteristic. In our system, we keep the K top eigenvectors, where K represents the number of important features from the vehicle eigenspace, and form the vehicle eigenspace ε (M rows, K columns).

Given a new image, Γ, we subtract the mean and compute the projection:

$$\hat{\Phi} = \sum_{i=1}^{K} w_i \mu_i \qquad (5)$$

where $w_i = \mu_i^T \Gamma$ are the coefficients of the projection. In this paper, $\{w_i\}$ are our eigenfeatures.

The projection coefficients allow us to represent images as linear combinations of the eigenvectors. It is well known that the projection coefficients define a compact image representation and that a given image can be reconstructed from its projection coefficients and the eigenvectors (i.e., basis). The eigenspace $\mathcal{E}$ representation of the images has been used in various applications such as image compression and face recognition, as well as vehicle detection [4].

3.2 Support Vector Machine (SVM) and Classifier

The Support Vector Machine [5] is a statistical learning theory which is becoming increasingly popular as the machine learning technique of choice for pattern recognition, regression, and classification. The main objective of training a support vector machine is to find the biggest possible classification margin, which indicates the minimum value of w in:

$$\frac{1}{2} w^T w + C \sum_{i=1}^{l} \xi_i \qquad (6)$$

where l is number of samples in the training set, $\xi_i \geq 0$ and C is the error tolerance level. In our vehicle classification case, the data is non-linear but the theory can be treated in the same way. Real vehicle images can be considered on one side and the non-vehicle images are on the other side. The training of instance-label pairs $L_i(w_i, y_i)$, where w_i is the weight vector and $y_i \in \{-1, 1\}$, $i = 1 \cdots l$ is the class label of w_i, will aim to derive the optimized support vectors on the vehicle eigenspace. These best support vectors will form a maximum margin $m = \dfrac{2}{|w|}$ between the two classes, and can be used as the SVM model, S, for future classification processes. The judgment of any unclassified images, U_i, will be based on the relationship between its projection weight, w_i, onto the eigenspace $\mathcal{E}$, and the margin distance from the SVM model S. Among several popular kernel functions which can be used to find the optimized value of margin m, the Radial Basis Function (RBF) kernel $K(x_i, x_j)$ has been selected for our vehicle classification system, due to its simplicity and proven capability for dealing with both non-linear and linear datasets as presented in [7].

$$K(x_i, x_j) = \exp\left(-\gamma \|x_i - x_j\|^2\right), \gamma > 0 \qquad (7)$$

The Radial Basis Function (RBF) kernel is used in this study (i.e., our experiments have shown that the KBF kernel outperforms other kernels in the context of our simulation).

Appling the RBF kernel function (7) for an SVM classifier target, we need to know the two vital parameters, RBF kernel parameter γ and error level C, depend on each particular data type and kernel. Different values of C and γ can lead to large differences in the classification result. It is not known beforehand which C and γ are best for any one problem; consequently some kind of model selection (parameter search) must be done. In order to choose the most suitable parameters, a five-fold cross validation described in [6] has been used.

4 Simulation Results

4.1 Dataset

Our vehicle training database contains images of 1185 vehicle images and 1125 non-vehicle images. Some of these images are from the public downloadable PASCAL, MIT vehicle datasets, and INRIA objects dataset, while the rest were taken by ourselves. We used an OpenCV function tool to convert all images onto JPG format and the resolution of each image is 64x64. There is some variability in the way the sample images were taken; for example, certain sample images cover the whole vehicle, others cover the vehicle partially, while some contain the vehicle and some other background objects (see Fig. 5).

Fig. 5. Some vehicle and non-vehicle training sample images

4.2 PCA and SVM Parameters

We ran a five-fold cross validation described in section 3.1 above with different values of eigenvectors components (from 25 to 100 components). Then, we find a suitable parameters set for choosing the number of PCA components and (C, γ) pairs for SVM classifier training. Our best training set are:

Number of components:	30 eigenvectors
C	3.363586
γ	0.707107
The accuracy of SVM training:	94.9351%

4.3   Simulation of Vehicle Detection

We present here some exemplary simulation results of our approach for vehicle detection. The complete results were analyzed and simulated using a Visual C++ program in combination with an Open CV computer-vision tool and LIBSVM version 2.88 [7]. As shown in these examples, our vehicle detection algorithm suitably detects both whole and partial vehicles even in the cases of road scenes with complex backgrounds.

Example 1 – An example with one small, distant vehicle.

Fig. 6. Example 1 - HG result (left image) and HV result (right image)

Example 2 – Many different kinds of vehicles (trucks, car) are moving on the road.

Fig. 7. Example 2 - HG result (left image) and HV result (right image)

Example 3 – An actual road with a complex background.
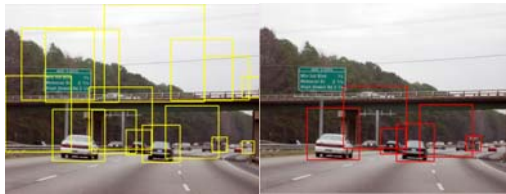
Fig. 8. Example 3 - HG result (left image) and HV result (right image)

Example 4 – An example road with partial vehicles and many transportation signs.

Fig. 9. Example 4 - HG result (left image) and HV result (right image)

Example 5 – Crowded road scene with a complex background.
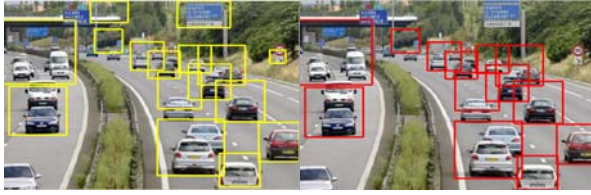
Fig. 10. Example 5 - HG result (left image) and HV result (right image)

5   Conclusion and Future Work

In this paper, a systematic approach has been proposed and implemented to provide a robust solution for preceding vehicle detection (front and rear vehicle view detection). We have proposed a two-stage approach to robustly detect preceding vehicles in traffic surveillance. The first stage is hypothesis generation (HG) which used a simple technique based on prior knowledge about front and rear vehicle appearance to generate vertical and horizontal edge maps. Next, we choose long vertical and long horizontal edges using some threshold guidelines. Then, a new method that repairs and connects the neighborhood of long horizontal edges was proposed to indicate the bottom location that a vehicle may appear in the scene. Finally, we generated the potential target locations of the vehicles in the image scene. In the hypothesis verification (HV) stage, all the hypotheses were verified by a strong classifier using vehicle eigenspace and support vector machine. We have tested the proposed methods on many different kinds of real roads with complex background scenes. The complete results were analyzed and simulated on a Visual C++ program in combination with the Open CV computer-vision tool, and LibSVM library. The simulation results obtained from different runs are believe to be very satisfactory and have proven the robustness of the preceding vehicle detection framework.

Although the current proposed scheme works reasonably well, the process manually chooses the number components or numbers top of eigenvectors to build the vehicle eigenspace model. For future work, we plan to investigate a proposed feature selection scheme across different feature extraction space. We believe that a Genetic Algorithm (GAs) will be able to select complimentary information offered by a different feature extraction method and improve the vehicle detection system performance. In the near future, we plan to detect pedestrians, develop a mobile robot and test our system with obstacle avoidance conditions on real world roads.

References

1. Papageorgiou, C., Poggio, T.: A Trainable System for Object Detection. International Journal of Computer Vision 8(1), 15–33 (2000)
2. Han, F., Shan, Y., Cekander, R., Sawhney, H.S., Kumar, R.: A Two-stage Approach to People and Vehicle Detection with HOG-Base SVM. In: PerMIS proceeding, pp. 133–140 (2006)

3. Turk, M., Pentland, A.: Eigenfaces for Recognition. Journal of Cognitive Neuroscience 3(1), 71–86 (1991)
4. Chen, Z.X., Chen, W.B.: A PCA-based Vehicle Classification Framework. In: Proc. IEEE Conference on Data Engineering, pp. 1–17 (2006)
5. Steve, R.G.: Support Vector Machine for Classification and Regression Web Site, Technical report of Faculty of Engineering, Science and Mathematics School of Electronics and Computer Science, University of Southampton, USA (1998),
   `http://users.ecs.soton.ac.uk`
6. Hsu, C.W., Chang, C.C., Lin, C.J.: Apractical Guide to Support Vector Classification, Department of Computer Science, National Taiwan University,
   `http://www.csie.ntu.edu.tw/~cjlin`
7. Hsu, C.W., Chang, C.C., Lin, C.J.: LIBSVM: A Library for Support Vector Machines (2001), `http://www.csie.ntu.edu.tw/cjlin/libsvm`

A Novel Method Using Contourlet to Extract Features for Iris Recognition System

Amir Azizi[1] and Hamid Reza Pourreza[2]

[1] Islamic Azad University Mashhad Branch
[2] Ferdowsi University of Mashhad
`Amirazizi_b@Yahoo.com`, `hpourreza@um.ac.ir`

Abstract. In different areas of Biometrics, recognition by iris images in nowadays has been taken into consideration by researchers as one of the common methods of identification like passwords, credit cards or keys. Iris recognition a novel biometric technology has great advantages such as variability, stability and security. Although the area of the iris is small it has enormous pattern variability which makes it unique for every one and hence leads to high reliability. In this paper we propose a new feature extraction method for iris recognition based on contourlet transform. Contourlet transform captures the intrinsic geometrical structures of iris image. It decomposes the iris image into a set of directional sub-bands with texture details captured in different orientations at various scales so for reducing the feature vector dimensions we use the method for extract only significant bit and information from normalized iris images. In this method we ignore fragile bits. At last, the feature vector is created by using Co-occurrence matrix properties. For analyzing the desired performance of our proposed method, we use the CASIA dataset, which is comprised of 108 classes with 7 images in each class and each class represented a person. And finally we use SVM and KNN classifier for approximating the amount of people identification in our proposed system. Experimental results show that the proposed increase the classification accuracy and also the iris feature vector length is much smaller versus the other methods.

Keywords: Biometric–Iris Recognition, Contourlet Transform, Co-occurrence Matrix, Support Vector Machine (SVM).

1 Introduction

There has been a rapid increase in the need of accurate and reliable personal identification infrastructure in recent years, and biometrics has become an important technology for the security. Iris recognition has been considered as one of the most reliable biometrics technologies in recent years [1, 2]. The human iris is the most important biometric feature candidate, which can be used for differentiating the individuals. For systems based on high quality imaging, a human iris has an extraordinary amount of unique details as illustrated in Fig.1. Features extracted from the human iris can be used to identify individuals, even among genetically identical twins [3]. Iris-based recognition system can be noninvasive to the users since the iris is an internal organ as well as externally visible, which is of great importance for the real-time
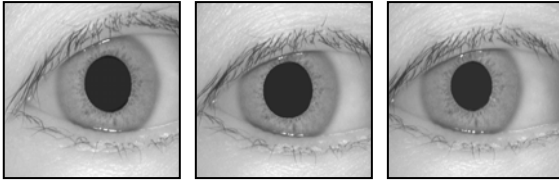
Fig. 1. Samples of iris images from CASIA [7]

applications [4]. Based on the technology developed by Daugman [3, 5, 6], iris scans have been used in several international airports for the rapid processing of passengers through the immigration which have pre registered their iris images.

1.1 Related Works

The usage of iris patterns for the personal identification began in the late 19th century; however, the major investigations on iris recognition were started in the last decade. In [9], the iris signals were projected into a bank of basis vectors derived by the independent component analysis, and the resulting projection coefficients were quantized as Features. A prototype was proposed in [10] to develop a 1D representation of the gray-level profiles of the iris. In [11], biometrics based on the concealment of the random kernels and the iris images to synthesize a minimum average correlation energy filter for iris authentication were formulated. In [5, 6, 12], the Multiscale Gabor filters were used to demodulate the texture phase structure information of the iris. In [13], an iris segmentation method was proposed based on the crossed chord theorem and the collarette area. An interesting solution to defeat the fake iris attack based on the Purkinje image was depicted in [16]. An iris image was decomposed in [17] into four levels by using the 2D Haar wavelet transform, the fourth-level high-frequency information was quantized to form an 87-bit code, and a modified competitive learning neural network (LVQ) was adopted for classification.

Fourth-level high-frequency information was quantized to form an 87-bit code, and a modified competitive learning neural network (LVQ) was adopted for classification. In [18], a modification to the Hough transform was made to improve the iris segmentation, and an eyelid detection technique was used, where each eyelid was modeled as two straight lines. A matching method was implemented in [19], and its performance was evaluated on a large dataset. In [20], a personal identification method based on the iris texture analysis was described. The remainder of this paper is organized as follows: Section 2 deals with proposed method. Section 3 deals with Feature Extraction method discussion. Section 4 deals with feature subset selection and vector creation techniques, Section 5 shows our experimental results and finally Section 6 concludes this paper.

2 Proposed Method: The Main Steps

Fig. 2 illustrates the main steps of our proposed Approach. First the image preprocessing step performs the localization of the pupil, detects the iris boundary, and isolates the collarette region, which is regarded as one of the most important areas of the iris
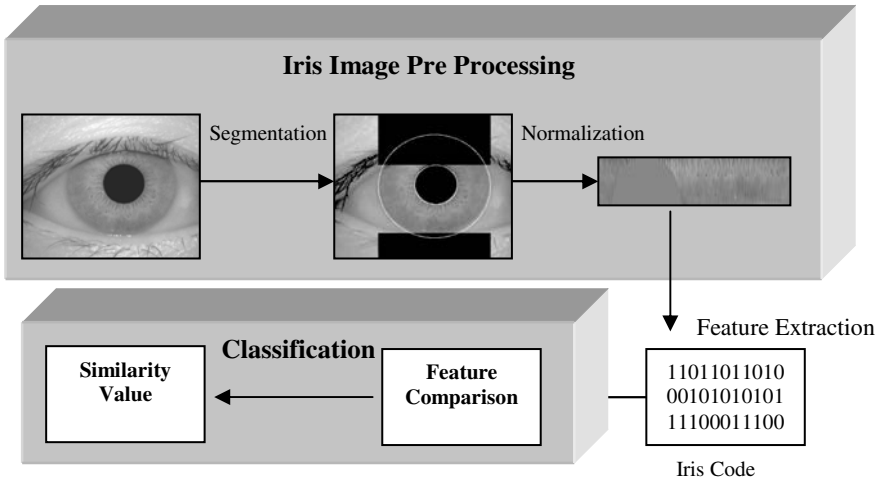
Fig. 2. Flow diagram of the proposed iris recognition scheme

complex pattern. The collarette region is less sensitive to the pupil dilation and usually unaffected by the eyelids and the eyelashes [8]. We also detect the eyelids and the eyelashes, which are the main sources of the possible occlusion. In order to achieve the invariance to the translation and the scale, the isolated annular collarette area is transformed to a rectangular block of fixed dimension.

The discriminating features are extracted from the transformed image and the extracted features are used to train the classifiers. The optimal features subset is selected using several methods to increase the matching accuracy based on the recognition performance of the classifiers.

2.1   Iris Image Preprocessing

First, we outline our approach, and then we describe further details in the following subsections. The iris is surrounded by the various non relevant regions such as the pupil, the sclera, the eyelids, and also noise caused by the eyelashes, the eyebrows, the reflections, and the surrounding skin [9].We need to remove this noise from the iris image to improve the iris recognition accuracy.

2.1.1   Iris / Pupil Localization
The iris is an annular portion of the eye situated between the pupil (inner boundary) and the sclera (outer boundary). Both the inner boundary and the outer boundary of a typical iris can be taken as approximate circles. However, the two circles are usually not concentric [20].

2.1.2   Eyelids, Eyelashes, and Noise Detection
- Eyelids are isolated by first fitting a line to the upper and lower eyelids using the linear Hough transform. A second horizontal line is then drawn, which intersects with the first line at the iris edge that is closest to the pupil [20].

- Separable eyelashes are detected using 1D Gabor filters, since a low output value is produced by the convolution of a separable eyelash with the Gaussian smoothing function. Thus, if a resultant point is smaller than a threshold, it is noted that this point belongs to an eyelash.
- Multiple eyelashes are detected using the variance of intensity, and if the values in a small window are lower than a threshold, the centre of the window is considered as a point in an eyelash as shown in Fig .3.
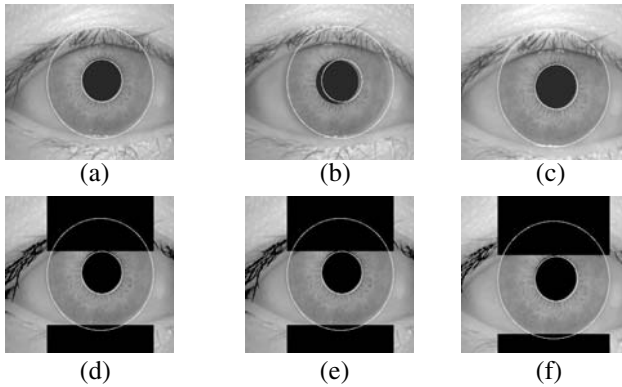
Fig. 3. CASIA iris images (a), (b), and (c) with the detected Collarette area and the corresponding images (d), (e), and (f) after Detection of noise, eyelids, and eyelashes

2.1.3 Iris Normalization

We use the rubber sheet model [12] for the normalization of the isolated collarette area. The center value of the pupil is considered as the reference point, and the radial vectors are passed through the collarette region. We select a number of data points along each radial line that is defined as the radial resolution, and the number of radial lines going around the collarette region is considered as the angular resolution. A constant number of points are chosen along each radial line in order to take a constant number of radial data points, irrespective of how narrow or wide the radius is at a particular angle. We build the normalized pattern by backtracking to find the Cartesian coordinates of data points from the radial and angular positions in the normalized pattern [3, 5, and 6]. The normalization approach produces a 2D array with horizontal dimensions of angular resolution, and vertical dimensions of radial resolution form the circular-shaped collarette area (See Fig.4I). In order to prevent non-iris region data from corrupting the normalized representation, the data points, which occur along the pupil border or the iris border, are discarded. Fig.4II (a) (b) shows the normalized images after the isolation of the collarette area.

3 Feature Extraction and Encoding

Only the significant features of the iris must be encoded so that comparisons between templates can be made. Gabor filter and wavelet are the well-known techniques in

θ

r

Black portion represents region of interest
Of the unwrapped iris image

White region denotes noise

(I)

(a)   (II)   (b)
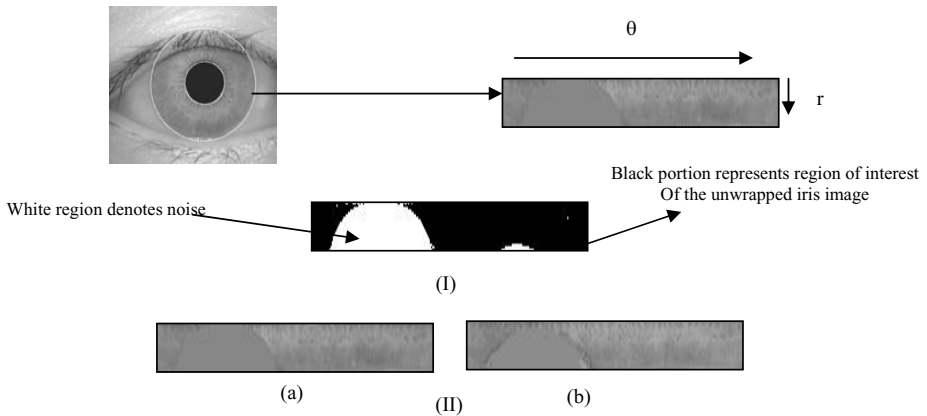
Fig. 4. (I) shows the normalization procedure on CASIA dataset; (II) (a), (b) Show the normalized images of the isolated collarette regions

texture analysis [5, 19, 20, 21]. In wavelet family, Haar wavelet [22] was applied by Jafer Ali to iris image and they extracted an 87-length binary feature vector. The major drawback of wavelets in two-dimensions is their limited ability in capturing Directional information. The contourlet transform is a new extension of the wavelet transform in two dimensions using Multi scale and directional filter banks.

The feature representation should have information enough to classify various irises and be less sensitive to noises. Also in the most appropriate feature extraction we attempt to extract only significant information, more over reducing feature vector dimensions, the processing lessened and enough information is supplied to introduce iris feature vectors classification.

3.1   Contourlet Transform

Contourlet transform (CT) allows for different and flexible number of directions at each scale. CT is constructed by combining two distinct decomposition stages [33], a multistage decomposition followed by directional decomposition. The grouping of wavelet coefficients suggests that one can obtain a sparse image expansion by applying a multi-scale transform followed by a local directional transform. It gathers the nearby basis functions at the same scale into linear structures. In essence, a wavelet-like transform is used for edge (points) detection, and then a local directional transform for contour segments detection. A double filter bank structure is used in CT in which the Laplacian pyramid (LP) [23] is used to capture the point discontinuities, and a directional filter bank (DFB) [24] to link point discontinuities into linear structures. The combination of this double filter bank is named pyramidal directional filter bank (PDFB) as shown in Fig.5.Benefits of Contourlet Transform in the Iris Feature Extraction To capture smooth contours in images, the representation should contain basis functions with variety of shapes, in particular with different aspect ratios. A major challenge in capturing geometry and directionality in images comes from the discrete nature of the data; the input is typically sampled images defined on rectangular grids.
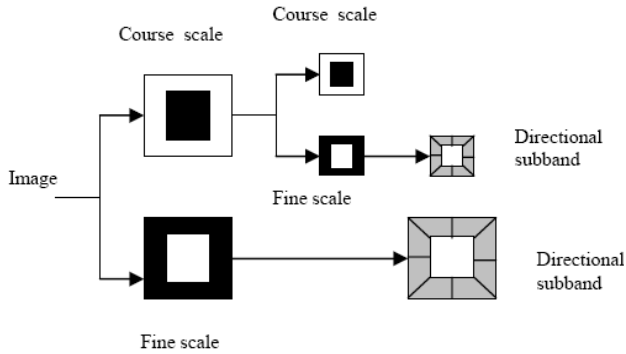
Fig. 5. Two Level Contourlet Decomposition [33]

Because of pixelization, the smooth contours on sampled images are not obvious. For these reasons, unlike other transforms that were initially developed in the continuous domain and then discretized for sampled data, the new approach starts with a discrete-domain construction and then investigate its convergence to an expansion in the continuous-domain. This construction results in a flexible multi-resolution, local, and directional image expansion using contour segments. Directionality and anisotropy are the important characteristics of contourlet transform. Directionality indicates that having basis function in many directions, only three direction in wavelet. The anisotropy property means the basis functions appear at various aspect ratios where as wavelets are separable functions and thus their aspect ratio is one. Due to this properties CT can efficiently handle 2D singularities, edges in an image. This property is utilized in this paper for extracting directional features for various pyramidal and directional filters.

3.2 The Best Bit in an Iris Code

Biometric systems apply filters to iris images to extract information about iris texture. Daugman's approach maps the filter output to a binary iris code. The fractional
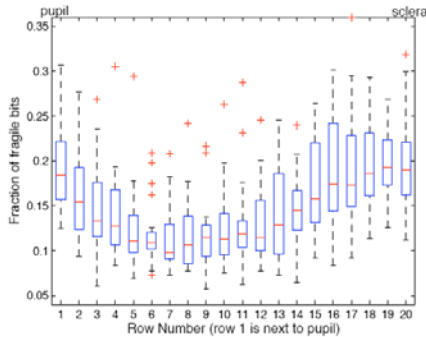
Fig. 6. Percent of Fragile Bit in Iris Pattern [25]

Hamming distance between two iris codes is computed and decisions about the identity of a person are based on the computed distance. The fractional Hamming distance weights all bits in an iris code equally. However, not all the bits in an iris code are equally useful. For a given iris image, a bit in its corresponding iris code is defined as "fragile" if there is any substantial probability of it ending up a 0 for some images of the iris and a 1 for other images of the same iris. According to [25] the percentages of fragile bits in each row of the iris code, Rows in the middle of the iris code (rows 5 through 12) are the most consistent (See Fig. 6.)

4   Feature Vector Creation in Proposed Method

According to the method mentioned in section 3.2, we concluded the middle band of iris normalized images have more important information and less affected by fragile bits, so for introducing iris feature vector based on contourlet transform the rows between 5 and 12 in iris normalize image are decomposed into eight directional subband outputs using the DFB at three different scales and extract their coefficients. In our method we use using the Grey Level Co-occurrence Matrix (GLCM).The technique uses the GLCM (Grey Level Co-occurrence Matrix) of an image and it provides a simple approach to capture the spatial relationship between two points in a texture pattern. It is calculated from the normalized iris image using pixels as primary information. The GLCM is a square matrix of size G * G, where G is the number of gray levels in the image. Each element in the GLCM is an estimate of the joint probability of a pair of pixel intensities in predetermined relative positions in the image. The $(i, j)^{th}$ element of the matrix is generated by finding the probability that if the pixel location (x, y) has gray level I_i then the pixel location $(x+dx, y+dy)$ has a gray level intensity I_j. The dx and dy are defined by considering various scales and orientations. Various textural features have been defined based on the work done by Haralick [26]. These features are derived by weighting each of the co-occurrence matrix values and then summing these weighted values to form the feature value. The specific features considered in this research are defined as follows:

1) $\text{Energy} = \sum_i \sum_j p(i, j)^2$

2) $\text{Contrast} = \sum_{n=0}^{N_g-1} n^2 \left[\sum_{i=1}^{N_g} \sum_{j=1}^{N_g} P(i, j) \|i - j\| = n \right]$

3) $\text{Correlation} = \dfrac{\sum_i \sum_j (ij) P(i, j) - \mu_x \mu_y}{\sigma_x \sigma_y}$

4) $\text{Homogeneity} = \sum_i \sum_j \dfrac{1}{1 + (i - j)^2} P(i, j)$

5) $\text{Autocorrelation} = \sum_i \sum_j (ij) P(i, j)$

6) Dissimilarity $= \sum_i \sum_j |i - j|.P(i, j)$

7) Inertia $= \sum_i \sum_j (i - j)^2 P(i, j)$

Here $\mu_x, \mu_y, \sigma_x, \sigma_y$ are mean and standard deviation along x and y axis.

5   Experimental Results

For creating iris feature vector we carried out the following steps:

1) Iris normalized image (Rows in the middle of the iris code (rows 5 through 12)) is decomposed up to level two.(for each image, at level one, 2 and at level two, 4 sub band are created).
2) The sub bands of each level are put together, therefore at level one a matrix with 4*120 elements, and at level two a matrix with 16*120 elements is created. We named these matrixes: Matrix1 and Matrix 2.
3) By putting together Matrix1 and Matrix 2, a new matrix named Matrix3 with 20*120 elements is created. The co-occurrence of these three matrixes with offset one pixel and angles 0, 45, 90 degree is created and name this matrix: CO1, CO2 and CO3.in this case for each image 3 co-occurrence matrixes with 8*8 dimensions are created.
4) According to the Haralick's [26] theory the co-occurrence matrix has 14 properties, of which in iris biometric system we used 7 properties which are used for 3 matrixes , so the feature vector is as follow:

F=[En1,Cont1,cor1,hom1,Acor1,dis1,ine1,   En2,Cont2,cor2,   hom2,Acor2,dis2,ine2 En3,Cont3,cor3,hom3,Acor3,dis3,ine3] In other word the feature vector in our method has only 21 elements. Also for improving results, for each sub bands and scale we create a feature vector by using GLCM.in other words for each eight sub bands in level 3 of Contourlet transform we computed GLCM properties, separately and then by combining these properties the feature vector is created. In this case the feature vector has 56 elements. In Table 1 you can see the result of implementing our proposed method:

Table 1. Result of Implementing Proposed Method

The Number Of Classes	The Correct of Percentage Classification (%)		
	KNN Classifier	SVM Classifier(Kernel 1)	SVM Classifier(Kernel 2)
20	96.6	100	100
40	88.3	94.3	96.3
60	90.8	91.6	95.6
80	89.3	90.1	95.8
100(GLCM)	88.5	90.07	94.2
100(GLCM (Combining Sub bands)	87.5	91.3	96.3

In Table 2 we compared our proposed method with some other well known methods from 2 view points: feature vector length and the correct of percentage classification:

Table 2. Comparison Between Our Proposed Method and Some well- known Method

Method	The Correct Of Percentage Classification (%)	The Feature Vector Length(Bit)
Dugan[3]	100	2048
Lim[17]	90.4	87
Ma[20]	959	1600
Jafar Ali[22]	92.8	87
Our Methods		
(GLCM)	94.2	21
GLCM (Combining Sub bands)	96.3	56

6   Conclusions

In this paper we proposed an effective algorithm for iris feature extraction using con-tourlet transform Co-occurrence Matrix have been presented. The GLCM proved to be a good technique as it provides reasonable accuracy and is invariant to iris rotation. For Segmentation and normalization we use Daugman methods .Feature extraction in our proposed method includes: sub bands proper composition from Contuorlet pyra-mid and co-occurrence calculations and finally selecting a set of Haralick's properties that form the Maximum distance between inter classes and Minimum distance between intra classes. Our proposed method can classify iris feature vector properly. The rate of expected classification for the fairly large number of experimental date in this paper verifies this claim. In the other words our method provides a less feature vector length with an insignificant reduction of the percentage of correct clas-sification.

References

1. Wildes, R.P.: Iris recognition: an emerging biometric technology. Proceedings of the IEEE 85(9), 1348–1363 (1997)
2. Jain, A., Bolle, R., Pankanti, S.: Biometrics: Personal Identification in a Networked Soci-ety. Kluwer Academic Publishers, Norwell (1999)
3. Daugman, J.: Biometric personal identification system based on iris analysis. US patent no. 5291560 (1994)
4. Mansfield, T., Kelly, G., Chandler, D., Kane, J.: Biometric product testing. Final Report, National Physical Laboratory, Middlesex, U (2001)

5. Daugman, J.G.: High confidence visual recognition of persons by a test of statistical independence. IEEE Transactions on Pattern Analysis and Machine Intelligence 15(11), 1148–1161 (1993)
6. Daugman, J.: Demodulation by complex–valued wavelets for stochastic pattern recognition. International Journal of Wavelets, Multiresolution and Information Processing 1(1), 1–17 (2003)
7. CASIA.Chinese Academy of Sciences – Institute of Automation. Database of 756 Grayscale Eye Images. Versions 1.0 (2003), http://www.sinobiometrics.com
8. He, X., Shi, P.-f.: An efficient iris segmentation method for recognition. In: Singh, S., Singh, M., Apte, C., Perner, P. (eds.) ICAPR 2005. LNCS, vol. 3687, pp. 120–126. Springer, Heidelberg (2005)
9. Bae, K., Noh, S., Kim, J.: Iris feature extraction using independent component analysis. In: Kittler, J., Nixon, M.S. (eds.) AVBPA 2003. LNCS, vol. 2688, pp. 1059–1060. Springer, Heidelberg (2003)
10. Boles, W.W., Boashash, B.: A human identification technique using images of the iris and wavelet transform. IEEE Transactions on Signal Processing 46(4), 1185–1188 (1998)
11. Chong, S.C., Teoh, A.B.J., Ngo, D.C.L.: Iris authentication using privatized advanced correlation filter. In: Zhang, D., Jain, A.K. (eds.) ICB 2005. LNCS, vol. 3832, pp. 382–388. Springer, Heidelberg (2005)
12. Daugman, J.: Statistical richness of visual phase information: update on recognizing persons by iris patterns. International Journal of Computer Vision 45(1), 25–38 (2001)
13. He, X., Shi, P.-f.: An efficient iris segmentation method for recognition. In: Singh, S., Singh, M., Apte, C., Perner, P. (eds.) ICAPR 2005. LNCS, vol. 3687, pp. 120–126. Springer, Heidelberg (2005)
14. Jeong, D.S., Park, H.-A., Park, K.R., Kim, J.H.: Iris recognition in mobile phone based on adaptive gabor filter. In: Zhang, D., Jain, A.K. (eds.) ICB 2005. LNCS, vol. 3832, pp. 457–463. Springer, Heidelberg (2005)
15. Vijaya Kumar, B.V.K., Xie, C., Thornton, J.: Iris verification using correlation filters. In: Kittler, J., Nixon, M.S. (eds.) AVBPA 2003. LNCS, vol. 2688, pp. 697–705. Springer, Heidelberg (2003)
16. Lee, E.C., Park, K.R., Kim, J.H.: Fake iris detection by using purkinje image. In: Zhang, D., Jain, A.K. (eds.) ICB 2005. LNCS, vol. 3832, pp. 397–403. Springer, Heidelberg (2005)
17. Lim, S., Lee, K., Byeon, O., Kim, T.: Efficient iris recognition through improvement of feature vector and classifier. Electronics and Telecommunications Research Institute Journal 23(2), 61–70 (2001)
18. Liu, X., Bowyer, K.W., Flynn, P.J.: Experiments with an improved iris segmentation algorithm. In: Proceedings of the 4th IEEE Workshop on Automatic Identification Advanced Technologies (AUTO ID 2005), Buffalo, NY, USA, October 2005, pp. 118–123 (2005)
19. Liu, X., Bowyer, K.W., Flynn, P.J.: Experimental evaluation of iris recognition. In: Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR 2005), San Diego, Calif, USA, vol. 3, pp. 158–165 (2005)
20. Ma, L., Tan, T., Wang, Y., Zhang, D.: Personal identification based on iris texture analysis. IEEE Transactions on Pattern Analysis and Machine Intelligence 25(12), 1519–1533 (2003)
21. Daugman, J.: How Iris Recognition works. IEEE Transactions on Circuits and systems for video Technology 14(1), 21–30 (2004)

22. Jafar, M.H.A., Hussanien, A.E.: An Iris Recognition System to Enhance E–security Environment Based on Wavelet Theory. AMO – Advanced Modeling and Optimization 5(2) (2003)
23. Do, M.N., Vetterli, M.: The contourlet transform: An Efficient directional multiresolution image representation. IEEE Transactions on Image Processing 14(12), 2091–2106
24. Burt, P.J., Adelson, E.H.: The Laplacian pyramid as a compact image code. IEEE Trans. Commun. 31(4), 532–540 (April 1983); Bamberger, R. H., Smith, M.J.T.: A filter bank for the directional decomposition of images: Theory and design. IEEE Trans. Signal Proc. 40(4), 882–893(1992)
25. Hollingsworth, K.P., Bowyer, K.W.: The Best Bits in an Iris Code. IEEE Transactions on Pattern Analysis and Machine Intelligence, PAMI (April 2008)
26. Haralick, R.M., Shanmugam, K., Stein, L.D.: Textural features for image classification. IEEE Transactions on Systems, Man, and Cybernetics 3(6), 610–621 (1973)

Vehicle License Plate Detection Algorithm Based on Color Space and Geometrical Properties

Kaushik Deb[1], Vasily V. Gubarev[2], and Kang-Hyun Jo[1]

[1] Dept. of Electrical, Electronics and Information Systems, University of Ulsan,
680 - 749 San 29, Mugeo 2-dong, Nam-ku, Ulsan, Korea
[2] Faculty of Automation Engineering, Novosibirsk State Technical University,
630 - 092, Karl Marks avenue 20, Novosibirsk, Russia
{debkaushik99,jkh2009}@islab.ulsan.ac.kr,
gubarev@vt.cs.nstu.ru

Abstract. In this paper, an algorithm for vehicle license plate detection (VLPD) is proposed, to select automatically statistical threshold value in HSI color space. The proposed VLPD algorithm consists of two main stages. Initially, HSI color space is adopted for detecting candidate regions. According to different colored LP, these candidate regions may include LP regions; geometrical properties of LP are then used for classification. The proposed method is able to deal with candidate regions under independent orientation and scale of the plate. Finally, the decomposition of candidate regions contains predetermined LP alphanumeric characters by using position in the histogram to verify and detect vehicle license plate (VLP) region. In experiment more than 150 images were used, they were taken from the variety of conditions such as complex scenes, illumination changing, distances and varied weather etc. Under these conditions, success of LP detection has reached to more than 94%.

Keywords: Vehicle license plate detection (VLPD), HSI color space, geometrical properties and position histogram.

1 Introduction

Humans can perform usual target recognition without too much effort. However, by computer the task of recognizing specific object in an image is one of the most difficult topics in the field of computer vision or digital image processing. VLPD task is quite challenging from vehicle images due to the multi-style plate formats, view point changes and the nonuniform outdoor illumination conditions during image acquisition [1, 3]. In addition, VLPD system should operate fast enough (real time) to satisfy the needs of ITS and not to miss a single interest object from the vehicle image. VLPD is also very interesting in finding license plate area from vehicle image. The vehicle license plate detection is widely used for detecting speeding cars, security control in restricted areas, in unattended parking zone, traffic law enforcement and electronic toll collection. With the rapid development of highway and the wide use of vehicles, people have started to

pay more and more attention to the advanced, efficient, and accurate intelligent transportation systems (ITSs). Recently, the necessity of VLPR has increased significantly. The license plate detection is an important research topic of VLPR system. One of the major problems in LP detection is determining LP systems. This system must guarantee robust detection under various weather and lighting conditions, complex scenes, independent of orientation and scale of the plate.

As far as detection of the plate region is concerned, researchers have found many methods of locating license plate. For example, a method for multi-style LP recognition has been presented in [1]. This method has introduced the density-based region growing algorithm for detecting LP location. The horizontal and vertical projections are scanned using a search window to locate the license plate in [2]. Survey paper [3], offers to researchers a link to a public image database to define a common reference point for VLPR algorithmic assessment. In addition, this survey paper discusses about current trends and anticipated research in VLPR system. A region-based license plate detection method was presented in [4], which firstly applies a mean shift procedure in spatial-range domain to segment a color vehicle image in order to get candidate regions. A method based on image segmentation technique named as sliding windows (SW) has also been proposed for detecting candidate region (LP region) in [5]. Currently, some researchers prefer a hybrid detection algorithm, where license plate location method based on corner detection, edge detection, characteristics of license shape, character's connection and projection is presented in [7, 8].

During the past few years, developments dealing with simple images have been achieved acceptable results. However, recent researches have been addressed to processing complex images with unconstrained conditions [9]. The proposed algorithm deals with such vehicle images.

The focus of this paper is on the integration of a new method to select automatically statistical threshold value in HSI color space for detecting candidate regions. Generally, as a common way of color-based VLPD system, threshold value is defined by predetermined coefficients or by user. It provides stable result, but in poor lighting condition it is too sensitive. Whereas in our experiments we calculate threshold value in a statistical way, 20% of sample data (only green, yellow and white LP areas) are randomly selected for training. After training from those sample data, the mean and standard deviation values of hue are computed for detection of green and yellow LP pixels. Detecting white license plate pixels, the mean and standard deviation values of saturation and intensity are computed to detect green, yellow and white LP from vehicle images.

Moreover, candidate regions may include LP regions; geometrical properties of LP are then used for classification. Finally, decomposing of candidate region which contains predetermined LP alphanumeric character, using position in the histogram to verify and detect vehicle license plate region is performed.

This paper is organized as follows. In Section 2, the specific features of Korean VLP have been described. The enhanced VLPD algorithm is proposed in section 3. Implementation of the enhanced VLPD algorithm is presented in Section 4. Finally, experimental results and some conclusions are reported in Section 5.

2   Specific Features of Korean VLP

In this section, the color arrangement of the license plate and outline of the Korean VLPs that are considered in this study have been discussed.

Color arrangement of the plate: Korean license plates is well classified as shown in Fig. 1. Each style has a different plate color and/or character color. However, in all, only five distinct colors like white, black, green, yellow, and deep blue are used in these license plates. It is worth paying attention to three different plate colors while searching for LP in an input image. Other types of vehicles, such as diplomatic cars and military vehicles, are not addressed since they are rarely seen. Color arrangements for the Korean VLPs are shown in Table 1.

Table 1. Styles of license plates

Vehicle type	Plate color	Character color
Private automobile 2-3	White	Black
	Green	White
2-3 Taxi, truck and bus	Yellow	Deep blue
Government vehicle	Yellow	Black

Outline of the Korean VLP: Standard LP contains Korean alphabets and numbers which are shown in Fig. 1. Few LPs contain Korean alphabets and numbers in two rows; in future these kinds of LPs are to be converted into a single-row types. Where plate color is white and character color is black, they contain seven alphanumeric characters written in a single line. In Fig. 1, Where plate color is green and yellow then character color is white and black or deep blue, respectively, they contain Korean LP in two rows. When plate color is yellow, some LP contains all alphanumeric characters written in a single line.

3   Proposed Algorithm

In the author's previous work [6], HSI color based vehicle license plate detection method was presented. We propose in this section an enhanced version of VLPD algorithm shown in Fig. 2. Like the traditional LP detection method, automatic focus and white balancing of camera often cause the changing illumination. To overcome this problem, we propose an adaptive LP detection method for detecting white license plate pixels; we use it in the case of really high- or low-illumination condition as shown in Fig. 4. And also distinguish with the traditional LP detection method, as license plates can appear at many different

Fig. 1. Outline of the Korean license plate
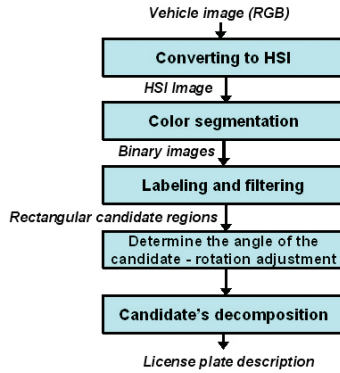
Fig. 2. Main algorithm scheme for detecting license plate region

angles to the camera's optical axis, each rectangular candidate region is rotated until they are all aligned in the same way before the candidate decomposition. The proposed algorithm can efficiently determine and adjust the rotated plate as shown in Fig. 5. Measurements such as center of area and the least second moment are employed to solve the rotation adjustment problem. The least second moment provides the principal axis as the orientation with the candidate object. General scheme for detecting LP region is shown in Fig. 2. In the proposed algorithm, detection is based on color properties of LP, shape-based verification and position histogram.

4   Vehicle License Plate Detecting Module

In this part, the four primary stages of the proposed VLPD algorithm, i.e., color segmentation, labeling and filtering, determine the angle of the candidate - rotation adjustment, and candidate decomposition of candidate region have been discussed in details.

4.1   Color Segmentation

In the proposed method, input vehicle images are converted into HSI color images. Then the candidate regions are found by using HSI color space on the basis of using hue, saturation and/or intensity. Many applications use the HSI color model. Machine vision uses HSI color space in identifying the color of different objects. Plate color information is used to detect candidate regions in our experiments, and shape properties of LP allow reducing number of LP-like candidates. One of the common ways of color-based vehicle license plate detection can be formalized as follows:

$$R(x,\ y) > \alpha_R;\ G(x,\ y) > \alpha_G;\ B(x,\ y) > \alpha_B\ , \tag{1}$$

$$R(x,\ y) - G(x,\ y) > \beta_{RG};\ R(x,\ y) - B(x,\ y) > \beta_{RB}\ , \tag{2}$$

where R, G and B are red, green and blue components of $x \times y$ image. α and β are predefined coefficients. Equation (1) sets up limitations for the minimal values of pixel components. Equation (2) formalizes dependencies between pixel components for LP. Generally, common way of using color-based vehicle license plate detection was based on two types of restrictions: first, restriction is based on equations (1) and (2). It provides good results in good lighting conditions. However, it is not good for low-contrast images. Pixel belongs to green and yellow LP, respectively like following equations (3) and (4)

$$b_{green} = \begin{cases} 1, [\{R(x, \ y) \ \leq \ 0.85 \cdot G(x, \ y)\} \ \& \ \{B(x, \ y) \ \leq \ 0.90 \cdot G(x, \ y)\}] \\ 0, \qquad\qquad\qquad\qquad otherwise \end{cases} \quad (3)$$

$$b_{yellow} = \begin{cases} 1, [\{B(x, \ y) \leq \ 0.90 \cdot R(x, \ y)\} \ \& \ \{B(x, \ y) \leq \ 0.80 \cdot G(x, \ y)\}] \\ 0, \qquad\qquad\qquad\qquad otherwise \end{cases} \quad (4)$$

where b_{green} and b_{yellow} are green and yellow candidate binary masks. The second restriction is based on equations (3) and (4), and a threshold value is taken heuristically. It provides stable result whereas in bad lighting condition it is too sensitive.

In this proposed method, LP detection is based on its color properties, namely mean and standard deviation values of hue. For detection of green and yellow LP pixels, hue parameter of HSI color is used in our experiment. To detect white LP pixels hue value is meaningless, hence only saturation and intensity parameters are important for this case. To estimate these properties, we used 30 images of LP taken under different lighting and weather conditions. After training from those sample data, the mean and standard deviation values of hue are computed for detection of green and yellow LP pixels. Detecting white license plate pixels, the mean and standard deviation values of saturation and intensity are computed to detect green, yellow and white LP from vehicle images. For detection of green and yellow LP pixels, the binarization process can be formulated as follows:

$$b_{green} = \begin{cases} 1, \ [\{\mu_H - \sigma_H \ \leq \ H(x, \ y) \ \leq \ \mu_H + \sigma_H\}] \ , \\ 0, \qquad\qquad\qquad otherwise \end{cases} \quad (5)$$

$$b_{yellow} = \begin{cases} 1, \ [\{\mu_H - \sigma_H \ \leq \ H(x, \ y) \ \leq \ \mu_H + \sigma_H\}] \ , \\ 0, \qquad\qquad\qquad otherwise \end{cases} \quad (6)$$

where $H(x, \ y)$, $S(x, \ y)$ and $I(x, \ y)$ are hue, saturation and intensity components of xth, yth pixel respectively. μ_H, σ_H are mean hue and hue standard deviation values for green and yellow LP of sample data, respectively.

However, the automatic focus and white balancing of camera often cause the changing illumination. Our proposed LP detection method can work well in normal illumination condition, but it seems not good enough to work in bad illumination conditions. To overcome this problem, we use an adaptive LP detection method; we use it in the case of really high- or low-illumination condition. For

normal, low- and high- illumination conditions of white license plate pixels, the binarization process can be formulated as follows, respectively:

$$b_{white(n)} = \begin{cases} 1, & [S(x,\ y) \ \leq \ (\mu_S + \sigma_S) \ \& \ I(x,\ y) \ \geq \ (\mu_I + 0.25 \cdot \sigma_I)] \ , \\ 0, & otherwise \end{cases} \quad (7)$$

$$b_{white(l)} = \begin{cases} 1, & [S(x,\ y) \ \leq \ (\mu_S + \sigma_S) \ \& \ I(x,\ y) \ \geq \ (\mu_I - 0.33 \cdot \sigma_I)] \ , \\ 0, & otherwise \end{cases} \quad (8)$$

$$b_{white(h)} = \begin{cases} 1, & [S(x,\ y) \ \leq \ (\mu_S + \sigma_S) \ \& \ I(x,\ y) \ \geq \ (\mu_I + 0.50 \cdot \sigma_I)] \ , \\ 0, & otherwise \end{cases} \quad (9)$$

where $S(x,\ y)$, $I(x,\ y)$, are saturation and intensity components of xth, yth pixel respectively. μ_S, μ_I are mean values for saturation and intensity, σ_S, σ_I are standard deviation values for saturation and intensities of white LP of sample data, respectively. $b_{white(n)}$, $b_{white(l)}$ and $b_{white(h)}$ are white candidate binary masks. Vehicle image and its color segmentation results are depicted in Fig. $3(a) - (c)$ (green, yellow and white back ground LP), respectively.

Color segmentation parameters are very sensitive in order to detect as much candidates as possible. All false candidates will be filtered out on the next stages. After the segmentation, there may still exist noises such as small holes or/and small bulges of the target candidate regions. These problem may be resolved by using mathematical morphology closing operation which is dilation followed by erosion to fill in holes and gaps smaller than the structuring element on the plate image. Removing those holes play an important role in calculating bounding box region. Implementation of morphological closing operation is depicted in Figure 4(c).

4.2 Labeling and Filtering

After the candidate regions are obtained by applying color segmentation, features of each region are to be extracted in order to correctly differentiate the LP regions from others. Next step of proposed algorithm is labeling the connected components. In the proposed method, a recursive algorithm is implemented for connected component labeling operation. Recursive algorithm works on one component at a time, but can move all over the image. In this step we

(a)   *(b)*   *(c)*
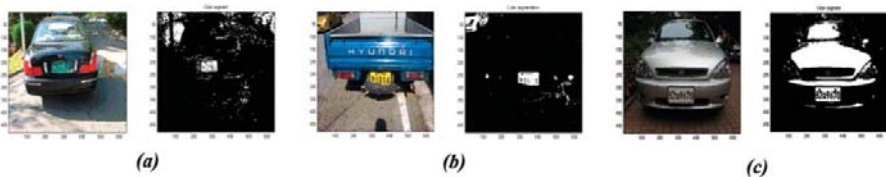
Fig. 3. An LP image (left) and its color segmentation results (right) using HSI color model
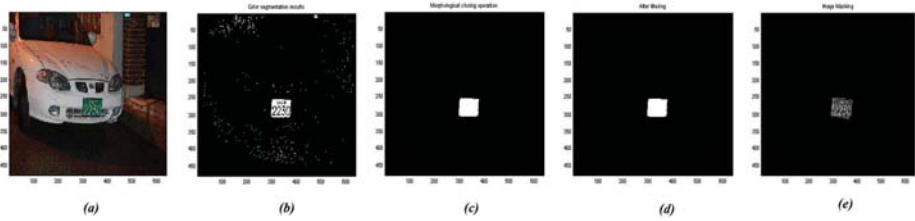
Fig. 4. Illustration of license plate segmentation: (a) an LP image in a night time, (b) color segmentation result, (c) implementation of morphological closing operation for removing small holes in candidate region, (d) detected candidate after filtering, and (e) extracted candidate

extract candidate regions which may include LP regions from the binary mask obtained in the previous step. During this step, main geometrical properties of LP candidate such as area, bounding box, and aspect ratio are computed. These parameters are used filtering operation to eliminate LP-like objects from candidate list. Filtering operation is done on geometrical properties of LP regions. Figure 4 illustrates the steps for license plate segmentation.

4.3 Determine the Angle of the Candidate Region-Rotation Adjustment

As license plates can appear at many different angles to the cameras optical axis, each rectangular candidate regions is rotated until they are all aligned in the same way before the candidate decomposition. Following the successful filtering operation in image, measurement such as center of area and the axis of least second moment are employed to solve the rotation adjustment problem.

The least second moments provides the principal axis as the orientation with the candidate object. For getting principal axis of detected candidate region, we compute central moments of detected candidate region. The central moments are defined as:

$$\mu_{pq} \;=\; \sum_{r=0}^{N-1} \sum_{c=0}^{N-1} (r - \overline{r})^p (c - \overline{c})^q I(r,c) \;, \tag{10}$$

where $\overline{r}$ and $\overline{c}$ are the centroid for the candidate object. We apply this result to obtain a direction of principal axis by centroid of detected candidate region. Angle of principal axis moments is obtained as

$$\theta \;=\; \frac{1}{2} arctan(\frac{2\mu_{11}}{\mu_{20} - \mu_{02}}) \;, \tag{11}$$

where θ denotes an angle between basis horizontal coordinate and principal axis of region. Figure 5 portrays a sequence of successful license plate identification.

4.4 Candidate Decomposition

Information extracted from image and intensity histograms plays a basic role in image processing, in areas such as enhancement, segmentation and description.
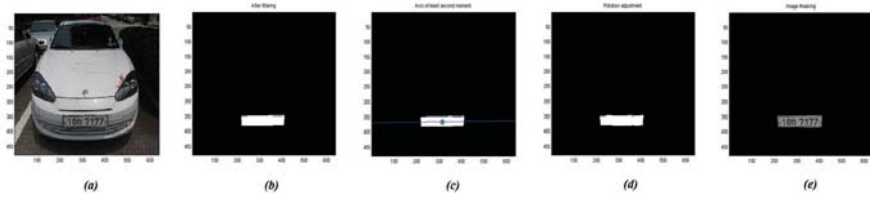
Fig. 5. Illustration of license plate segmentation: (a) an LP image, (b) detected candidate after filtering, (c) principal axis, (d) rotation adjustment, and (e) extracted candidate

In this section, verification and detection of the VLP region as well as character segmentation are considered and discussed in this study. Once the candidate area is binarized the next step is to extract the information. At first, regions without interest such as border or some small noisy regions are eliminated; the checking is made by height comparison with other plate characters height. Fig. 7 shows the results for verifying predetermined alphanumeric characters.

5   Experimental Results and Conclusions

All experiments have been done on Pentium-IV 2.4 GHz with 1024 MB RAM under Matlab environment. In the experiments, 150 images were used the size is 640*480 pixels, some images which are shown in Fig. 6. The images are taken from (a) different illuminations (strong sunshine, shadow and night time), (b) complex scenes where several objects such as trees, light post in front of vehicles, (c) various environments in campus parking, access areas and more than one license plates in the same image and (d) damaged LP as bent or old. They were taken in distance from 2 up to 8 meters and the camera was focused in the plate region. Under these conditions, success of LP detection has reached to more than 94%.

A common drawback of the proposed VLPD system is the failure to detect the boundaries or border of license plates. This occurs when vehicle bodies and their license plate possess similar colors. In our experiments, in 19 images vehicle bodies and license plate have similar color; among them in 11 images license

Fig. 6. Example images: (a) different illuminations, (b) complex scenes, (c) various environments and (d) damaged license plates
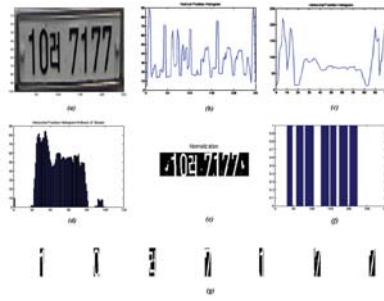
Fig. 7. Steps for verify predetermined alphanumeric characters (white back ground LP): a) extracting candidate region, (b) vertical position histogram with LP border, (c) horizontal position histogram with LP border, (d) horizontal position histogram without LP border, (e) view of normalization candidate region after removing border and noisy area, (f) vertical position histogram (seven peaks for predetermined seven alphanumeric characters in LP region), and (g) character extraction
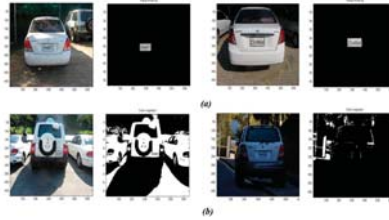
Fig. 8. Sample images: (a) successful plate identification and (b) algorithm failure to detect license plates

plates were detected successfully. Fig. 8(a) shows successful plate identification, where vehicle bodies and their license plates possess similar colors. Fig. 8(b) shows the result in which our detecting algorithm failed to detect license plates.

The proposed method, an adaptive LP detection method, is used for detecting white license plate pixels in nonuniform outdoor illumination conditions. Constants on equations $(7-9)$ are determined in empirically. In our experiments, in 55 images the LP color is white; among them using those constants in 53 images candidate region detected successfully.

In conclusion, a new method is adopted in this paper to select automatically statistical threshold value in HSI color space. In the proposed method candidate regions are found by using HSI color space. These candidate regions may include LP regions; geometrical properties of LP are then used for classification. The proposed method is able to deal with plates (candidate regions) under independent orientation and scale of the plate. More than one license plate can be extracted in the same image. Finally, VLP regions containing predetermined LP alphanumeric character are verified and detected by using position histogram.

While conducting the experiments, different illumination conditions, complex scenes, varied distances between vehicle and camera often occurred. In that case,

the result that has been confirmed is very much effective when the proposed approach is used. However, the proposed method is sensitive when vehicle bodies and their license plates possess similar colors. We leave these issues to be considered in future studies.

References

1. Jiao, J., Ye, Q., Huang, Q.: A configurable method for multi-style license plate recognition. Pattern Recognit 42(3), 358–369 (2009)
2. Huang, Y.P., Chen, C.H., Chang, Y.T., Sandnes, F.E.: An intelligent strategy for checking the annual inspection status of motorcycles based on license plate recognition. Expert Syst. with Applications 36(5), 9260–9267 (2009)
3. Anagnostopoulos, C., Anagnostopoulos, I., Loumos, V., Kayafas, E.: License plate-recognition from still images and video sequences: A survey. IEEE Trans. Intell. Transp. Syst. 9(3), 377–391 (2008)
4. Jia, W., Zhang, H., He, X.: Region-based License Plate Detection. J. Network and comput. Applications 30(4), 1324–1333 (2007)
5. Anagnostopoulos, C., Anagnostopoulos, I., Loumos, V., Kayafas, E.: A License Plate-Recognition Algorithm for Intelligent Transportation System Applications. IEEE Trans. Intell. Transp. Syst. 7(3), 377–392 (2006)
6. Deb, K., Jo, K.H.: HSI Color based Vehicle License Plate Detection. In: IEEE IC-CAS, pp. 687–691. IEEE Press, New York (2008)
7. Xu, Z., Zhu, H.: An Efficient Method of Locating Vehicle License Plate. In: IEEE ICNC, pp. 180–183. IEEE Press, New York (2007)
8. Zhang, C., Sun, G., Chen, D., Zhao, T.: A Rapid Locating Method of Vehicle License Plate based on Characteristics of Characters Connection and Projection. In: IEEE Conf. on Industrial and Applications, pp. 2546–2549. IEEE Press, New York (2007)
9. Matas, J., Zimmermann, K.: Unconstrained Licence Plate and Text Localization and Recognition. In: IEEE Int. Conf. on Intell. Transp. Syst., New York, pp. 255–230 (2005)

A Video-Based Indoor Occupant Detection and Localization Algorithm for Smart Buildings

Ling Chen, Feng Chen, and Xiaohong Guan

Department of Automation, Tsinghua University, Beijing, 100084, China
Chenling.thu@gmail.com

Abstract. In buildings, a practical sensing system to collect occupant location information has great importance in improving occupants' comfort and utilizing energy more efficiently by optimizing control strategies of lighting, HVAC devices and elevators. We implement a practical algorithm for occupant detection in use of existing video camera hardware. In our system, we present a novel blob segmentation method based on rule and propose a fast template-based head detection algorithm that matches directly on gradient maps other than edge maps. The accuracy is improved and can satisfy the need of the control system in smart buildings. The speed is about twice as fast as traditional algorithms.

Keywords: Head detection, occupant detection, smart building.

1 Introduction

In smart buildings, occupant location information is valuable in many applications, such as saving energy, improving occupant comfort and training human behavior models for research purpose. *First*, each year, approximately 39% of energy and 68% of electricity are consumed by residential and commercial buildings in USA[1]. In other countries, buildings also account for a significant part of energy consumption. To improve the efficiency of energy utilization, optimizing system control by sensing occupancy has been proven effective. *Second*, to satisfy the comfort demand of building occupants, the knowledge of occupant locations can help adjust illumination, temperature and humidity automatically. *Third*, understanding regular patterns for human behavior in buildings is the basis for many scientific researches, such as the optimal design of low energy consumption architectures. However, the large quantity of real location data from daily life is difficult to obtain manually. Thus, a sensing system that provides occupant locations automatically will be an effective tool.

Passive infrared motion sensors, RFID, wireless sensor networks (WSN) and video cameras are all feasible approaches for occupancy detection. But infrared motion sensors cannot tell the exact number of occupants. RFID and WSN both require occupants to actively wear tags before entering the building. As for video camera, it has the apparent advantages that it doesn't require the awareness of occupants and that video surveillance systems are now already available in most buildings.

However, video-based indoor occupants detecting and localization also face lots of challenges because traditional approaches are not practical in our application. *First*,

tracking techniques are not suitable for real-time daily use. Although it helps improve accuracy, a robust tracking algorithm requires to process images continuously in high frame rates (e.g. 30fps in [2]), which has not been reached by the processing speed of today's video surveillance systems yet. For light control, HVAC and elevator dispatching, it is not necessary to record occupants' walking trajectories frame by frame. *Second*, in indoor environment, lots of occlusions occur due to desks and chairs, and there are various appearances of occupant gestures like sitting, standing and walking. So, those approaches based on full-body models are not feasible, such as building hierarchical shape templates proposed in [3]. *Third*, as for the approaches based on face detection, capturing frontal face cannot be guaranteed. Yohei [4] combines face and head detection for a surveillance system, but requires large amounts of training samples and may be sensitive to viewpoint variations.

Considering the above challenges, our research focuses on head contour based detection. Since most of the surveillance cameras are installed at high positions, heads are usually visible without occlusions. Besides, head contours are not sensitive to viewpoint variations, so the algorithm can be transplanted into a new scene with a few adjustments. Additionally, when occupant rotates the orientation of the face or changes his gesture, head contours also keep stable. Our main contribution in this paper is that we present a novel blob segmentation method based on rule and a fast template-based head detection algorithm that matches directly on gradient maps. The speed of the fast head detection algorithm is about twice as fast as traditional algorithms and the accuracy is also improved.

2   Overview of the System Structure

The scheme of the processing steps includes Background Updating, Foreground Segmentation, Head Detection and Occupants Locating.

In Background Updating, we adopt Gauss Background Modeling. Since in offices occupants often sit along desks for a long period without moving and thus are easily being updated into background, we slow down the updating speed within a certain area, like those places in front of the desks and computers.

Foreground Segmentation and Head Detection will be introduced in details in the next two sections. As for Occupant Localization, we *first* compute corresponding coordinates of foot based on head coordinates. The foot plane can be treated as an affine transformation from the head plane. *Second*, to obtain a unified coordinate system, we also employ a matrix to transform the coordinates from camera view to the coordinate system of top-down view. The affine transformation includes scale, translate, shear, rotate, and trapezoid. The matrix is estimated by the least squares method.

By converting the detection results of different cameras into a unified coordinate system, we finally come up with an integrated result combining several scenes, dynamically showing the locations of all the occupants appearing in the surveillance area, as showed in Fig. 1(c).

(a)   (b)   (c)
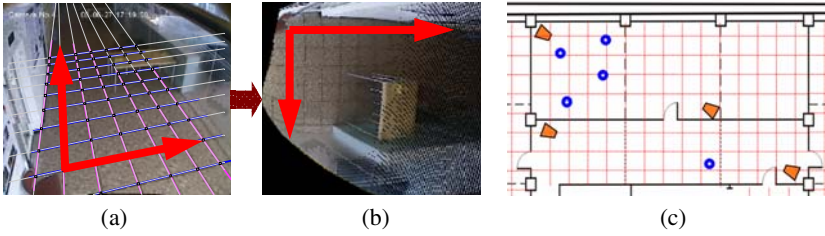
Fig. 1. (a) Shows the scene from the camera viewpoint; (b) Shows the scene after affine transformation from (a); (c) is the floor plan with detected occupants marked as circles, while the trapezoid represents cameras

3   Foreground Segmentation

Our approach of foreground segmentation makes use of geometry constraints of an occupant candidate at one particular position of the scene.

In the preprocessing phase, foreground extracted from background subtraction is eroded, expanding the gaps between adjacent blobs.

We treat each blob as a continuous set of vertical lines with certain lengths. So at each coordinate x, there lie k_x line segments. The jth line segment's top and down coordinate is denoted as $y_{top}(x, j)$ and $y_{down}(x, j)$. The parameters of one occupant candidate are defined by $\{ X_{left}^{(p)}, X_{right}^{(p)}, Y_{top}^{(p)}, Y_{down}^{(p)} \}$, denoting the left, right, top and down boundary's coordinates of the pth segmented blob respectively.

The basic idea of the algorithm is based on rule. $D_{top}^{(TH)}$, $D_{down}^{(TH)}$, $\Delta_{top}^{(TH)}$, $\Delta_{down}^{(TH)}$ are all corresponding threshold values.

From left to right and up to down, search among blobs' vertical lines for the first unanalyzed segment, denoting the starting line of the first occupant candidate $p = 0$.

While still exists unanalyzed line

    Initialization:
$$p = p +1, \ X_{left}^{(p)} = x, \ Y_{top}^{(p)} = y_{top}(x, j_x), \ Y_{down}^{(p)} = y_{down}(x, j_x)$$

Repeat

    Update candidate's parameters:
$$Y_{top}^{(p)} = \max\{ y_{top}(x, j_x), Y_{top}^{(p)}\}$$
$$Y_{down}^{(p)} = \min\{ y_{down}(x, j_x), Y_{down}^{(p)}\}$$

    Find next unanalyzed line $L'(x+1, j_{x+1})$

Until cannot find line L' or L' does not belong to current occupant:
$$\left| Y_{top}^{(p)} - y_{top}(x+1, j_{x+1})\right| > D_{top}^{(TH)}$$
$$\text{or } \left| Y_{down}^{(p)} - y_{down}(x+1, j_{x+1})\right| > D_{down}^{(TH)}$$
$$\text{or } \left| y_{top}(x-1, j_{x-1}) - y_{top}(x+1, j_{x+1})\right| > \Delta_{top}^{(TH)}$$
$$\text{or } \left| y_{down}(x-1, j_{x-1}) - y_{down}(x+1, j_{x+1})\right| > \Delta_{down}^{(TH)}$$

If cannot find line L'
      If the pth occupants' height and width doesn't satisfy requirements
         Discard: $p = p - 1$
      End If
      Denote lines $X_{left}^{(p)}..X_{right}^{(p)}$ as 'analyzed'
      Search next unanalyzed line $L(x', j_{x'})$
Else if L' doesn't belong to current occupant
      $p = p - 1$
      Reset this line $L'(x+1, j_{x+1})$ as the starting line of a new candidate
End If
End While

(a)   (b)   (c)
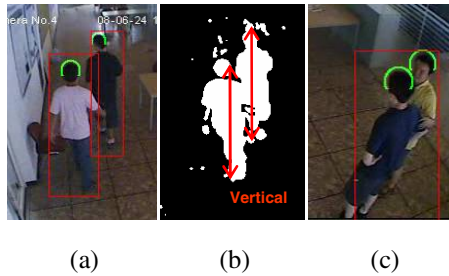
Fig. 2. Examples for Foreground Segmentation. (a) Input image and segmentation result; (b) Background subtraction in (a); (c) Examples that foregrounds are difficult to segment.

In simple cases, the above algorithm is effective, shown in Fig. 2(a) and (b). But in some complex situations, e.g. shown in Fig. 2(c), when the two occupants were captured walking closely to each other or even one person stands in the interior of a foreground blob, they can't be segmented based on above algorithm and will be detected in following head detection part.

4   Head Detection

For the obtained candidate area, that is, the upper part of segmented foreground blobs, we filtered through hair color model first. Then, dilate and erode those remaining areas, eliminating noises.

There are two main categories of template based contour detection: the approach based on Poisson process and another based on gradient. Although the Poisson [5] approach takes into account both clutter distribution of background and the target, it requires a highly precise target model. Due to head contours' differences between individuals in various scenes, this approach is more suitable for detecting rigid objects rather than head contours. The gradient approach [6] evaluates the similarity between target and template based on both distance and direction. We adopt this approach.

The gradient approach proposed in [2] is based on Canny edge detection [7] and Euclidian Distance Transformation (EDT), which has following two defects: *First*, Canny edge detection and EDT both need huge computational cost. The last step of edge linking in Canny edge detection is implemented through recursion. As for Distance Transformation, many literatures are dedicated to lower its computation complexity, e.g. [8] proposed a fast and exact method with linear time complexity. *Second*, the last step in Canny edge detection is two-threshold edge linking. Threshold setting will directly affect the quality of edge detection. If the threshold is set too high, those contours with low contrast to background will not be linked, for example, when hair color is close to the color of background; on the contrary, if the threshold is set too low, lots of useless edges will be linked. Our approach of head detection that directly matches on the gradient map, which does not need Canny edge detection and EDT, and thus avoids above defects.

4.1 Template Based Head Detection

First, compute the gradient map with Sobel operators and eliminate those noise points. Second, match the template at candidate locations. According to the conclusion of [9], the shape of upper 60% omega is the optimal shape of the template. The size of the template is determined by the camera-to-scene calibration.

| | | | | |
| (a) | (b) | (c) | (d) | (e) |
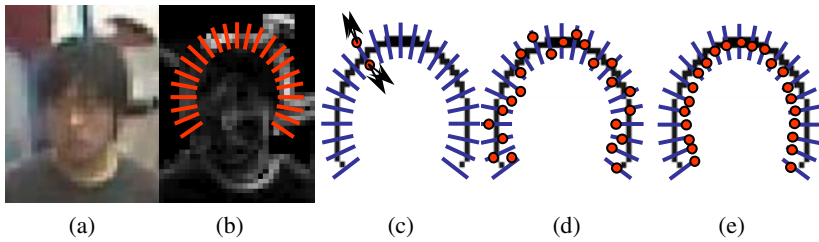
Fig. 3. (a) Input image; (b) Gradient map with Measure Paths; (c) Detection process; (d) An illustration for incoherent optimal point sequence; (e) Coherent optimal point sequence

As shown in the Fig. 3(b), assuming the template is superimposed underneath the head candidate *(x, y)*, $\bar{m}_i$ and $\bar{v}_i$ are the coordinates and the unit normals of the ith pixel in the template. We call the short lines shown in Fig. 3 Measure Path. They are line segments along the ith pixel's normal. Then search for the point with local maximum magnitude in gradient map along the Measure Path. Assume that within the range of *(-p, +p)*, *k* points of local maximum magnitudes are found. We evaluate the optimal point among these local maximums in the following way.

$$s_i = \max\{e^{-\lambda D(j)}\vec{v}_i \cdot \vec{O}(\vec{C}(D(j), \vec{m}_i))\} \ \lambda = 0.25 \ j \in \{1, 2, .., k\} \tag{1}$$

In which, *D (j)* represents the distance between the template pixel *i* and the jth local maximum point along its normal direction. $\vec{C}(D, \vec{m}_i)$ represents the coordinates of the

point that is D away from template pixel i in the direction of normal. And $\vec{O}$ denotes the unit gradient vector on the gradient map.

The above method takes into account jointly the distance between the optimal point to the template pixel i and the gradient direction of this point. s_i represents the score of each template pixel i.

Additionally, the optimal points found in above methods may not belong to one continuous curve actually, probably due to the clustered background or other texture. For example, in the figure below, comparing the optimal points represented by red dots in the left figure and in the right, we find that although their $\sum s_i$ may be almost the same, the right one is more likely to be a real head than the left, because the deviation in the right figure is probably caused by a thin occupant. To evaluate the coherence of the sequence of selected optimal points, compute as following first:

$$D_i'(j^*) = D_i(j^*) - \frac{1}{n}\sum_{i=1}^{n} D_i(j^*) \tag{2}$$

Then count in the sequence $D_i'(j^*), i \in \{1, 2, .., n\}$ how many two adjacent number pairs have opposite signs and denote it as β. For example, for the Fig. 3(d), $\beta = 12$, for Fig. 3(e), $\beta = 1$. β/n can be the indicator to measure the coherence of the optimal points sequence. So, at candidate (x, y), the matching score is:

$$S(x, y) = [(1/n)\sum_{i=1}^{n} s_i] * [1 + (0.5 - \beta/n) * \gamma] \tag{3}$$

In our experiment, $\gamma = 0.3$, which weights the importance between the basic score and the smoothing factor β.

4.2   Experiment for Head Detection Algorithm

To test this part of algorithm of head detection, we select a video sequence of pedestrian crowds captured in the scene of underground to compare the approach based on

                    (a)   (b)

Fig. 4. (a) Results from our approach that directly matches on gradient map. Total pedestrians: 35, detected: 30, false alarm: 2, computing time in Matlab: 52s; (b) Results from the approach based on Canny Edge Detection and EDT. Total pedestrians: 35, detected: 25, false alarm: 3, computing time in Matlab: 108s. (Circles in the figures mark those undetected and false alarms.)

Canny edge detection and EDT with our approach on a computer with Dual Core 1.83GHz CPU.   We only search filtered hair areas.

We tested 10 pictures with 366 heads in total. Fig.4 is one of the testing pictures. Our approach detected 306 heads, reaching 84% detection rate, while the Canny & EDT method detected only 279 with a detection rate of 76%. In all the pictures, the false alarm rate of our method is lower. Besides, the time spent for our method is 1.35 second/head on average, while the other needs 2.87 second/head. Fig.5. shows the testing result for each picture.
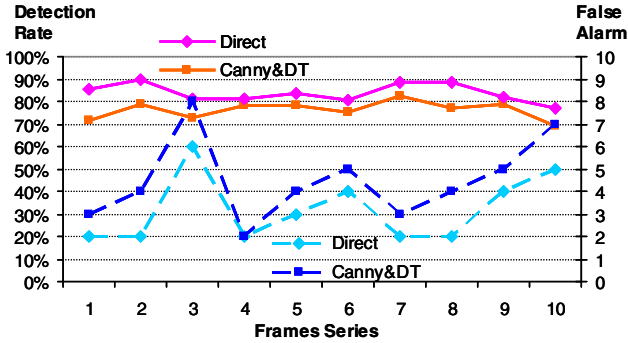
Fig. 5. Comparison between the two approaches

The higher accuracy of our approach is due to following reasons. First, in Canny edge detection, the Gaussian smoothing step has the side effect of eliminating small gradient changes and smoothing those originally rough contours which might cause false alarms. Second, during Distance Transform, for those edge pixels, their values will be set to zero. But in case when the real head contour is not detected in Canny edge detection because its magnitude is lower than the linking threshold, for these undetected contour pixels, their distance values will be contrarily given a high number, which will lead to low score for matching and failure in head detection. In our approach, the edges are not detected in advance, thus we don't need a global threshold for edge linking. Instead, we locally look for the relative maximum magnitude within a certain range in gradient map. This helps scoring reasonably in those low contrast circumstances.

5 Results of the Whole System

In our experimental building, we installed one camera in a classroom, two cameras in the opposite corners in a large room and another two cameras in the aisle, recording videos of daily life of the occupants working in that building. The frame size is 352*288 and the sample rate is 12fps. Some experiment results of this video system are shown in the figure below.

(a) No.1 Aisle   (b) No.2 Aisle   (c) No.3 Office   (d) No.4 Office (e) No.5 Classroom

Fig. 6. Selected frames of detection results captured from five cameras in the building

We select clips from each camera for test. From the results summarized in table 1, we see that in the aisle the detection rate reaches 93% on average with false alarm rate 10%. Most false alarms are caused by shadows or having taken shoulders as detected heads. The reason why the result in No.1 Aisle is superior to No.2 is that in No.2 Aisle, when occupants pass through the scene, their heads are only visible within a short distance.

In rooms, the correct detection rate drops to 81% on average due to occlusions from desks, occupants sitting long time and weak illumination, while the false alarms rise to 13% due to shadows, occupants' complex gestures and clustered background. Another thing to mention is that the results only consider occupants who are completely and clearly visible in the scene. Those who sit far away from the camera (like in No.3, those sitting at up left part in the picture) or those whose figures are incomplete (like in No.5 the occupant in the lower left corner) are not taken into account. The result of No.3 is worst because whenever occupants come in or out, the moving chair will change position and become foreground.

Table 1. Detection Results for Different Scenes

Camera No.	Frame Numbers	Total occupants	Detection Rate	False Alarm Rate
No. 1 (Aisle)	550	932	88%	11%
No. 2 (Aisle)	1800	2934	95%	10%
No. 3 (Office)	850	5173	78%	18%
No. 4 (Office)	1200	1760	86%	12%
No. 5(Classroom)	950	5419	82%	8%

The performance of the system is better than the underground experiment because head detection, as the last module of the whole system, only searches candidate areas filtered by previous modules. So the accuracy of the motion detection and the foreground segmentation modules will also have significant effect on the final results.

The biggest problem for testing whole day long is the abrupt changing of illumination, for example, turning on light, which results in vital destruction to the background model and cause failure in the motion detection step. If the system is designed for practical use, a set of algorithms of automatically illumination detection and background correction will play a key role in enhancing the system's robustness.

6   Conclusions

This work designs a video-based occupant detection and localization algorithm to satisfy the demand for localization information in smart buildings, used for energy saving, human comfort improving and real data collecting for the research of human behavior model. Considering the specialty in this scene, we focus on template based head detection. A segmentation method based on rule and a fast head detection algorithm that matches directly on the gradient map are proposed and implemented. The results demonstrated these approaches are efficient and accurate for practical use.

Acknowledgement

References

1. Optimize Energy Use, National Institute of Building Sciences (2008),
   http://www.wbdg.org/design/minimize_consumption.php
2. Zhao, T., Nevatia, R., Wu, B.: Segmentation and Tracking of Multiple Humans in Complex Environment. IEEE Transactions on Pattern Analysis and Machine Intelligence 30(7) (2008)
3. Lin, Z., Davis, L.S., Doermann, D., DeMenthon, D.: Hierarchical Part-Template Matching for Human Detection and Segmentation. In: Proc. International Conference on Computer Vision (2007)
4. Ishii, Y., Hongo, H., Yamamoto, K., Niwa, Y.: Face and Head Detection for a Real-Time Surveillance System. In: International Conference on Pattern Recognition (2004)
5. Wu, Y., Hua, G., Yu, T.: Switching Observation Models for Contour Tracking in Clutter. In: Computer Vision and Pattern Recognition (2003)
6. Birchfield, S.: Elliptical Head Tracking Using Intensity Gradients and Color Histograms. In: IEEE Conference on Computer Vision and Pattern Recognition (1998)
7. Canny, J.: A Computational Approach to Edge Detection. IEEE Transactions on Pattern Analysis and Machine Intelligence 8(6) (1986)
8. Cuisenaire, O., Macq, B.: Fast and Exact Signed Euclidean Distance Transformation with Linear Complexity. In: International Conference on Acoustics, Speech, and Signal Processing (1999)
9. Garcia, J., Lobo, N.V., Shah, M., Feinstein, J.: Automatic Detection of Heads in Colored Images. In: Proceedings of the 2nd Canadian Conference on Computer and Robot Vision (2005)

Spatial Relation Model for Object Recognition in Human-Robot Interaction

Lu Cao, Yoshinori Kobayashi, and Yoshinori Kuno

Graduate School of Information and Computer Science, Saitama University
255 Shimo-Okubo, Sakura, Saitama 338-8570, Japan
caolu@cv.ics.saitama-u.ac.jp

Abstract. Carrying out user commands entails target object detection for service robots. When the robot system suffers from a limited object detection capability, effective communication between the user and the robot facilitates the reference resolution. We aim to develop a service robot, assisting handicapped and elderly people, where most of the user requests are directly or indirectly linked to some objects in the scene. Objects can be described using features like color, shape, size etc. For simple objects on simple backgrounds, theses attributes can be determined with satisfactory results. For complex scenes, position of an object and spatial relation with other objects in the scene, facilitate target object detection. This paper proposes a spatial relation model for the robot to interpret user's spatial relation descriptions. The robot can detect a target object by asking the user the spatial relationship of the object and some known objects automatically recognized.

1   Introduction

1.1   Background and Motivation

Service robots are designed to be able to perform preprogrammed physical tasks, act under the direct control of a human or autonomously under the control of a preprogrammed computer. In recent years, with aging and social development, they are expected to carry out user requirements through intuitive instructions, for instance, fetching objects for handicapped and elderly people, rather than needing to be programmed by experts.

In a typical service robotics scenario, a robot is instructed by a human user to act upon a specific object. To achieve this aim, both participants need to negotiate their internal conceptual representations linguistically to identify the referent [1]. Between humans, such communication is fairly simple and straightforward, because humans have the ability to specify reference objects by their class names such as "the cup is on the table". Even in a situation where it is difficult to name all of the trivial objects, humans naturally communicate by employing pointing gestures. Robots, however, have limited perceptual capabilities that often preclude accurate recognition of broadly similar objects and, moreover, may not have access to the necessary world knowledge that would identify the object by class [2]. A corresponding to the HRI instruction, an accommodated perceptual way to the robot may be more like that: "the

white little reflecting object is on the brown round table (Assuming that the robot has known about table)." By means of this elaborate description, it complicates to establish an effective joint between human and robot. Thus, the spatial configuration and the position of the object relative to the robot itself can be used for linguistic reference. Our research therefore concentrates on such an option. Namely, to interpret the user command, the robot locates the object by spatial differentiation between objects.

1.2   Previous Work

Dating back to the work by Winograd [3], there has been a great deal of research that inquires into the ways robot systems understand the scene or task through interaction with the user [4][5][6]. These studies, however, have dealt with objects that can be described by simple word combinations such as 'blue box' or 'red ball'. In our application domain, objects are usually more complex, and are thus not so easily recognized by the robot. For example, we may want the robot to bring us a bag of potato chips, in which the package has various colors. However, we may not seem to use complex expressions to describe such complex objects. Thus, we performed observation experiments of human-human interactions and have found that humans usually use simple expressions to describe complex objects [7]. For example, humans usually use only one major color to describe multicolor objects. The system should understand such human expressions to achieve user-friendly interaction. Based on this finding, we have proposed a vision system identifying multicolor objects even when the user mentions one color [7]. We have developed an integrated vision system by combining an automatic object recognition method and an interactive one. The basic strategy is that the system first tries to recognize the object asked by its user. If the system cannot detect the object or make a mistake, the system turns into the interactive mode.

   In the interactive mode, a human user may be asked to instruct the robot to describe the target object by its shape, color and spatial relationship. However, we considered interaction mainly about color in our previous work [7]. In this paper, we address the problem in interaction about spatial relationship.

1.3   Qualitative Spatial Knowledge as a Communicative Method

The psychologist Jackendoff [8] has noted that there are many ways to describe what an object is, but few ways to describe where an object is. Human can specify an object by its shape, color and texture but hard to describe the location. For example, Levin [9] has demonstrated a case where a person could image what an object looked like but could not image the spatial relationship of objects.

   To address this problem, a powerful strategy for achieving reference in human-human communication should be considered more closely. Qualitative spatial reference then serves as a necessary bridge between the metric knowledge required by the robot, and more 'vague' concepts that build the basis for natural linguistic utterances, as suggested by Hernàndez [10]. Whereas many objects may have some particular color, size or texture—which therefore give rise to more potential confusion, or 'distractors', for a referential expression—the position of an object is generally uniquely

defined; if identified sufficiently restrictively, only one object is in a given place at a time. This could make the use of explicit positional information a good strategy for achieving unique reference in the human-robot communicative situation as well [2]. In this paper, we focus on positional information for reference resolution and propose an appropriate solution to locate objects by means of spatial relationship representation.

1.4   Related Work

Although there has been considerable research on the linguistics of spatial language for humans, there has been only limited work done in using spatial language for interacting with robots. Some researchers have proposed a framework for such an interface. Moratz et al [11] investigated the spatial references used by human users to control a mobile robot. An interesting finding is that the test subjects consistently used the robot's perspective when issuing directives, in spite of the 180-degree rotation. At first, this may seem inconsistent with human-to-human communication. However, in human-to-human experiments, Tversky et al. observed a similar result and found that speakers took the listener's perspective in tasks where the listener had a significantly higher cognitive load than the speaker [12].

2   Spatial Reference Representation

In addition to mentioning attributes of each single object, we may describe relationships among multiple objects. We may use descriptions comparing a certain attribute such as "the largest object." However, most common expressions may be those representing spatial relationships among objects.

Fig. 1. Object

Fig. 2. Experimental setting

In our previous work [7], we assembled ordinary objects that we may want a service robot to bring, such as food and drink as shown in Fig. 1. Humans can usually recognize such objects when the object name (e.g. potato chips) is mentioned. We examined how humans describe objects when they were not allowed to mention the object by name.

Fig. 2 is a representation of the experimental setting. Ten pairs of participants took part in the experiment. A board was used to separate them. We placed about 20 objects (among those shown in Fig. 1) on top of the table on participants B's side. We also placed one of the same objects on participant A's side. We asked participant A to describe the object without naming the object, and asked participant B to choose

the correct object. Participant A continued with the description until participant B selected the correct object. We videotaped the experiments and examined the descriptions. We examined 227 utterances that described objects.

The results reveal that participants mentioned color, shape, texture/pattern, size, attachment, and material, and the most frequent was color. Although most of the objects shown in Fig. 1 have multiple colors, humans often describe an object by one of the colors of the object. This color is typically either the background color or the largest area color.

In our current research, we performed experiments in the same setting as shown above to collect expressions about spatial relationships. However, we were not able to obtain many examples. We observed a small number of spatial relationship expressions such as "right to the object that I indicated last time." However, under the restriction that object names cannot be mentioned, it was hard for the participants to describe spatial relationships. Thus, we investigated linguistics literature about spatial relationship usage.

2.1 Reference Systems

Humans use reference systems to describe object positions. The relation between human's spatial cognition and language expression has been well studied in the field of psychology, linguistics, and other related fields. Levinson [13] has proposed that humans use three kinds of reference systems: intrinsic, relative, and absolute. In the intrinsic reference system, the relative position of one object (the referent) to another (the relatum) is described by referring to the relatum's intrinsic properties such as front or back. For example, the expression such as "the book in front of you" is good enough to describe the position of the book since the front of a human body is intrinsically determined. On the other hand, in the relative reference system, we use a position of a third entity as origin instead of referring to inbuilt features of the relatum. An example is "viewed from the cup, the pen is to the left of the box." In the absolute system, neither a third entity nor intrinsic features are used for reference. Instead, we use some absolute direction specification terms, for example, such as north and south.

We mainly consider the relative and intrinsic reference systems in our service robot domain. Fig. 3 shows an example scene. In this case, we may say, "The book is in front of the computer display," by using the intrinsic reference system, or "The book is to the left of the mouse (viewed from the speaker or listener)," by the relative reference system. The front direction of a computer display can be determined regardless the viewpoint, whereas we need to specify the viewpoint to determine the left or right of a mouse. The viewpoint is usually omitted when it is either the speaker or listener.
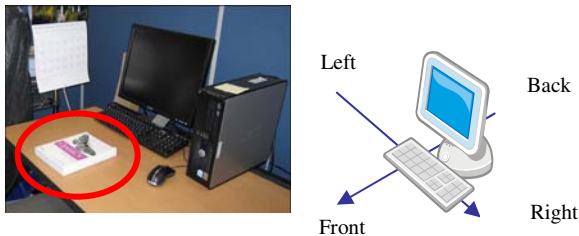
Fig. 3. Intrinsic vs. Relative

2.2   Group-Based Reference

In addition to the above three reference systems, the group-based reference system has been proposed [1][9]. When there are multiple same or similar objects in the scene, humans consider them as a group, describing the position of an object in the group by the spatial relation between the object and the total group. The group-based reference system is considered to be the relative reference system using the group as a relatum. The notion of group can be extended. There are cases that we consider objects as a group even if they are different kinds. For example, we may indicate the can marked by the circle in Fig. 4 as the rightmost object by considering all objects on the table as a group.

Fig. 4. Group-based reference system

2.3   Viewpoint

The relative reference system needs the viewpoint. Depending on the viewpoint, the orientation may be different. In our scenario, if the user and the robot are at the different locations as shown in Fig. 5, the can (indicated by the green box in Fig. 5) is described in two ways. For the user, the can is left to the blue cookie box while for the robot; the can is in front of the blue cookie box. In our current implementation, the robot assumes the viewpoint from the robot if the viewpoint is not explicitly mentioned.

Fig. 5. Viewpoints in the relative reference system

3   Model Designed for Robot System

3.1   Relative Reference System

We have implemented the relative system, which may be most often used to represent spatial relationships. The relative system has three entities: referent (target object),

relatum (reference object), and origin (viewpoint). The origin is often omitted and the default origin is usually the listener, sometimes the speaker. In our implementation, we assume that the origin is the robot (listener). If the user (speaker) and the robot are looking in almost the same direction, the speaker origin coincides with the listener origin. If we take the robot's point of view as origin, all objects are represented in an arrangement resembling a plan view. Thus, the reference axis is a combination of two directed lines through the center of the object as a relatum, which is demonstrated in Fig. 6. Note that the vertical divides the reference plan as left and right parts while the horizontal manages the front and back parts.
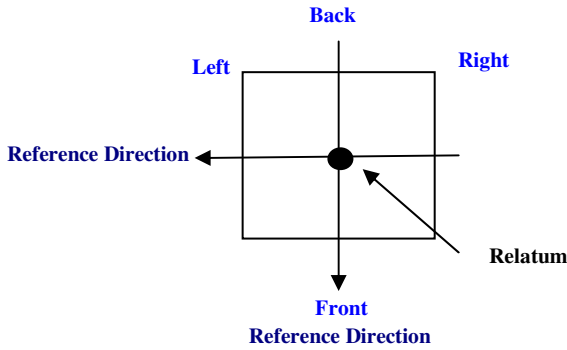
Fig. 6. Relatum and reference direction

For more finely partition, the reference axis is rotated for 45 degrees, respectively, new orientation relations are found, which are called left-front, left-back, right-front and right-back. For combined expressions like "left-front" vs. precise expressions like "strict front", we use the partition presented in Fig.7.
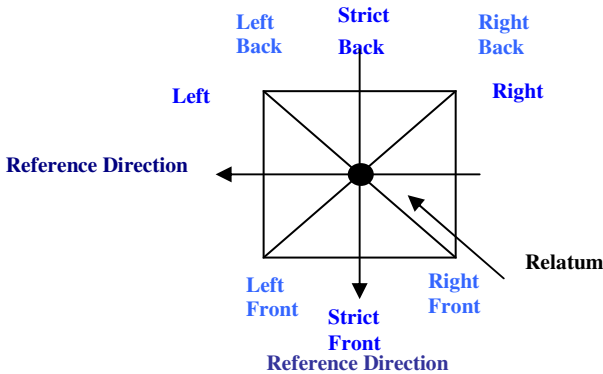
Fig. 7. Combined expressions

Thus, to define the partitions formally, we take the angle Ø between the reference direction and the directed straight line from the relatum to the referent is defined (see Fig.8). The relations between spatial prepositions and Ø can be defined as:

*Referent **front** relatum*   : = -π/2 < Ø < π/2

*Referent **left** relatum*   : = 0 < Ø < π

*Referent **back** relatum*   : = π/2 < Ø < 3/2 π

*Referent **right** relatum*   : = -π < Ø < 0

*Referent **left-front** relatum*   : = 0 < Ø <π/2

*Referent **left-back** relatum*   : = -π/2 < Ø <π

*Referent **right-front** relatum*   : = -π/2 < Ø <0

*Referent **right-back** relatum*   : = -π < Ø < -π/2
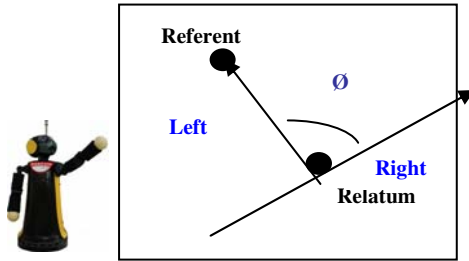
Fig. 8. Relative reference model

3.2 Intrinsic Reference System

The robot needs to know the intrinsic direction of the reference object such as the front direction of a car to use the intrinsic reference system. To do this, the robot needs to recognize the object and to obtain its orientation. We have not developed this orientation detection method. Other things are common with the relative reference system. When the viewpoint is omitted in user's instruction, the expression of the relative system is the same as in the intrinsic system. However, the system can differentiate these two systems easily (although not implemented yet) because objects used for the intrinsic system are restricted and can be registered in advance.

3.3 Group-Based Reference System

The group-based reference system is considered to be the relative reference system using a group as the relatum. We assume the origin is the robot as in the relative reference system. We consider the centroid of all group objects as the center of the relatum as shown in Fig. 9 and determine left, right, or middle. For example, "the middle object" is interpreted as the object closest to the group centroid.

The problem here is that the robot needs to know which objects form the group mentioned by the user. In the current implementation, the robot can consider the same
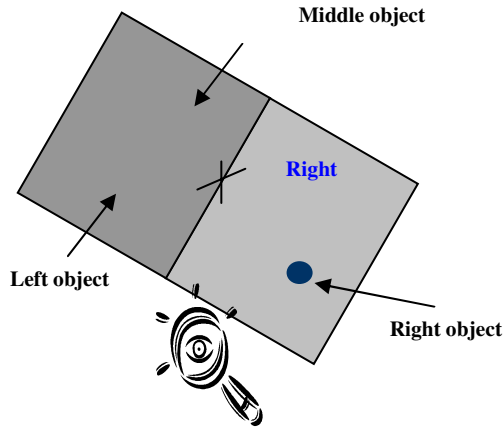
Fig. 9. Group based reference system

attribute objects as a group. For example, if the robot finds two red objects, the robot says that it finds two red objects and asks the user which one. If the robot can automatically recognize multiple objects, the robot can also recognize them as a group. For example, if the robot can recognize three cups, the robot can deal with these cups as a group.

4   Experimental Study

As we mentioned in Section II, it is difficult for us to use the relative system if we are not allowed to mention object names. However, the system may not be able to recognize the objects mentioned by their names. Thus, we adopt the robot initiative way. The robot tells the user the names of objects that it can recognize automatically, and asks the positional information of the target object with respect to the objects mentioned.

   Fig.10 shows an example case. In this case, the robot is supposed to be able to recognize "coffee bottle" and "can" (denoted by the red circles). The target is "cup" (denoted by the light blue rectangle)

   The dialogs to detect the target in this image are depicted below.

> User: Bring me the cup.
> Robot: I can see the bottles and can.
> User: The object is in front of the bottle (or right front of the bottle).
> Robot: Which bottle?
> User: The rightmost bottle.
> Robot: I can see it.

   The robot then pointed to the intended target object.

Fig. 10. Example of interaction by spatial relationships. The system tells the user that it knows the bottles (Coffee) and the can present. Then the user says that the target is in front of the rightmost bottle. The system detects the target indicated by "+" with purple.
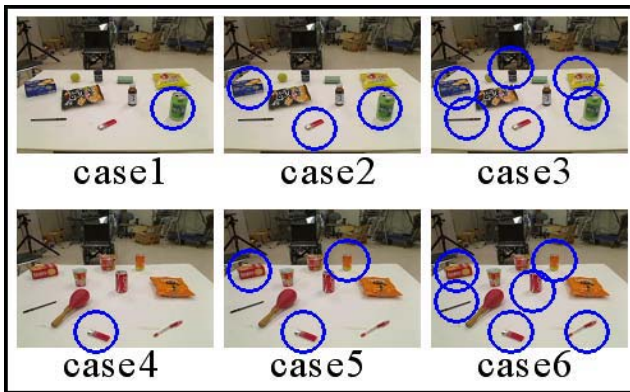
Fig. 11. Examples of experimental scenes. The autonomous method is assumed to be able to recognize the objects indicated by the circles.

We have integrated the spatial relation interaction presented here and the previous color-based interaction [7]. It is desirable that the number of interactions is as small as possible. The system uses the variation of colors in a given image (scene) and the number of known objects to determine what questions to ask the user. If there are various color objects (the number of colors >= 3 in the current implementation), the system first asks the user what color the target object is. If the number of known objects is large (>= 3 in the current implementation), the system says that it knows that such and such objects are present. The user cannot remember many object names if the robot mentions them at once. Thus, the robot mentions two objects even when it has several entries in the known object list. It selects the two that are furthest away around the unknown objects. If both color variation and number of known objects exceed the thresholds, the system takes the color based interaction first.

We have performed experiments with this integrated system. We prepared 20 scenes where 10 different objects were placed on a table. Various different color objects were included in 10 scenes (the number of colors >= 3), whereas color variations were small in the other 10 scenes. In each scene, we assumed three cases that the autonomous object recognition method could recognize one, three, and six objects among the ten objects. Thus, the experimental cases are divided into six cases as shown in Fig. 11. Cases 1 to 3 are large color variation cases, whereas cases 4 to 6 are
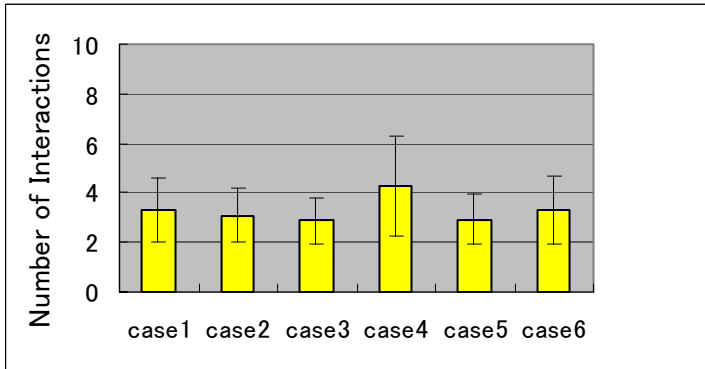
Fig. 12. Average number of interactions in each case

small color variation cases. In cases 1 and 4, the autonomous recognition method is assumed to be able to recognize one of the 10 objects. In cases 2 and 5, three objects; and in cases 3 and 6, six objects.

We used 10 participants. All were university students. Each participant experienced three scenes for each case, that is, 18 trials in total. Thus, we obtained 30 data for each case, 180 data in total. Fig. 12 shows the average number of interactions necessary for the robot to determine the target object. The robot detected the target objects in all cases through about three interactions on average except in case 4 where the number of known objects is only one and the color variation is small. The experimental results confirms the usefulness of interactive object recognition.

5   Conclusion and Discussion

We have proposed a spatial relation model that is applied for the robot to interpret user's spatial relation descriptions. The model represents the target object position by its spatial relation with the relatum (reference object) in the reference system. The relatum can be a group of objects. We have considered two reference systems: the relative reference system and the intrinsic reference system. The robot asks the user the spatial relationship of the target object and some known objects automatically recognized. From the user's utterances, the robot determines the relatum and selects the appropriate reference system to detect the target object.

Experimental results show promising results for our interactive approach. However, we need to discuss two issues. In the current implementation, the robot system takes the initiative and asks the user questions about the target object. This may not seem to be a human-centered design. However, since the purpose of the system is limited to object recognition, this robot-initiative interaction can be acceptable way for users. The important point in our system is that the system can understand natural expressions in the user's answers. Thus, the user can make smooth interaction with the system even though in the robot-initiative way. The second issue is our current assumption that the robot vision can separate at least a part of target object from the

background. In other words, the robot cannot detect an object through interaction if any part of the object is not separated from the background or all parts of the object are merged into other objects. This is related to an old but fundamental problem in object recognition: the relation between segmentation and recognition. In old conventional object recognition methods, segmentation comes first. Then, each segmented region is recognized. Segmentation errors are fatal in such methods. Recent object recognition methods [14] directly try to recognize objects. Our autonomous object recognition method [7] also adopts this approach. However, some segmentation is necessary to make interaction with the user. We think that proper interaction with the user enables the system to recover from segmentation errors. This is left for future work and we will also consider other attributes such as object pose for intrinsic system.

References

1. Tenbrink, T., Reinhard, M.: Group-based Spatial Reference in Linguistic Human-Robot Interaction. Spatial Cognition and Computation 6(1), 63–64 (2006)
2. Moratz, R., Tenbrink, T., Bateman, J.A., Fischer, K.: Spatial knowledge representation for human-robot interaction. In: Freksa, C., Brauer, W., Habel, C., Wender, K.F. (eds.) Spatial Cognition III. LNCS (LNAI), vol. 2685, pp. 263–286. Springer, Heidelberg (2003)
3. Winograd, T.: Understanding Natural Language. Academic Press, New York (1972)
4. Kawaji, T., Okada, K., Inaba, M., Inoue, H.: Human Robot Interaction through Integrating Visual Auditory Information with Relaxation Method. In: Proc. IEEE Int. Conf. Multisensor Fusion on Integration for Intelligent Systems, pp. 323–328 (2003)
5. McGuire, P., Fritsch, J., Steil, J.J., Roothling, F., Fink, G.A., Wachsmuth, S., Sagerer, G., Ritter, H.: Multi-modal Human Machine Communication for Instruction Robot Grasping Tasks. In: Proc. IROS 2002, pp. 1082–1089 (2002)
6. Takizawa, M., Makihara, Y., Shimada, N., Miura, J., Shirai, Y.: A Service Robot With Interactive Vision-objects Recognition Using Dialog with User. In: Proc. First Int. Workshop Language Understanding and Agents for Real World Interaction (2003)
7. Mansur, A., Sakata, K., Kobayashi, Y., Kuno, Y.: Human Robot Interaction Through Simple Expressions for Object Recognition. In: Proc. 17th IEEE RO-MAN, pp. 647–652 (2008)
8. Jackendff, R.: Languages of the Mind. MIT Press, Cambridge (1992)
9. Levine, D., Warach, J., Farah, M.: Two Visual Systems in Mental Imagery: Dissociation of 'What' and 'Where' in Imagery Disorders Due to Bilateral Posterior Cerebral Lesions. Neurology 35, 1010–1018 (1985)
10. Hernández, D.: Qualitative Representation of Spatial Knowledge. LNCS (LNAI), vol. 804. Springer, Heidelberg (1994)
11. Moratz, R., Fischer, K., Tenbrink, T.: Cognitive Modeling of Spatial Reference for Human-Robot Interaction. Intl. Journal on Artificial Intelligence Tools 10(4), 589–611 (2001)
12. Tversky, B., Lee, P., Mainwaring, S.: Why Do Speakers Mix Perspective. Spatial Cognition and Computation 1, 399–412 (1999)
13. Levinson, S.C.: Frames of reference and Molyneux's Question: Crosslinguistic Evidence in Language and Space. MIT Press, Cambridge (1999)
14. Ponce, J., Hebert, M., Schmid, C., Zisserman, A.: Toward Category-Level Object Recognition. LNCS, vol. 4170. Springer, Heidelberg (2006)

Window Extraction Using Geometrical Characteristics of Building Surface

Hoang-Hon Trinh, Dae-Nyeon Kim, Suk-Ju Kang, and Kang-Hyun Jo

Graduate School of Electrical Engineering, University of Ulsan, Korea
San 29, Mugeo-Dong, Nam-Ku, Ulsan 680 - 749, Korea
{hhtrinh,dnkim2005,sjkang,jkh2009}@islab.ulsan.ac.kr

Abstract. This paper describes an approach to extract windows by analyzing geometrical characteristics of building surface. Firstly, building surfaces are detected and then wall region is extracted by using hue color of pixel; this step was well described in our previous works. The non-wall regions are considered as candidates of other components of building such as windows, doors, columns and so on. To extract the windows, the image of candidates is recovered in rectangular shape. Then the ambiguous candidates which have irregular shape, for example, long and thin or very small are coarsely rejected. The geometrical characteristics such as the center coordinates, area, aspect ratio and the aligned coexistence are used for extracting the windows. The proposed approach has been experimented for a database with 150 building surfaces comprising 1607 windows. We obtained 93.34% extraction rate.

Keywords: Window extraction, Aspect ratio, Geometrical characteristics of building surface.

1   Introduction

Landmark detection and reconstruction are very important when we create an intelligent outdoor robot. Because they are the core functions of the other ones such as recognition, exploration, navigation, localization and mapping and so on. Among the natural landmarks of urban environment, building is the best one with dense appearance, rich of geometrical structure and usually locate beside the road and street. To design the detection and recognition functions, the authors usually used multiple images or image sequence, [1,3,4,5,6], where the information of each image can supplement or help one another. The question is: which information can be supported from each image?. So that many authors tried to exploit the information of single image [2,7,8]. In our previous works, [12,13,14,15], we detected and recognized the building surface, extracted wall regions which is hopeful that it makes the available information for designing all functions of outdoor robot even we use single, multiple or sequence images. In this paper, the geometrical structure of building is analyzed more detail by extracting the windows. Fig. 1 shows an overview of proposed method where wall region and surface detection have been done by previous works.
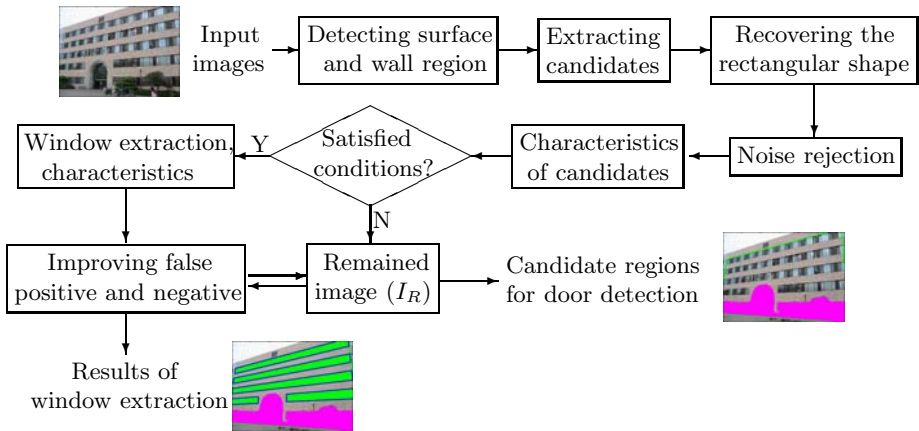
Fig. 1. An overview of proposed method

There are also a few researches for extracting window, [3,4,7,8], but the strate-
gies are different from our proposed method. The works of S. Pu and G. Vos-
selman [7,8] extracted window of building surface from terrestrial laser scanning
data which support the cloud point data with known 3D coordinates. So that the
user easily find building features such as wall, roof, door and so on by comparing
a set of characteristics of position, color, topology, etc [7]. The location of win-
dows did not reflect the laser beam or reflected data locates on another plane so
that it makes the holes on the wall segment. Finally, the windows are extracted
by recovering these holes. The works of S.C. Lee *et al* [3,4] used the combined
information of line segments and calibrated facade to detect and reconstruct the
windows in 3D coordinate. Their results were sound with centimeter error, but
some cases also needs user assistance (semi-automatic). All the above methods
had performed with a small data set; and the occlusion was not mentioned. Here,
we used just only single image and fully automatically extract the windows with
150 images of data set. Then the occluded windows in estimated by context
information.

2   Surface and Wall Region Detection

The processes for detecting building surface and estimating wall regions were ex-
plained in detail in our previous works [12,13,14]. We first detected line segments
and then roughly rejected the segments which come from the scene as tree, bush,
sky and so on. MSAC (m-estimator sample consensus, [10]) algorithm is used
for clustering segments into the common dominant vanishing points comprising
one vertical and several horizontal vanishing points. The number of intersec-
tions between the vertical line and horizontal segments is counted to separate
the building pattern into the independent surfaces. Finally, the boundaries of
surface were found as the green frame in the second row of Fig.2.

Fig. 2. Detection of building surface and wall region

To extract wall region, we used color information of all pixels in the detected surface [14]. Firstly, a hue histogram of surface's pixels is calculated, then it is smoothed several times by 1D Gaussian filter. The peaks in the smoothed histogram are detected. The continuous bins that are larger than 40(%) of the highest peak are clustered into separate groups. The pixels indexed by each continuous bin group are clustered together. The pixels of each group are segmented again where the hue value is replaced by gray intensity information. Finally, the biggest group of pixels is chosen as wall region. Fig.2 illustrates several examples; the first row is the original images; the second row shows the results where the green frames are the boundaries of detected surfaces and the red regions are extracted wall.

3   Natural Characteristics Based Window Extraction

For each surface, a binary image is constructed where the wall regions are considered as background and non-wall regions are considered as the other components of building such as windows, doors, columns and so on, Fig.3(a).

3.1   Noise Rejection

Firstly, the binary image is recovered into rectangular shape where any two windows, with the same size in the real world, have also the same size in the image. From the convex quadrangle of detected surface, we calculate a rectangle whose length and width equal the average lengths of the opposite edges of the quadrangle, respectively. A 2D transformation matrix, H_r, is computed by the quadrangle and the rectangle. The binary image is transformed by H_r and 2D interpolation method, the results can be shown in Fig.3(b). The ambiguous candidates which have irregular shape, for example, long and thin or very small are coarsely rejected. To do so, each column and row of pixels is separately labeled by using labeling connected component. A label whose length, L_{label}, is not satisfied that $T_{Lmin} \leq L_{label} \leq T_{Lmax}$ will be ruled out; where T_{Lmin} is a certain threshold that used for rejecting a thin connection and small

objects, T_{Lmax} is used for discarding the long vertical noises. In the experiments, $T_{Lmin} = 20$ pixels. T_{Lmax} is selected such that the column noise whose length is over $\frac{2}{3}$ the high of recovered surface will be rejected. In practice, several windows spread all the width of surface so that the threshold T_{Lmax} is only used for the vertical direction. The results after rejecting noise is illustrated in the Fig.3(c). The survived foregrounds, I_F, are considered as the candidates of windows.

(a) Non-wall regions

(b) Recovered rectangular shape

(c) Candidates of windows (I_F), after rejecting the noises
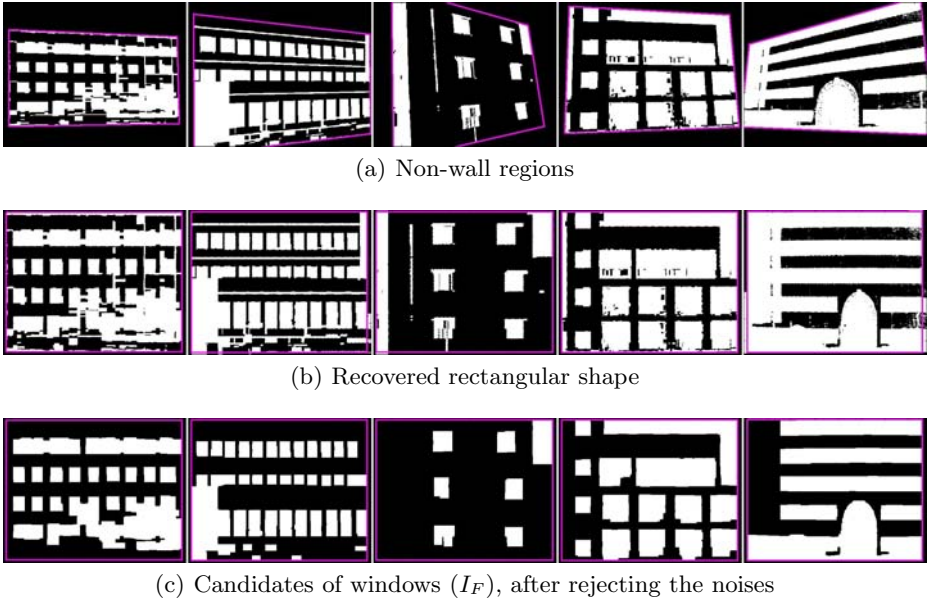
Fig. 3. Noise rejection (each column according to the building in Fig.2)

(a) Geometrical characteristics   (b) Aligned windows: 3 rows, 11 columns
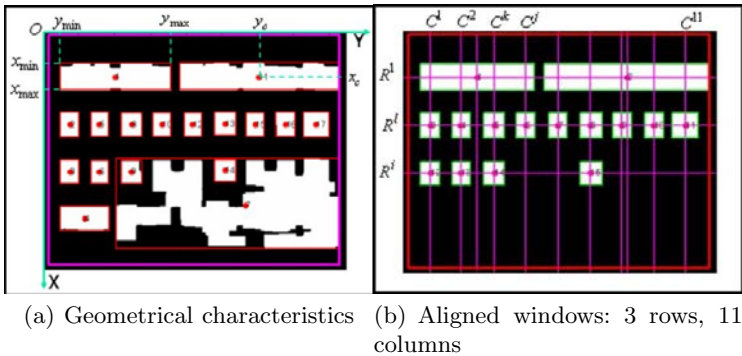
Fig. 4. Illustration of geometrical characteristics of windows

3.2   Geometrical Characteristics of Window

The geometrical characteristics of each candidate are calculated as in Fig.4(a) as follows,

- **Position:** The position is estimated by six parameters including maximum, minimum and centroid coordinates $x_{max}, y_{max}, x_{min}, y_{min}$ and x_c, y_c, respectively. The centroid is calculated by average coordinates of all pixels.
- **Dimension:** The dimension is represented by the high (h) and wide (w) values of each candidate,

$$\begin{cases} h = x_{max} - x_{min} \\ w = y_{max} - y_{min} \end{cases} \tag{1}$$

- **Size:** The size of candidate is defined by the bounding box area, $A = hw$.
- **Aspect ratio:** The aspect ratio, A_{asp}, is calculated that $A_{asp} = \frac{h}{w}$.

3.3   Window Extraction

Let W be a set of windows where R^i ($i = 1, 2, \cdots, M$) are the rows and C^j ($j = 1, 2, \cdots, N$) are the columns of windows. The windows of building are usually coexisted to each other. So that the constraints and relation of two candidates (a, b) are described as follows,

$$a \in W \Rightarrow A^a_{asp} \leq \tau_{asp} \tag{2}$$

$$a, b \in R^i (\text{or } C^j) \Rightarrow \frac{|A^a - A^b|}{A^a + A^b} \leq \tau_A \tag{3}$$

$$a, b \in R^i \Rightarrow \frac{|x^a_{max(min,c)} - x^b_{max(min,c)}|}{h^a + h^b} \leq \tau_c(\tau_c, \frac{\tau_c}{2}) \tag{4}$$

$$a, b \in C^j \Rightarrow \frac{|y^a_{max(min,c)} - y^b_{max(min,c)}|}{w^a + w^b} \leq \tau_c(\tau_c, \frac{\tau_c}{2}) \tag{5}$$

where τ_{asp}, τ_A, τ_c are certain thresholds according to 3.5, 0.25, 0.1, respectively. If two or more windows locate on the same row then their upper and under boundaries are re-calculated by iteratively reweighted least squares [9,11] of correspondent boundary pixels. Similarly, two windows locate on the same column then the left and right boundaries are re-calculated. Now, all windows of one row or column have the same x_{max}, x_{min}, x_c, h or y_{max}, y_{min}, y_c, w, respectively. The extracted windows are aligned into rows and columns as in Fig.4(b); this information will be used to accurate the extraction process in the next section. Fig.5(a) shows more several results of window extraction and alignment. After selecting the windows, the remain of foreground, I_R, is shown in Fig.5(b) which will be used for detecting the doors and improving the results of window extraction, it will be mentioned detail in the next sections. Fig.5(c) is several results after transforming the extracted windows into the original images by using matrix H_r. Here, several cases are falsely extracted; the false negative is shown in the yellow ellipses in the first building of Fig.5(c) where the windows were not extracted; the false positive is shown in the red ellipse of second building where the noises were falsely extracted as windows.

3.4   Improvement of False Positive and Negative

This section is mentioned about improving the false positive and false negative of window extraction. The false positive is usually appeared at the close location with the bottom boundary of the detected surface where many other objects on the yard such as tree, bush, ornament and trained trees and so on are appeared. The size of false positive windows is usually smaller than the size of the windows. So, to improve the false positive, we consider two conditions, the first one is x-coordinate of centroid and the second one is window area. After aligning the windows, we re calculate the area of extracted windows and compute the median of their area $\tilde{A}$. Similarly, the median of x-coordinate of centroid is calculated, $\tilde{x}_c$. A window not satisfied the inequations 6 and 7 is considered as a false positive and then rejected.

$$\Gamma_x = \frac{x_c - \tilde{x}_c}{\tilde{x}_c} \geq \tau_{\Gamma_x} \tag{6}$$

$$\Gamma_A = \frac{A - \tilde{A}}{\tilde{A}} \leq \tau_{\Gamma_A} \tag{7}$$

where τ_{Γ_x} and τ_{Γ_A} are selected thresholds according to 0.5 and -0.5, respectively, in the experiments. To recover the false negative, we consider each position where a column and row of windows intersect to each other as illustrating in Fig. 4(b). The false negative usually happens where a window is occluded by the other objects such as trees, sign and so on. Improving the false negative is very

(a) window extraction and alignment (I_W)

(b) Remain of foregrounds (I_R)
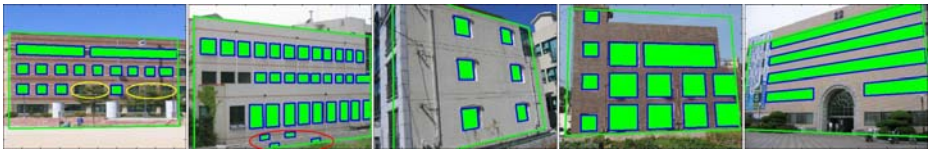
(c) Window extraction

Fig. 5. Examples of results of window extraction

useful for outdoor robot, specially, when robot processes the sequent images, 3D reconstruction and understanding environment and so on. This paper is just a part of our programme of outdoor robot design. So we just discuss about a single image. A intersection between a row R^i and a column C^j should be a position where a false negative window is occluded when it satisfied two conditions. The first condition is that it exists at least one window W^{lk} on another column C^k, which intersects with R^i by extracted window W^{ik}, such that $x_c^{W^{lk}} = x_c^{W^{lj}}$, where $k \neq j$, $i \neq l$ and W^{lj}, W^{lk} have been extracted, a illustration in Fig.4(b). The second condition is that the correlative factors of R^i and C^j are satisfied Inqe.8 (inequations).

$$\begin{cases} F_{w-corr}^{ij} = \frac{|mean(w^{ik})-w^j|}{mean(w^{ik})+w^j} \leq \tau_F, \ k = 1, 2, \cdots, N; k \neq j \\ F_{h-corr}^{ij} = \frac{|mean(h^{lj})-h^i|}{mean(h^{lj})+h^i} \leq \tau_F, \ \ l = 1, 2, \cdots, M; l \neq i \end{cases} \quad (8)$$

where w^j (h^i) is the width (high) of windows on the column C^j (row R^i), note that all windows of one column (row) have the same width (high); threshold τ_F equals 0.15 for experiments. Each intersection of R^i and C^j satisfied the conditions is used for recovering the false negative window. Call I_{cand} is binary image of all recovered windows, the final result I_{recov} is calculated by Eq.9.

$$I_{recov} = I_R \cap \text{not}(I_W) \cap I_{cand} \quad (9)$$

The final results are shown as in Fig.6 where the yellow quadrangles are the recovered windows from the false negative. Fig.6(b) shows the results after transforming into the original images by using matrix H_r, where the magenta regions are the candidates of the doors. It exceeds the performance of this paper and we will discuss later in the section 5.
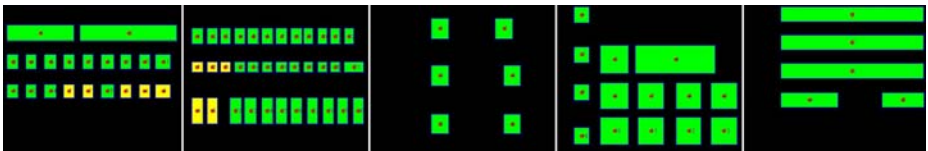
(a) Illustration for recovering the false negative
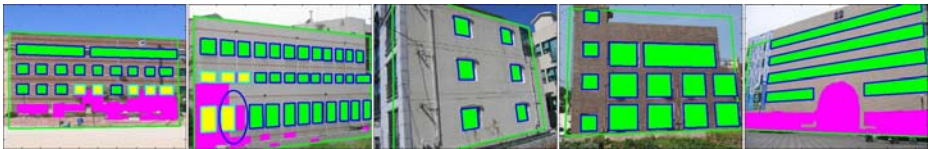
(b) Final results of window extraction

Fig. 6. Final results of window extraction: the green quadrangles are extracted windows, the yellow ones are the recovered windows of false negative, the magenta regions are the candidates of the doors

4   Experiments

The proposed method has been experimented by 150 facets or faces of building images, with total 1607 windows (T). The process given out 1544 extracted windows including 1521 correct cases (TP- true positive), 139 recovered false negative windows, 23 false positive (FP) window extraction. So that 84 windows are omissive extraction (FN-false negative) and 1382 correct windows are given by the first step (results of section 3.3). Fig.7 shows more example of results where the top figure in each sub-image, except Fig.7(f), is the original image with the boundary of detected surface; the under figure is results of extracted windows. Most of FN windows come from a small size window or a quadrilateral region which is not aligned to any row or column as shown in Fig.7(d). The second reason also affects to the FP results; for example, the window is in blue ellipse of second image of Fig.6(b) where a omissive extraction is replaced by the false positive window. Thus, this result gives out 1 for each FP and FN. The top row of Fig.7(f) shows two results of the same building image. The difference is caused by the value of T_{Lmin} in section 3.3. The results are obtained when the values of T_{Lmin} are 15 and 20 pixels, respectively. Therefore, if T_{Lmin} is decreased then a small window should be extracted but the number of FP is

(a)   (b)   (c)   (d)   (e)

(f)   (g)   (h)
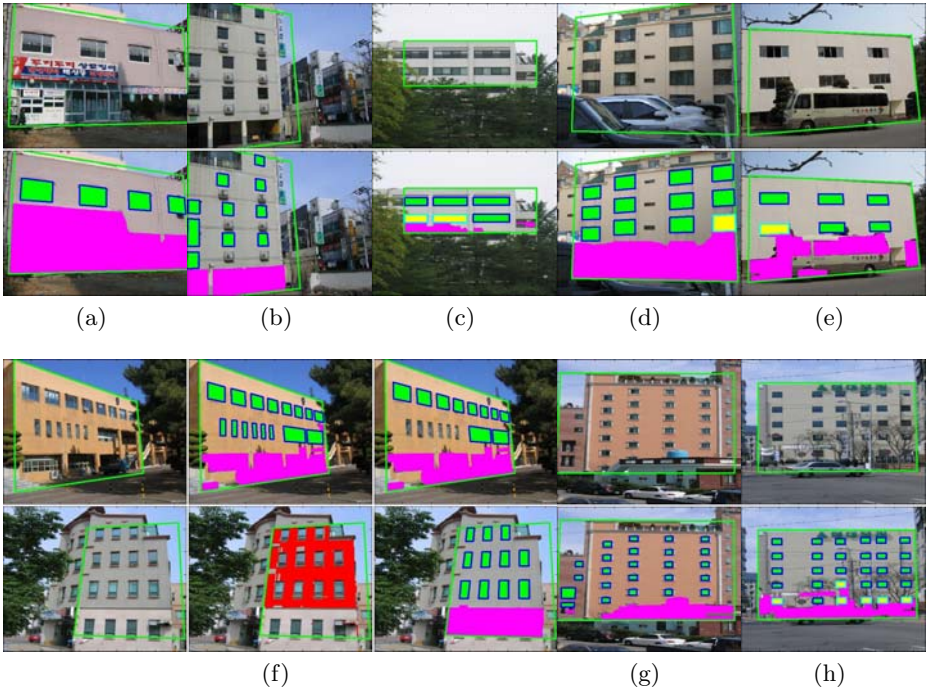
Fig. 7. More examples of results of window extraction; (g,h) according to 15 pixels for the value of T_{Lmin}

increased. In practice, a small window is usually appeared in a far building so their information is not so useful for robot. Another limitation of method is that it is strong dependent on result of wall detection. For example, the building in second row of Fig. 7(f), the wall has two regions with different color while only the bigger region was considered so that several windows can not be extracted. The extraction rate is estimated by,

$$\frac{TP}{T} = \frac{1521}{1607} = 94.65\% \tag{10}$$

or,

$$\frac{T - FN - FP}{T} = \frac{1607 - 84 - 23}{1607} = 93.34\% \tag{11}$$

5   Conclusions and Future Works

The proposed method performs the window extraction for understanding and exploring the environment for outdoor robot. Firstly, the surface and wall region are detected by using line segment and color information of pixels, respectively. The non-wall region is considered as the candidate of windows and then transformed into the rectangular shape. The natural properties of windows such as area, aspect ratio, coordinate position and co-existence in a row and/or column are estimated and used to coarsely extract the windows. The false positive is improved by comparing the median area and coordinate of extracted windows with the corresponding parameters of each window. The false negative is recovered by natural alignments of windows into row and column. We obtained 93.34% extraction rate for 150 building images. The information of results can help the robot to more understand and explore the surrounding environment. For example, the size of building is estimated by the number of floors or number of room on the same floor; robot can explores the door of building, etc.

We are going to detect the door of building and then apply the method for designing the outdoor robot. When the robot is close the building and needs to perform some special tasks, he should able to detect the door. After selecting the windows, the remained regions, magenta regions in Fig. 6 and Fig. 7, are considered as candidates of the doors. Different from the windows, the door usually appears alone in the building image. Therefore, to detect the door, we need more information such as the context and multiple cues from the neighbor regions. For example, the bush or ornament and trained trees appears beside the door; the location is at bottom of building surface and so on.

Acknowledgements

References

1. Cornelis, N., Leibe, B., Cornelis, K., Van Gool, L.: 3D Urban Scene Modeling Integrating Recognition and Reconstruction. In: IJCV, vol. 78, pp. 121–141 (2008)
2. Criminisi, A.: Single-view metrology: Algorithms and applications. In: Van Gool, L. (ed.) DAGM 2002. LNCS, vol. 2449, pp. 224–239. Springer, Heidelberg (2002)
3. Lee, S.C., Jung, S.K., Nevatia, R.: Automatic Integration of Facade Textures into 3D Building Models with a Projective Geometry Based Line Clustering. Computer Graphics Forum 21, 511–519 (2002)
4. Lee, S.C., Nevatia, R.: Extraction and Integration of Window in a 3D Building Model from Ground View images. In: Proc. of int'l Conf. on CVPR, vol. 2, pp. 113–120 (2004)
5. Li, Y., Shum, H.Y.: Stereo Reconstruction from Multiperspective Panoramas. Transactions on Pattern Analysis And Machine Intelligence 26(1) (2004)
6. Pope, A.R.: Model-Based Object Recognition-a Survey of Recent Research, Technical Report, 94-04 (1994)
7. Pu, S., Vosselman, G.: Automatic extraction of building features from terrestrial laser scanning. Int'l Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences 36, 5 (2006)
8. Pu, S., Vosselman, G.: Extracting windows from terrestrial laser scanning. In: Int'l Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences, vol. 36, part 3/W52, Espoo, Finland, September 12-14, pp. 320–325 (2007)
9. Rousseeuw, P.J., Leroy, A.M.: Robust Regression and Outlier Detection. Wiley Interscience, Hoboken (2003)
10. Torr, P.H.S., Zisserman, A.: MLESAC: A New Robust Estimator with Application to Estimating Image Geometry. In: CVIU, vol. 78, pp. 138–156 (2000)
11. Torr, P.H.S.: A Structure and Motion Toolkit in Matlab. "Interactive Advantures in S and M". Technical report MSR-TR-2002-56 (2002)
12. Trinh, H.H., Kim, D.N., Jo, K.H.: Structure Analysis of Multiple Building for Mobile Robot Intelligence. In: SICE Proc., Japan (September 2007)
13. Trinh, H.H., Kim, D.N., Jo, K.H.: Urban Building Detection and Analysis by Visual and Geometrical Features. In: ICCAS 2007, Seoul, Korea (October 2007)
14. Trinh, H.H., Kim, D.N., Jo, K.H.: Supervised Training Database by Using SVD-based Method for Building Recognition. In: ICCAS 2008, Seoul, Korea (October 2008)
15. Trinh, H.H., Kim, D.N., Jo, K.H.: Facet-based multiple building analysis for robot intelligence. Journal of Applied Mathematics and Computation (AMC) 205(2), 537–549 (2008)

Auto-surveillance for Object to Bring In/Out Using Multiple Camera

Taeho Kim[1], Dong-Wook Seo[2], Hyun-Uk Chae[1], and Kang-Hyun Jo[1]

[1] Graduate School of Electrical Engineering, University of Ulsan, Daehak-ro 102,
Nam-gu, Ulsan, 680 - 749, Korea
[2] MOTORWEL Corporation, 221, Automotive Parts Innovation Center, 8BL
Magegok-dong, Buk-gu, Ulsan, 683-420, Korea

Abstract. This paper describes an auto-surveillance system which tracks a person who comes in/out an office using multiple camera system. Furthermore it automatically recognize whether the person bring an object in/out. For this purpose, we set three steps. The first step is detecting a person using MBM(Multiple Background Model) and TMB(Temporal Median Background). The second step is calculation of correspondence between persons detected by different view-point cameras in the multiple camera system. We simply calculate the correspondence based on the principal axis and homography. The last step is generating global color model, which includes every local color model organized by GMM (Gaussian Mixture Model) from each camera, of the person. The global color model represented by GMM checks the temporally varied error and detects the object to bring in or out objects. In the experiment, we show the detected human silhouette by background subtraction and the tracking result by correspondence of multiple views. We also show the color segmentation using GMM and the recognition result for detecting objects brought in/out by the tracked person.

Keywords: MBM(Multiple Background Models), TMB(Temporal Median Background), Multiple camera system, Background subtraction, GMM(Gaussian Mixture Model).

1 Introduction

In recent years, human tracking via vision system is one of the important tasks for HCI(Human Computer Interaction). The simple application systems are already used in many fields and it will be applied to organize intelligent environment such as robot town, intelligent space and automated surveillance system [12,4,3]. For designing those systems, the multiple camera-based method is often used to overcome the limitation of field of view (FOV) that the demerit of using a single camera. Thanks to many researcher, we already have many method to solve problems which multiple camera system has like a calibration, a camera switching and a correspondence within multiple cameras [13]. In this paper, we regard the focus as correspondence of objects between different views and checking the

difference of temporally varied correspondence in the multiple camera system. The related works for corresponding objects in multiple camera system are introduced continuously with various methods. In the geometrical based method, the principal axis of detected object is used as feature. Then the correspondence of object is obtained by finding intersection of features extracted from each object using homography [8]. In another approach, the global color model generated by color histogram of each object is used to match objects by measuring the sum of distance in multiple camera systems [6]. Viewed in the calibration of color space in the different cameras, Porikli [6] proposes color transfer function to match objects among non-overlapping views. Javed et al. [11] also proposes the method to identify same human in non-overlapping views by using color transfer function and relationships of cameras. However, those methods are usually needed a training process in advance such as generating color model or a learning a relationship of cameras. Therefore, we propose a method to identify a same human in the multiple views without a learning process in advance. For obtaining correspondence, the process of proposed methods has three steps which are motion detection, color segmentation and correspondence of people. The process of motion detection by background subtraction is explained in section 2. Section 3 introduces the process of matching people between different views for identifying the same person. The color segmentation and generating blobs from segmented regions are described in section 4. The experimental results and the conclusion are presented in section 5 and section 6, respectively.

2   Motion Detection

In this paper, we use two different period background models, MBM and TMB. MBM has strong merit to construct several clusters of varying each pixel values in long period. However it is not each to recognize that a short period change is a temporally changed background or moving object. Therefore we use TMB to overcome the problem in this paper. The whole progress is described in section 2.1, section 2.2 and Fig.1. In the Fig.1, B_t and M_t are frame amounts of generating MBM and TMB respectively. The amount of frame depends on operator, however the less frame amount of B_t will generate the weaker MBM. Normally B_t is much higher than M_t.

2.1   Background Generation

In this paper, we protect to be blended background model by multiple clusters for each pixel. When we have the camera motion vector from an image pair, we generate MBM that classify clusters for variation of each pixel value. However original MBM is just considered the case of static camera. Therefore, we choose current pixel connected to a pixel in a previous image by the camera motion vector to generate. The whole process is fully described in the [10]. However we use updated σ(standard deviation of each cluster) to the threshold in this paper different with [10]. The next step is classification of clusters and eliminating
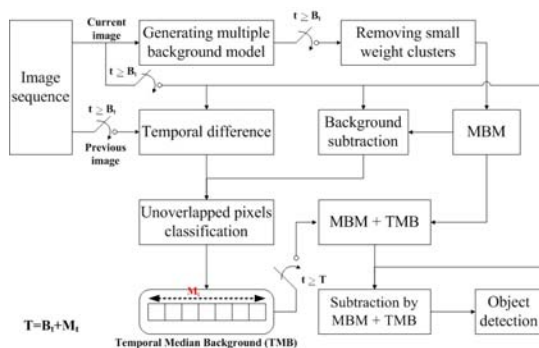
Fig. 1. Block diagram of generating background and detecting objects

small clusters which describe unexpectedly included clusters by moving objects. It is also described in the [10].

2.2   Updating Background by Temporal Median Filter

A background is normally and temporally changed. Therefore background update is important problem not only detection of moving object but also understanding environment in time sequence. One of simple classification of a background is long-term and short-term background discussed by Elías et al. [5]. In their updating strategy, they use temporal median filter to generate initial background and update background using stable information in long time (long-term background) and temporal changes (short-term background). For our case, the progress of generating multiple background concerns classifying the clusters of stable information. However, when a car will park in the scene or leave, it is difficult to arrange the cluster that has previous information. For this reason, we use temporal median filter to eliminate the effect of non-exist cluster. [2].

3   Correspondence between Different Views

It is one of the important security tasks that automatically surveillant systems to check a human bring in or take out objects. For this purpose, we have to track a trajectory of human who moves in the office. We track a moving trajectory of human by correspondence between different views with Kalman filter. The result is described in Fig. 7. To calculate correspondence, we try to detect the principal axis from human silhouette generated by background subtraction. The detected principal axis for one of the camera is translated to the other image plane by homography [13]. Fig. 2 shows the geometrical relationship among multiple views. We use this relationship to overcome the problem correspondence across multiple views. The detailed description is fully described in the [13].

We can use the correspondence information to track a person in each single camera view when tracked human is within the ground plane of any two views.
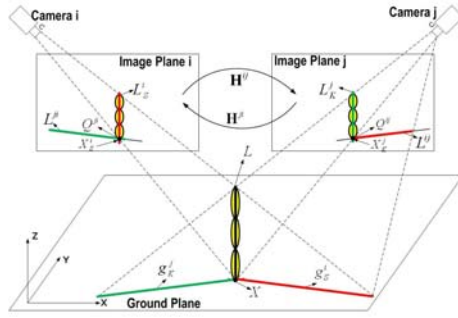
Fig. 2. The geometrical relationship among multiple views

As the principal axis of a person in each view can be detected robustly and accurately, the intersection of the principal axis of the person in one view and the line obtained by transforming the principal axis of the person from another view to the first view is robust and accurate. We use this intersection to track person in single view. As shown in Fig. 2, L_S^i and L_K^j, indicated to principal axes of person in camera i and j respectively, correspond to the same person, the intersection Q^{ij} is used to track a person in camera i.

4   Color Segmentation

The labeled human silhouette from the previous process is segmented to several regions for the each camera. The segmentation of human silhouette is divided by color information. Then the segmented regions are represented as blobs. The number of blobs in an image depends on each image properties and each person is composed of a set of blobs. The set of blobs in a person is denoted as Eq.(1).

$$O_l^{k,N} = \left\{ O_{l,1}^{k,N}, \cdots, O_{l,b}^{k,N} \right\} , \tag{1}$$

O_l is a detected object with a label l. N and k are indices of frame and camera, respectively. b is an index of the number of blobs in a set, i.e., a set of blobs belonged in a person. The pixel values in separated blobs are represented as Gaussian probability distribution with mean and variance. Therefore we obtain the model of probability distribution as follows:

$$p\left(O_{l,b}^{k,N}\right) = A_b \cdot exp\left(-\frac{(c_{i,j} - \mu_c)^2}{2\sigma_c^2}\right) , \tag{2}$$

p in the Eq.(2) is the probability model in the nth blob of the lth object, A is the sum of each blob. μ and σ are the mean and standard deviation as the color value c of the blobs coordinates, respectively. The model of a person is from the distributions of blobs as a GMM by using Eq.(3)

$$p\left(c|\Theta\right) = \sum_{j=1}^{b} p\left(c|\omega_j, \theta_j\right) p\left(\omega_j\right) \ , \tag{3}$$

c is a vector which has color values as components and $\Theta = \{\theta_1, \cdots, \theta_b\}$ are unknown parameters for each Gaussian distribution of blobs $p\left(c|\omega_j, \theta_j\right)$, $j = 1, \cdots, b$. ω_j is the jth labeled blob among known the number of blobs. $\theta_j = \{\mu_j, \sigma_j^2\}$ is statistical values of the mean (μ) and variance (σ^2) of the jth blob. It is able to estimation statistical values as GMM using an expectation-maximization (EM) [1]. For obtaining correspondence, the GMMs of each channel in the HSI color space is used all together.

5   Experiments

5.1   Correspondence and Tracking

We consider to track human that we see the normal movement of human under natural environment. However it has too many unexpected objects in real situation. Therefore we experiment recognition of indoor human movement to reduce noise and proof the performance of correspondence among multiple cameras. Because of indoors, it is no exaggeration to say that moving object is human. We use five cameras that labeled C1–C5 with 5000 images(5×1000) for each camera in the experiment.

Fig. 3 shows the result of multiple background model. We use 100 frames to generate the initial multiple background model. Fig. 4 shows the result of the median background. We use 31 frames to generate median background. Using the multiple background model and median background, we extracted moving a person. We apply this result to detect the principal axis of a person using
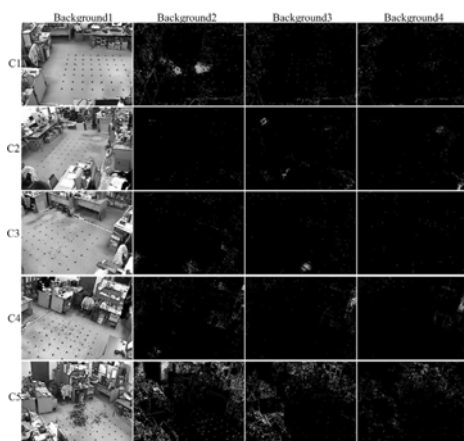
Fig. 3. Multiple background model, each row denotes cameras and each column describes the MBM ordered by their weight

Fig. 4. Median background

Fig. 5. Result of homography between two views

Fig. 6. Result of correspondence between two views

the principal axis moment. After detection of the principal axis of the person, we can obtain to match person between two views using homography constrain. Fig. 5 shows the result of the homography between two views. The bounding boxes of green and yellow are the detected moving person. The lines of green and yellow are the principal axis and the transformed principal by homography. The red points are the centroid and intersection with the principal axis and the transformed principal axis. Fig. 6 shows the result of correspondence between two views. First row is the correspondence between C1 and C2. Second row is the correspondence between C3 and C4. The bounding boxes of green and yellow color are detected a person for each camera. The lines of green and yellow are the detected principal axis and transformed principal axis using homography. The red points are the centroid point and the intersection point which principal axis and transformed principal axis.

Table 1 shows the rate of human region detection. We use 901 images in each camera. We can see the detected result over 99 %. If non-human region is detected larger than human region, human region process is failed. Also, the result of incorrectly detected region is 1.05 %. We can get the accurate result.

Table 1. The rate of human region detection

Camera Image	Total Image	Detected Region	Incorrectly Detected Region
C1	901	893/901, 99.11 %	13/893, 1.46 %
C2	901	885/901, 98.22 %	13/885, 1.47 %
C3	901	901/901, 100.00 %	12/901, 1.33 %
C4	901	901/901, 100.00 %	3/901, 0.33 %
C5	901	901/901, 100.00 %	6/901, 0.67 %
Total	4500	4481/4500, 99.58 %	47/4481, 1.05 %

(a) Trajectory in C1   (b) Kalman filtering
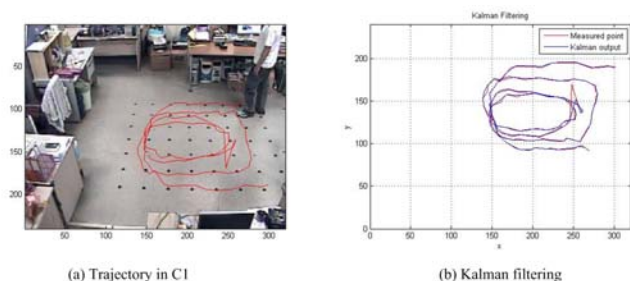
Fig. 7. The result of human tracking in 2D space

Table 2. The rate of correspondence between two views

Stereo Image	Total Object	Corresponding Rate
(C1, C2)	887	867/887, 97.75 %
(C3, C4)	901	875/901, 97.11 %
(C5, C1)	893	875/893, 97.98 %
Total	2681	2617/2681, 97.61 %

However we consider only one person to stay in structured space by multiple cameras. Table 2 shows the rate of correspondence between two views. In the structured space, total object means stayed moving objects at the same time. We can see the corresponding rated over 97 %. From this table, we know that the result obtained accuracy. Fig. 7 shows the result of Kalman filter. [14] We track a person by corresponding points on the plane. Fig. 7(a) is the trajectory in camera 1. Fig. 7(b) is the result of Kalman filtering. The red line is the measuring points. The blue line is the kalman output.

5.2 Global Color Model and Recognition of Objects

The tracked person silhouette is segmented to several blobs described in Section 4. The Fig. 8 shows the example of color blobs described in GMM. For the experiment, we take image sequence of two persons and three objects for two below situations.

(a) Image of detected person

(b) Comparison of GMM
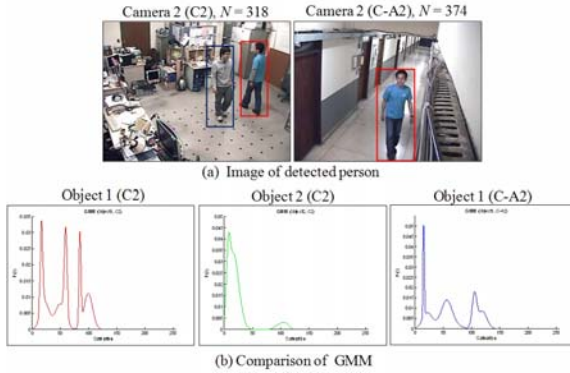
Fig. 8. The results of motion detection

Table 3. The recognition result of object to bring in or take out

| Situation | Tracking | Detection rate of object | |
		Bring in	Take out
Situation 1	97.9%	100%	100%
1-4 Situation 2	76.2%	67.4%	62.1%

  i. Each person moves an object for bring in or take out (total 12 cases)

 ii. Two persons come in together but only one person bring an object in or take out. (total 8 cases)

The recognition result of detecting object that the tracked person brings in or takes out is noted in Table 3.

6 Conclusions

We described how to correspond among multiple images which are taken by multiple structured cameras and recognize objects to bring in or take out. The correspondence of multiple cameras is finding correspondence among simultaneously existed objects in different view. Based on correspondence of multiple cameras, we estimate the relation of objects in different view and track objects in time sequence. We applied the principal axis and the homography to solve the problems correspondence across multi-view. For detecting moving silhouette of human, we use the multiple background model. The several variations for each pixel in time sequence are successfully contained by multiple background model. The principal axis is used to calculate the direction of principal axis for the detected silhouette. By using the principal axis of person, we match the person in structured space by multiple cameras based on homography. In the 3D space, we have the only one intersection of principal axis of human and the ground plane. The intersection is the invariant point in different views. Therefore

we estimate the invariant point of each view by a principal axis of human in a view and a transformed principal axis to the same view using homography. The detected intersections are shown in the experimentation result. The correspondence is constructed according to relationship between intersection points. The experimental result shows the accurate correspondence among multiple views. The result of recognizing object move by a person is described in Table 3. In the result we have 100% of recognition for the situation 1 described in Section 5.2. However the recognition result of detecting object for situation 2 not well enough. It causes by the low tracking result. Because we use gray image in the generating MBM and tracking sequences, it is weak to detect the difference if a person wears red or blue clothes. In the experiment of recognizing object, unfortunately, two persons in the image sequence wear red and blue shirts. Therefore the tracking and recognition result are not good. To overcome this problem, we are recomposing the color-based algorithm.

Acknowledgment

References

1. Dempster, A.P., Laird, N.M., Rubin, D.B.: Maximum Likelihood from Incomplete Data via the EM Algorithm. J. of the Royal Statistical Society Series B 39(1), 1–38 (1977)
2. Haritaoglu, I., Harwood, D., Davis, L.: W4: Who? When? Where? What? A real time system for detecting and tracking people. In: Proc. of Int. Conf. on Automatic Face and Gesture Recognition, pp. 222–227 (1998)
3. Collins, R.T., Lipton, A.J., Kanade, T., Fujiyoshi, H., Duggens, D., Tsin, Y., Tolliver, D., Enomoto, N., Hasegawa, O., Burt, P., Wixson, L.: A System for Video Surveillance and Monitoring, Carnegie Mellon University, Pittsburgh. In: PA, Tech. Rep., CMU-RI-TR-00-12 (2000)
4. Hashimoto, H.: Intelligent space: a commentary on research trends. In: Int. Symposium on Industrial Electronics, vol. 1, pp. 11–16 (2002)
5. Elías, H.J., Carlos, O.U., Jesús, S.: Detected motion classification with a double-background and a Neighborhood-based difference. Pattern Recognition Letters 24(12), 2079–2092 (2003)
6. Porikli, F.M.: Inter-Camera Color Calibration by Correlation Model Function. In: Int. Conf. on Image Processing, vol. 2, pp. 133–136 (2003)
7. Wang, C.K., Cheng, M.Y., Liao, C.H., Li, C.C., Sun, C.Y., Tsai, M.C.: Design and Implementation of a Multi-purpose Real-time Pan-tilt Visual Tracking System. In: Proc. of IEEE Int. Conf. on Control Applications, pp. 1079–1084 (2004)

8. Hu, W., Hu, M., Zhou, X., Tan, T., Lou, J., Maybank, S.: Principal Axis-Based Correspondence between Multiple Cameras for People Tracking. IEEE Trans. on Pattern Analysis and Machine Intelligence 28(4), 663–671 (2006)

9. Morioka, K., Mao, X., Hashimoto, H.: Global Color Model Based Object Matching in the Multi-Camera Environment. In: Proc. of the 2006 IEEE/RSJ Int. Conf. on Intelligent Robots and Systems, pp. 2644–2649 (2006)

10. Taeho, K., Kang-Hyun, J.: Detection of Moving Object Using Remained Background under Moving Camera. Int. J. of Information Acquisition 4(3), 227–236 (2007)

11. Javed, O., Shafique, K., Rasheed, Z., Shah, M.: Modeling inter- camera space-time and appearance relationships for tracking across non-overlapping views. Computer Vision and Image Understanding 109, 146–162 (2008)

12. Murakami, K., Hasegawa, T., Kurazume, R., Kimuro, Y.: A Structured Environment with Sensor Networks for Intelligent Robots. In: IEEE Conf. on Sensors, pp. 705–708 (2008)

13. Seo, D.W., Chae, H.U., Jo, K.H.: Multiple Camera-Based Correspondence of Ground Foot for Human Motion Tracking. Jour. of Institute Control, Robotics and Systems 14(8), 848–855 (2008)

14. Welch, G., et al.: An Introduction to the Kalman Filter. In: SIGGRAPH (2001)

Object Analysis for Outdoor Environment Perception Using Multiple Features

Dae-Nyeon Kim, Hoang-Hon Trinh, and Kang-Hyun Jo

Graduate School of Electrical Engineering, University of Ulsan,
Daehak-ro 102, Nam-Gu, Ulsan, 680 - 749, South Korea
{dnkim,hhtrinh,jkh2009}@islab.ulsan.ac.kr

Abstract. This paper describes the method to know objects for autonomous robot navigation in an unknown outdoor environment. The method segments the objects from an image taken by moving robot on outdoor environment. In the beginning object segmentation, this uses multiple features to obtain the objects of segmented region. Multiple features are color, context information, line segments, edge, Hue Co-occurrence Matrix (HCM), Principal Components (PCs) and Vanishing Points (VPs). The model of the objects for outdoor environment defines their characteristics individually. We segment the region as mixture using the proposed features and methods. Next the stage classifies the object into natural and artificial ones. We detect sky and trees of natural object and detect building of artificial object using the combination of appearance and context information. Then we estimate the dimensions of building. Extensive experiments with the object segmentation and analysis on outdoor environment confirm the validity of the approach.

Keywords: Object segmentation, Multiple features, Object analysis.

1 Introduction

When an autonomous the robot navigation on outdoor environment, it is likely for him to set specific a target. The robot also needs avoid objects when he encounters obstacle, and know where he is and know further path take he. The problem of long range navigation in unknown outdoors environments is not very frequently addressed. Systems that were actually experimented were demonstrated for road following, or motion in rather limited environment conditions. This paper focuses on objects analysis for autonomous robot navigation in an unknown outdoor environment. To analysis an object, we classify object into artificial and natural. Then we define their characteristics individually. We segment the object after the process of preprocessing. We propose a method to segment objects of outdoor environment using multiple features. To analyze and recognize specific object, our method used property of segmented objects. Among multiple features, we present a method to apply the texture with a color information model. To use the color or texture in existing research, we have various color models which is used widely. The kinds of color model are RGB, HSI, CIE, YIQ,

YCbCr, etc. We use HSI color model [1]. Image segmentation can become very difficult, as the image gray value or color alone are rarely good indicators for object boundaries due to noise, texture, shading, occlusion, or simply because the color of two objects is nearly the same. Zhang [2] proposed the color image segmentation method by intensity and color. we propose a method for detecting the faces of building using line segments and their geometrical vanishing points [5,8,9].

Haralick [3] used statistical features extracted from object using gray level co-occurrence matrix in analysis method of texture [3]. We developed and evaluated different implementations of GLCM, using co-occurrence matrix of hue-value instead of gray-level. This paper shortened a processing time taking a displacement vector which is specific direction 135 ° into accounts at HCM [8]. We analyze the features for image segmentation using candidates by definition of object. We use color, line segments, PCs, edge, vanishing point [8]. In additional, we use HCM to detect tree, where H is information of Hue. Segmentation of the image uses as mixture a feature of such multiple candidates. The methods to combine features segment the images as the characteristic of the object.

In the following sections, we describe feature extract for objects of image that present a color, line segment, PCs, edge, vanishing point and HCM. Section 3 presents the methods of region segmentation. Experimental results are shown in section 4. Section 5 concludes the paper.

2   Multiple Features

When the robot navigates on outdoor environment, we classify to segment the object from image to get as the prior knowledge, and then we apply the knowledge of an object. The objects are divided into the natural and the artificial object as a prior knowledge of the human. We present the candidate for a segmented region with natural and artificial object such as sky, trees and building. So, we segment the region by using multiple features. The features are color, context information, HCM, line segments, PCs and vanishing point [8,10]. The feature of **color** uses the range of HSI model. Before accomplishing color segmentation, we convert RGB color space to HSI. This color space describes direct color component in reality such as human vision similarly. It used HSI space at the application to interpret the color of the image. This paper normalizes as color values at 0∼255 range of saturation and intensity. Also, Hue color value has the range of 360 degrees. We found the HSI value with sky, cloud and trees of natural object through repeated experiment. **Context Information** uses information of habitual location in the image. We describes as the percentages of being at the *top, middle and bottom* of an image, (L_{Ti}, L_{Mi} and L_{Bi}, respectively). The y position of all pixels is obtained and the probability of each of them to belong to a certain position is computed. The main drawback of not using context is the overlap between classes, e.g. sky and water, both blues. The system can then easily confuse a water region, at the bottom of the image, with sky, since they have a very similar appearance. Two small image patches are ambiguous

at a very local scale but clearly identifiable inside their context. Specifically, we distinguish two kinds of context information: (i) Absolute context: referred to the location of objects in the image, (ii) Relative context: position of the objects respect to other objects in the images. Some proposals consider both kinds of context [6], while only the relative context is considered by He et al. [7]. **Hue Co-occurrence Matrix** uses co-occurrence matrix of hue-value instead of gray-level. To reduce the computational complexity, only some of these features were selected. The hue value of spatial dependence frequencies are a function of the angular relationship between the neighboring resolution pixels as well as a function of the distance between them [8, 10]. We analyze the spatial characteristics using HCM [8, 10]. HCM $P[i, j]$ is defined by specifying a displacement vector and counting all pairs of pixels separated by distance d and direction ϕ having hue level i, j. This paper uses the feature of HCM in the 135° diagonal directions from a simple original image having hue levels $0, 1$ and 2. **Line segments** have much component in artificial object such as building. The line segment is a straight part of edge of which its length is longer than a certain threshold [9]. The value of threshold is chosen to decide a workspace of robot. Here, the threshold is chosen with 10 pixels for detecting line segments. For each segment, two neighbored regions are extracted. We define a straight line segment with following two conditions. The first step of the line segment detection is the edge detection of image. We used the edge detection function with Canny edge detector algorithm. The function is run in automatically chosen threshold. The second step is line segment detection. We define a straight line as following. A straight line segment is a part of edge including a set of pixels which have number of pixels larger than the given threshold (T_1) and all pixels are alignment. That means, if we draw a line through the ends, the distance from any pixel to this line is less than another given threshold (T_2). **Principal Components** needs to analyze the geometric information in detail. For example, line segments and their lengths with a reference element employ the edges. In our approach, the windows, doors and wall region are detected [10]. The edges and lengths of windows are accurately estimated. The PCs are detected by merging the neighbor parallelograms which have the similar physical properties as color [5]. We based on the RGB color space to present the characters of basic parallelogram. In reality, the light energy coming from different region of the large objects including building to the camera is different. So their intensities with different region are not the same although they located on the same PCs, for example the same wall. The relatively geometrical information of PCs such as the quantity, height and alignment are computed. Those information is available for reconstruction of 3D model of building. After the result of line segment, we find the **Vanishing Points**. The survived segments are processed with assumptions that if an image contains building pattern then this pattern must be upright positioned. The apt direction of building is flowing y-axis. The edges of PCs usually parallel together in the real world. They are projected into 2D image with a common dominant vanishing point (DVP). MSAC (m-estimator sample consensus) algorithm [5] for clustering segments into the vanishing points as suggested in our

previous work [9]. In practice, the building faces within non-orthogonal direction are usually appeared in either multiple faces of building or multiple buildings in the images. So the maximum number of DVPs are calculated for vertical and horizontal directions according to one and five, respectively. So the maximum number of DVPs are calculated for vertical and horizontal directions according to one and five, respectively. For each horizontal cluster, the number of line segments must be larger than a certain threshold (6, in the experiments). The priority of horizontal DVPs is arranged from high to low due to the number of segments in corresponding groups. This paper calculates one dominant vanishing point for vertical direction and five dominant vanishing points in maximum for horizontal direction. The method of object segmentation uses the multiple features in outdoor environment.

3   Region Segmentation of Objects

The goal of segmentation is to simplify and/or change the representation of an image into something that is more meaningful and easier to analyze. We consider images of outdoor environment and we segment each object as sky, trees, building, etc. Region segmentation uses the mixture such multiple features according to the characteristic of the objects.

3.1   Sky and Cloud

This paper uses HSI color model and we find the value of sky and cloud in the image. This method finds the value of HSI to repeated experiment. Also, we use absolute context information for referee to the location of objects in the image. The image divides as a three part at the top, middle and bottom. If the robot travels as regular intervals on outdoor environment, we regard that sky and cloud exist at top in image. Here, we add context information as sky position at top of the image. If there is a different object in the sky, it regards as the region of sky. The method used HSI color model and found the part to correspond to the value of cloud and sky in the image. We find the value of HSI to repeated experiment. The range of sky and cloud is equal to hue, saturation and intensity by [8]. Region segmentation extracts the region of the cloud after the extraction does sky region. Region segmentation of sky and cloud shows in Fig. 1. The image to do region segmentation of sky and cloud is seen at a Fig. 1(b), Fig. 1(c). The merger of sky and cloud show Fig. 1(d).

3.2   Trees

To find the region of trees, we use the value of the HSI in compliance with the experiment which is repeated. Additionally, in order to estimate the similarity between different gray level co-occurrence matrices (GLCM), Haralick [3] proposed statistical features extracted from them. GLCM, one of the most known texture analysis methods, estimates the image properties related to second-order

Fig. 1. Segmentation of sky and cloud region: (a) original image (b) sky (c) cloud (d) the merger of sky and cloud
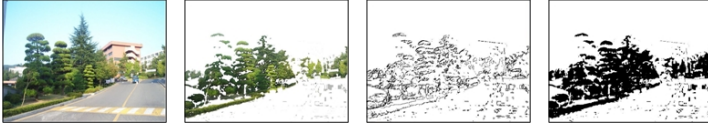
Fig. 2. Comparison results of diverse cues of segments region with trees: (a) original images (b) trees detection using HSI (c) trees detection using HCM (d) trees detection using HSI+HCM

statistics. Each entry (i, j) in GLCM corresponds to the number of occurrences of the pair of gray levels i and j which are a distance apart in original image. We use co-occurrence matrix of hue-value instead of gray-level. To reduce the computational complexity, only some of these features have select. We analyze the spatial characteristics using HCM [8,10]. The HCM $P[i, j]$ is defined by specifying a displacement vector and counting all pairs of pixels separated by distance d and direction ϕ having hue level i, j. Kim et al. [8,10] illustrated how to object HCM in the $135°$ diagonal direction from a simple original image having hue levels $0, 1$ and 2. We have get image segmentation using displacement vector of $135°$ diagonal direction in the HCM. This paper thus attempts to outline an alternative reading of GLCM. Thus, this paper proposes the method of HCM algorithm. HCM analyzes of appearance count of hue value pixel pairs at original image. At first, we use HSI and find the range of hue as shows in Fig. 2(b). Then, we define a range from CM for high frequency regions. Next, we obtain the value of HCM as shown in shows in Fig. 2(c). Finally, we use HCM and HSI together as shown in Fig. 2(d). The method of HSI alone has many noises at the image segmentation. We decrease such noise the method using HCM and HSI together. HSI has been desired from the repeated experiment trials to segment trees regions for the natural object.

3.3   Building

Building has a plane surface that contains PCs as doors, windows, wall region and columns. The processes for detecting building face and estimating wall regions explained in detail at our previous works [8,9]. The first step detects region of trees as algorithm of HSI and HCM described at previous chapter 3.2. The second step of the line segments detection use Canny edge detector. Line segments detection is a part of edge which satisfied two conditions [8,9]. In experiments,

we choose T_1 and T_2 as 10 and $\sqrt{2}$ pixels respectively. The result of line segments detection is shown in Fig. 3(b). Most of the low contrast lines usually do not locate on the edge of PCs because the edge of PCs distinguishes the image into two regions which have high contrast color. We base on the intensity of two regions beside the line to discard the low contrast lines [8,9]. The result is illustrated by Fig. 3(c). The vertical group contains line segments which create an acute angle $20\,°$ in maximum with the vertical axis. The remnant lines are treated as horizontal groups. For the fine separation stage, we use MSAC [5] robustly to estimate the vanishing point. Suppose that the line segments end points are x_1, x_2 such that $l = (a, b, c)^T$; $l = x_1 \times x_2$ and $x_1 = (\acute{x_1}, \acute{y_1}, 1)^T$, $x_2 = (\acute{x_2}, \acute{y_2}, 1)^T$. Given two lines, a common normal is determined by $v = l_i \times l_j$, where $v = (v_1, v_2, v_3)^T$. Hence given a set of n line segments belonging to the lines parallel in 3D, the vanishing point v is obtained by solving the following Eq. (1):

$$l_i^T v = 0; \qquad i = 1, 2, \ldots, n. \tag{1}$$

The robust estimation of v by MSAC has proven the most successful. We calculate five dominant vanishing points in maximum for horizontal direction [8]. The algorithm proceeds in three steps [4,8]. The priority of horizontal vanishing point is dependent on the number N_i of parallel lines in corresponding groups. They are marked by color as following red, green, blue, yellow and magenta color. They are illustrated by Fig. 3(d). The vertical line segments are extended to detect a vanishing points. We base on the number of intersection of vertical lines and horizontal segments to detect and separate planes as the faces of building. Fig. 3(e) shows us the results of face detection. The boundaries of faces define as three steps by Kim et al [8]. The minimum of horizontal lines in left and right faces is N_l and the number of points of intersection is N_i. The ratio of N_l and N_i is larger than given threshold satisfying Eq. (2) with N_T is 0.35.

$$N = \frac{N_i}{N_l} \geq N_T \tag{2}$$

Finally, the mesh of basic parallelograms is created by extending the horizontal lines. Each mesh represents one face of building. Fig. 3(e) shows the results of mesh of face detection.

Fig. 3. The result of building detection: (a) original images (b) line segments detection and trees region (c) survived line segments reduction (d) dominant vanishing points detected by MSAC (e) mesh of basic parallelograms of face

4   Experiment

The image database used in the experiment consist about 1800 images. All the images are taken in Ulsan metropolitan city in South Korea under general condition. These images consist on object of natural and artificial on outdoor environment. The image to do region segmentation of sky and cloud is seen at a Fig. 1. We mark sky in a Fig. 1 by black and cloud in a Fig. 1(c). The result images are presented in a Fig. 1(d) which it merges sky and cloud. Normally, around the leaf for the trees has high frequency. So, we search the object which has proposed HCM algorithm for trees. The result of region segmentation with trees is presented in Fig. 2. Also, we used HSI color model, conversion of RGB to HSI, and found a part to correspond to the value of the trees in image. At last, we find the region of trees as combined features of HSI, context information and HCM. The result of segmentation of trees region is preprocessing for detection of building. Then we remove the high frequency in trees region. Line segments for detecting building were used because of noise reduction. For detecting the faces of building used line segments and their geometrical vanishing points. MSAC algorithm is used to find the vanishing points not only for the multiple faces of building but also for the face having noises as branches of trees or electrical lines. We can see that well result in Fig. 3(e). The meshes of parallelograms can help us to detect more PCs as window, door and so on. In addition, the relation of geometrical properties as the height and the number of windows can be exploited to analyze more information of building in Fig. 4. For example, how many rooms the building has. More examples are illustrated by Fig. 5 where the first row is original images and the second one is result of wall and window detection. We summarize a performance analysis about building characters of Trinh et al. [9]. For each image, the line segments were detected and then the segments coming from the scene as trees, bush, sky and so on were roughly rejected. MSAC algorithm was used for clustering segments into the common dominant vanishing points comprising one for vertical and several for horizontal directions. The vertical segments were extended across the image and the number of intersections between the vertical lines and horizontal segments was counted to separate the building pattern into the independent facets. Finally, the boundaries of facet were found.

4.1   Building Analysis

This paper uses an approach that first estimates environment understanding through region segmentation, then performs building analysis by classifying each region using multiple features derived from the prior knowledge. After knowing the area of an object, an object has analyzed as wall regions and non-wall regions. The face of building usually contains three principal components such as doors, windows and wall. In those components, the color of the windows is sensitive. However, the wall region is insensitive so it is chosen to represent the face of building. It detects by merging the neighbor parallelograms which have similar colors. We use RGB color to represent a skewed parallelogram. A model of each
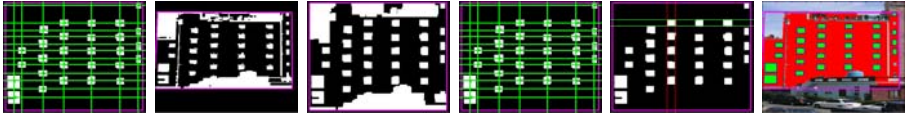
Fig. 4. Geometrical facet analysis as parameter estimation of PCs.: (a) Wall region (b) Candidates of windows (c) Rectangular shape (d) Window alignment (e) Accuracy of boundaries (f) Detection of wall and windows
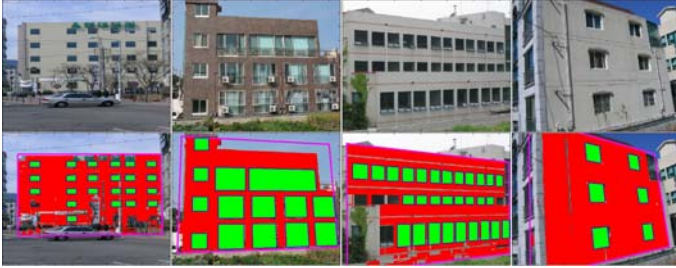
Fig. 5. Some examples of wall and window detection: (a) original images (b) result of wall and window detection

skewed parallelogram is constructed with the RGB color model with the average intensities in the order of strength into six groups. In practice, if two values of component colors are approximate then their order is usually changed by the noise or illumination condition. Thus we define a seventh when any reciprocal differences between three averaged values are less than a given threshold [9]. The wall region detects by merging a neighbored skewed parallelogram. Any two neighbored skewed parallelograms are clustered when they have same order of colors and their average values satisfy conditions as denoted by Eq. 3.

$$|\bar{I}_C^i - \bar{I}_C^j| \leq T; \quad C \in R, G, B \tag{3}$$

where T =35 pixels is threshold. To reduce an noisy region like a cluttered trees which lies in front of a building, we select only a skewed parallelogram having at least one of average intensities larger than 75. Finally, the wall is the cluster which contains largest number of pixels. The color information of pixels corresponding to wall region is used to calculate a hue histogram and quantize into 32 bins. The PCs are apparently appeared in the image if the robot is close the building. The result of detected wall is binaries as shown in Fig. 4(a) and 4(b). The wall region is the background and non-wall region is the foreground. To estimate windows, the facet is recovered in rectangular shape [9]. Fig. 4(c) shows the recovered rectangular facet in relative coordinate and different scale with Fig. 4(b). The noises such as the thin connection, small regions of background and foreground are rejected by several operators like "NOT", "AND" and connected component labeling. Each foreground is considered as a candidate PCs. By computing the center coordinate, area and aspect ratio, the candidates of

Table 1. Estimated size of buildings in Fig. 5

Buildings (from left to right)	Number of floors	Number of room in the same floor
1	6	6
2	4	5
3	3	12
4	3	2

windows are aligned into row and column. The windows of building usually have similar aspect ratio and area so some candidates with far different size and aspect ratio with majority of the others are rejected. Fig. 4(d) shows 28 candidates of windows aligned into 12 rows and 7 columns. The windows with similar area in the same row or column are used for accuracy of window boundaries as in Fig. 4(e). This information uses to estimate the size of building. The number of floors or the numbers of rooms with the same floor are estimated by maximum number of windows in the same columns or rows respectively. Finally, the accurate windows are projected into the original image. Fig. 5 is result of window and wall estimation. Table 1 summaries the estimated size of building. This information is more confident when we apply to reconstruct building by multiple images or image sequences of robot's movement.

5 Conclusion

This paper proposed a method of object segmentation on outdoor environment by using multiple features. Multiple features are color, context information, edge, line segments, HCM, PCs and VPs. Mixing those features, we segment the image to several regions such as sky and trees of natural object, building of artificial object. Here, we use features of color and absolute context information for extract of sky and cloud region. And, we use features of color, edge and HCM for extract of trees region. Also, we use to extract building as color, edge, line segments, PCs and VPs. Then we remove the high frequency in trees region. The meshes of parallelograms can help us to detect more PCs as window, door and so on. Overall the system of this paper segments the region of the object as a mixture by using multiple features. The detailed information was analyzed by estimating the size of a building. The number of floors or the numbers of rooms with the same floor are estimated by maximum number of windows in the same columns or rows respectively. We accomplished the process of preprocessing to know objects from an image taken by moving robot on outdoor environment. In future, we will study how to the objects respect geometric relationships on outdoor environment between objects as well as to apply the method in a set of images containing more objects (road, car, people, etc.).

NARC, the Human Resource Training project for regional innovation through KOTEF, the Human Resource Development Program for Convergence Robot Specialists through KIAT and post BK21 at University of Ulsan.

References

1. Lievin, M., Luthon, F.: Nonlinear color space and spatiotemporal MRF for hierarchical segmentation of face features in video. In: IEEE in Transaction on Image Processing, vol. 13, pp. 63–71. IEEE Press, New York (2004)
2. Zhang, C., Wang, P.: A New Method of Color Image Segmentation Based on Intensity and Hue Clustering. In: Proceeding of International Conference on Pattern Recognition, vol. 3, pp. 613–616 (2000)
3. Haralick, R.M., Shanmugam, K., Dinstein, I.: Texture features for image classification. IEEE Transaction on System Man Cybern. 3(6), 610–621 (1973)
4. Zhang, W., Kosecka, J.: Localization based on building recognition. In: International Conference on Computer Vision and Pattern Recognition, vol. 3, pp. 21–28 (2005)
5. Hartley, R., Zisserman, A.: Multiple view geometry in computer vision, 2nd edn. Cambridge University Press, UK (2004)
6. Singhal, A., Jiebo, L., Weiyu, Z.: Probabilistic spatial context models for scene content understanding. In: IEEE Computer Society Conf. on Computer Vision and Pattern Recognition, vol. 1, pp. 235–241 (2003)
7. He, X., Zemel, R.S., Carreira-Perpinan, M.A.: Multiscale conditional random fields for image labeling. In: IEEE Computer Society Conference on Computer Vision and Pattern Recognition, vol. 2, pp. 695–702 (2004)
8. Kim, D.N., Trinh, H.H., Jo, K.H.: Object Recognition of Outdoor Environment by Segmented Regions for Robot Navigation. In: Huang, D.-S., Heutte, L., Loog, M. (eds.) ICIC 2007. LNCS, vol. 4681, pp. 1192–1201. Springer, Heidelberg (2007)
9. Trinh, H.H., Kim, D.N., Jo, K.H.: Facet-based multiple building analysis for robot intelligence. Applied Mathematics and Computation 205(2), 537–549 (2008)
10. Kim, D.N., Trinh, H.H., Jo, K.H.: Region Segmentation of Outdoor Scene Using Multiple Features and Context Information. In: International Conference on Intelligent Computing, CCIS, vol. 15, pp. 200–207 (2008)

Pseudo Invariant Line Moment to Detect the Target Region of Moving Vessels

Jia Ke[1,2], Yongzhao Zhan[2], Xiaojun Chen[2], and Manrong Wang[2]

[1] School of Business Administration, JiangSu University
[2] School of Computer Science and Telecommunication Engineering, Jiangsu University
Zhenjiang, 212013, Jiangsu, China
`kejia@ujs.edu.cn`

Abstract. In order to get the features of moving vessels at the port correctly and track the target quickly and efficiently, we combined the advantages of traditional invariant moments and invariant line moments, and proposed a object recognition algorithm based on pseudo invariant line moments. Using this algorithm, first we get the calculation regions of the objects in an image, then do edge detection to the calculation regions and get the pseudo invariant line moments by calculating binary image. The experimental results show that the algorithm can not only get the regions of moving objects quickly and accurately, but also can predict the directions of moving objects effectively. This algorithm is applied in the intelligent video monitoring system of moving vessels successfully.

Keywords: moment; line moment; pseudo invariant line moment; intelligent video monitoring; moving target detecting.

1 Introduction

Intelligent video monitoring is an important research field of computer vision, and it has a wide range of applications in the traffic information monitoring and the monitoring of many important places such as bank, hotel, station and so on[1-2].The intelligent monitoring system of customs house dock is oriented to the application of practical event detecting, and it has an in-depth research into the methods for detecting moving objects and predicting movement directions under the condition that the position of the camera is fixed.

As for expressing of the feature of a image target, domestic and foreign experts have proposed many methods. All of these methods can be divided into two types, and one of them is based on regional feature and the other based on border feature. The usual methods of object recognition based on regional feature involve invariant moments first introduced by Hu [3], polar radius invariant moment [4] and etc. The methods according to border feature mainly include rectangular approximation [5], polygonal approximation [6], and invariant line moment [7], improved invariant moment [8-11] and so on. A method based on regional characteristic has a reflection of overall information. Therefore, when the serial images collected by camera are influenced by light or noise, the information will have great change and we can't get correct result. In addition, the calculation of regional moment is in terms of the existent regions of the

entire image, so the work of calculation is considerable and it is not conducive to real-time processing. The methods based on border characteristic utilize the fringe information of the targets to analyze and interpret images. As regional information, fringe information is affected by factors such as light much less. Moreover, the sum total of the pixels on the edge is much less than that of the region, thus it's beneficial to calculating and accessing. However, the method based on border characteristic usually requires extracting the outside contour closed curve of a target effectively, and it is a task costing calculation to extract the closed curve of the serial images. If affected by light or noise, the algorithmic perhaps gets invalidated. What's more, inside contour is very significant for some objects as well. After considering a variety of factors, this paper presents a fast algorithmic based on pseudo invariant line moment, which has combined the integrated characteristics of invariant moment representing regional feature with the quick features of the invariant line moment representing border feature, and realizes the method of calculating several contours inside and outside synthetically with the application of the formula for boundary moment. As the results of the experiments demonstrated, the algorithmic proposed in this paper can not only get the features of the region the moving target located in relatively exactly, it can also improve the speed of calculation.

2 Features of Pseudo Invariant Line Moment

2.1 Features of Pseudo-line Moment

As for digital image $f(x,y)$, if it is piecewise continuous and only has a finite number of non-zero points in the X-Y plane, it can be proved that all of its moments exist. The moments of a region are calculated by the use of all the points in the region, thus it's hardly influenced by noise. The $p+q$ -order moment of $f(x,y)$ is defined as follows:

$$m_{pq} = \sum_x \sum_y x^p y^q f(x,y) \Delta x \Delta y \quad \text{where} \Delta x = 1, \Delta y = 1. \tag{1}$$

We use the Canny operator to get the binary edge image of a target region. Fig. 1 (a) is the classical image of fruits, Fig. 1 (b) is the binary edge image of Fig. 1 (a). It can be concluded that the binary image has many contours inside and outside which represent

S	T	S
T	M	T
S	T	S

(a) fruits   (b) edge detection diagram   (c) connective pixel diagram
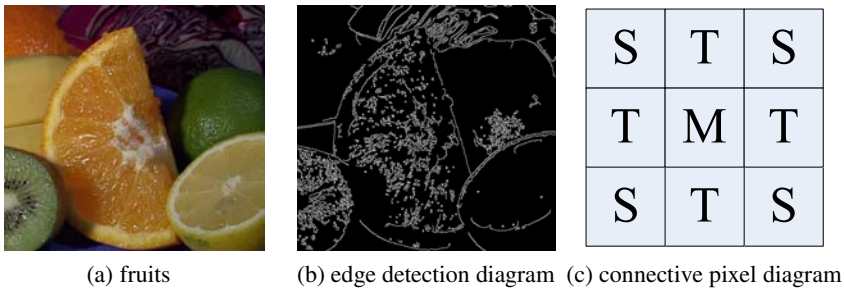
Fig. 1. Picture fruits' edge detection diagram and connective pixel diagram

the features of the target. However, the calculation of line moments can only be aimed at a certain closed contour, so we can't do the calculation of line moments directly. In this paper, a method of expressing features by pseudo-line moments is proposed and it can be used to calculate on several contours inside and outside.

After calculating the binary edge, $f(x,y)$degenerate into:

$$f(x, y) = \begin{cases} 1 & Contour \\ 0 & other \end{cases} \tag{2}$$

If the pixels of M, S and T meet the condition shown in Fig. 1(c), M and T are connective; M and S are Indirect- connective.

If we know that the number of a image's pixels in connective area is N, the $p+q$ -order moment (pseudo-line moment) of the image is defined as follows:

$$m_{pq} = \sum_{i=1}^{N} x^p y^q \Delta l \tag{3}$$

Since all the fringe pixels are continuous, Δl meets the condition of$\Delta l \to 1$, the geometric significance of pseudo-line moment can be described as the moment calculation of all the discrete contours of the binary image showed as Fig. 1 (b). Since pseudo-line moment can do calculation on a number of connective contours and needn't distinguish inside contour from outside one, it has a good commonness.

The central moment of a pseudo-line moment is defined as follows:

$$\mu_{pq} = \sum_{i=1}^{N} (x - x_0)^p (y - y_0)^q \tag{4}$$

Where $x_0 = \dfrac{m_{10}}{m_{00}}$, $y_0 = \dfrac{m_{01}}{m_{00}}$ is the Central coordinate of $f(x,y)$.Unitary center pseudo-line moment can be expressed as follows:

$$N_{pq} = \frac{\mu_{pq}}{\mu_{00}^r} \text{ Where } \gamma = \frac{p+q}{2} + 1, \mu_{00} = m_{00}. \tag{5}$$

2.2   Features of Pseudo Invariant Line Moment

According to the theory of algebraic invariants, Hu proposed a serial of moment invariants for the use of shape recognition. The "invariant" here means that some eigenvalues of an image remain unchanged in the following conditions: translation, rotation and scale transformation [3].

Hu has defined seven values which consist of the non-linear combination of standardized 3-order center moments, and they still remain unchanged when the position, scale or direction of a target have changed. The formulae of pseudo invariant line moment and those of invariant moment defined by Hu have the same form, and they are defined as follows:

$$T_1 = N_{20} + N_{02} \tag{6}$$

$$T_2 = (N_{20} - N_{02})^2 + 4N_{11}^2 \tag{7}$$

$$T_3 = (N_{30} - 3N_{12})^2 + (3N_{21} - N_{03})^2 \tag{8}$$

$$T_4 = (N_{30} + N_{12})^2 + (N_{21} + N_{03})^2 \tag{9}$$

$$T_5 = (N_{30} - 3N_{12})(N_{30} + N_{12})[(N_{30} + N_{12})^2 - 3(N_{12} + N_{03})^2] + \\ (3N_{21} - N_{03})(N_{21} + N_{03})[3(N_{30} + N_{12})^2 - (N_{21} + N_{03})^2] \tag{10}$$

$$T_6 = (N_{20} - N_{02})[(N_{30} + N_{12})^2 - (N_{21} + N_{03})^2] + 4N_{11}(N_{30} + N_{12})(N_{21} + N_{03}) \tag{11}$$

$$T_7 = (3N_{21} - N_{03})(N_{30} + N_{12})[(N_{30} + N_{12})^2 - 3(N_{21} + N_{03})^2] + \\ (3N_{12} - N_{30})(N_{21} + N_{03})[3(N_{30} + N_{12})^2 - (N_{21} + N_{03})^2] \tag{12}$$

$$m_{pq} = \sum_{i=1}^{N} x^p y^q \Delta l_i \quad \text{Where} \quad \Delta l_i = \sqrt{(x_i - x_{i-1})^2 + (y_i - y_{i-1})^2} \tag{13}$$

$$\mu_{pq} = \sum_{i=1}^{N} (x - x_0)^p (y - y_0)^q \Delta l_i \tag{14}$$

$$m_{np} = \frac{1}{A} \sum_{i=1}^{A} \left(\frac{r_i}{r_0} \right)^p \quad \text{Where A is region area,} \quad r = \sqrt{(x - x_0)^2 + (y - y_0)^2} \tag{15}$$

$$\mu_{np} = \frac{1}{A} \sum_{i=1}^{A} \left(\frac{r_i - r_0}{r_0} \right)^p \quad \text{Where A is region area,} \quad r = \sqrt{(x - x_0)^2 + (y - y_0)^2} \tag{16}$$

2.3 Computational Complexity Analysis for Pseudo Invariant Line Moment

In order to demonstrate the rapid character of the computation for pseudo invariant line moment, we need to analyze its computational complexity and compare it with other kinds of invariant moments. In this paper, formula (1) shows the calculation of Hu's invariant moment; formulae (13) and (14)[7] show that of invariant line moment; formulae (15) and (16)[4] show that of polar radius invariant moment. The computational complexity concluded is shown in Table 1. From Table 1, we know that the computational complexity of Hu's invariant moments is $O(8N^2)$, extremely time-consuming. For polar radius invariant moments, due to the need for solving the center of mass of targets, there is also a higher computational complexity and the total complexity is $O(7N^2)$. Both invariant line moments and pseudo invariant line moments need to calculate contours. Besides, invariant line moments need to calculate the contour points of the outermost contour line. Even though the amount of contour points invariant line moments finally get is less than that of pseudo invariant line moments, the calculation of contour points for pseudo invariant line moments is a linear increase. So the total computational complexity of pseudo invariant line moments, which is $O(2N^2+7N)$, is less than that of invariant line moments, which is $O(3N^2+7N)$. By

Table 1. Comparison of invariant moment' Time complexity

Type	Domain	Contour line	Edge	Invariant moment
Hu's invariant moment	$O(N^2)$			$O(7N^2)$
Invariant line moment	$O(N^2)$	$O(N^2)$	$O(N^2)$	$O(7N)$
pseudo invariant line moment	$O(N^2)$	$O(N^2)$		$O(7N)$
Polar radius invariant moment	$O(N^2)$			$O(5N^2)$

comprehensive analysis, we conclude that the computational complexity of pseudo invariant line moments is minimum.

3 Moving Object Recognition Technology Based on Pseudo Invariant Line Moments

By the derivation above, we get a rapid method to calculate pseudo invariant line moments. Next we will use the method to detect moving objects and predict objects' moving direction.

3.1 Getting the Calculation Domain of the Moving Object

Usually we can use an external rectangle to represent the calculation domain of a target [5]. If an object's shape is asymmetry, the center of gravity of the external rectangle is largely different from that of the target. If the target doesn't fill most of the regions of the rectangular box, the external rectangle also can't represent the target effectively. In Fig. 2 (a) A is the center of the external rectangle, and B is the actual center of gravity of the object. The two centers have great deviation and there is interference on the calculation of the target. Fig. 2 (b) has shown the area of a external rectangle. The shape of a common ship is close to shuttle. And if the middle section of the ship were much narrow, the entire ship would only occupy a much small part of the external rectangle. Therefore, if we use the external rectangle to represent the object approximatively, there will be very great error.

In this paper, we use an external polygon to represent the calculation domain of an object approximatively, and the polygon could approach the object's boundary by maximum when the amount of calculation is small. In Fig. 2 (c), $L1,L1'$ are parallel lines whose slope is K1, so are $L2,L2'$ and their slope is $K2$.All these lines are tangent to the movement region. From the picture, we know that the domain surrounded by the two sets of parallel lines and the external rectangle could depict the contour of the movement region effectively. The more the amount of tangent lines, the closer to the object's actual contour the border described. In this paper, we use the domain surrounded with n sets of parallel lines, which are tangent to the movement region and have equal interval in slope, and the external rectangle of the movement region to represent the object's calculation domain. The shadow region showed in the picture is the calculation domain of the moving object. The method described above is relatively applicable to the objects with simple shape, such as the ship.
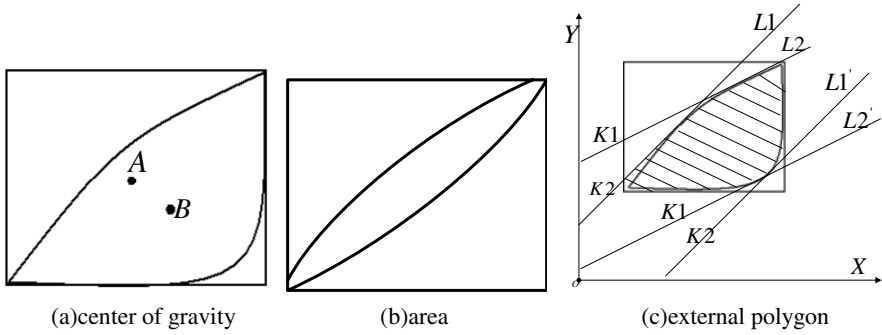
Fig. 2. Calculation domain diagram

3.2   Calculation of Similarity for Pseudo Invariant Line Moments

In the actual scene, there are rotation, translation and scale transformation among different images of the same object. In other words, the differences among different images of the same object involve. Therefore, we could get the invariants that can ascertain an object's shape if we get the pseudo invariant line moments, and these invariants are only relative to the object's shape, not relative to its position, direction and scale. In this way, we can use moment invariants to depict the domain of target image.

Since pseudo invariant line moments remain unchanged when the object rotating, translating or transforming scale, and pseudo invariant line moment can be expressed as the form of vector, by using the model of vector space, we can take pseudo invariant line moment as a point in the vector space and measure the similarity between the moments by calculating the closeness degree of two objects. Since there is intense relativity among pseudo invariant line moments and they have the same weight to the shape description of an object, here we use Euclidean distance to measure the distance between two vectors. It is defined as follows:

$$Dis = \left[\sum_{i=1}^{7} (T_{1,i} - T_{2,i})^2 \right]^{\frac{1}{2}} \tag{17}$$

where $T_{1,i}$ is the i-th pseudo invariant line moment for one of the two matching objects, and $T_{2,i}$ is that of the other one. $T_{1,i}$, $T_{2,i}$ are got by formulae (6)~(12).Therefore, we could calculate the pseudo invariant line moments for the different image regions got by regional segmentation, and use the method above to match the features of the target images with rotation and scale transformation, which takes pseudo invariant line moments as the characteristic quantity. In this way, we can calculate and predict the object's moving situations.

4   Simulation Testing and Analysis

In the experiment, the test system is Visual Studio 6.0, and the machine is configured with Pentium Dual-Core 2.5G CPU,1G SDRAM and 5400RPM IDE hard drive. The operating system running on the computer is Windows2000 Server.

First, we tested the effect of the detection for the moving object's calculation domain. Second, in order to validate the effectiveness of pseudo invariant line moment in detecting moving objects, we have tested on many different moving objects and compared it with several kinds of moments existent. Finally, we summarized the actual application of the algorithm, and the testing results are as follows.

4.1 Domain Detecting for Moving Objects

In Fig. 3 (a) and Fig. 3 (d) are the images needing to detect the moving objects. Fig. 3 (b) and Fig. 3 (e) show the movement domain. After getting the movement domain, mark the objects in it. Then describe the marked objects with their external polygons. Fig. 3 (c) shows the area of the object's external polygon. Just as the picture shows, since the color of some parts of the ship body is much close to the background color, there is a large part of defect. And it has better effect using an external polygon to represent the moving object's calculation domain. The detection result of Fig. 3 (d) is Fig. 3 (e) which shows the movement domain. What is different from Fig. 3 (a) is that there are two moving objects in Fig. 3 (e). After marking the objects and getting the external polygons we got Fig. 3 (f), which shows the domain of the external polygons for the two moving objects. The method proposed in this paper can distinguish a couple of objects in the same scene and accordingly get a couple of calculation areas used to calculate the pseudo invariant line moments.

(a)   (b)   (c)   (d)   (e)   (f)
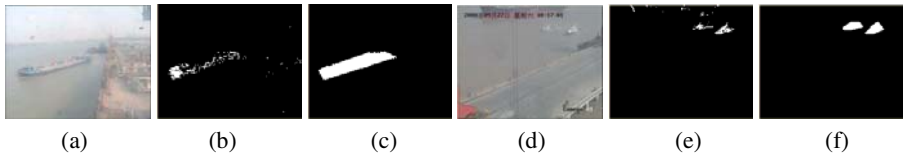
Fig. 3. object extraction experiments

4.2 Performance Testing for Pseudo Invariant Line Moments

In this paper, we have tested on 100 consecutive video images. The data of invariant moments got by dealing with one of the 100 images, which is shown as Fig. 3 (a), is showed in Table 2. The statistics for the computing time is shown in Table 3.

Table 2. Comparison of invariant moment' value

	T_1	T_2	T_3	T_4	T_5	T_6	T_7
Hu's invariant moment	0.4329	3.2134	4.4313	2.9345	8.3245	4.5673	7.3456
Invariant line moment	0.1543	1.3234	2.1233	1.4324	6.3245	3.3453	5.4543
Pseudo invariant line moment	0.2345	2.3221	2.9727	1.3321	5.1231	3.2331	5.3212

	V_1	V_2	V_3	V_4	V_5
Polar radius invariant moment	4.9343	2.8541	3.2384	4.1035	0.2312

From Table 2, we can know that pseudo invariant line moments also have the ability to depict constant values. The numerical range of pseudo invariant line moments meets actual application.

From Table 3, we know that Hu's invariant moments which need to calculate the entire gray image are the most time-consuming. When the processing time of one method is more than one second, it is improper to apply it in a real-time system. The processing time of invariant line moments is less than that of Hu's invariant moments, because invariant line moments only need to deal with the boundary points of a image. The statistic time for measuring the boundary points is about 0.9234 second, so the algorithm spends most of the time in calculating the boundary points. The total processing time also exceeds one second and it is improper to apply it in a real-time system as well. The polar radius invariant moment algorithm spends most of the time on the calculation for polar radius and the total processing time is nearly one second. The pseudo invariant line moment algorithm presented in this paper calculates fewer pixels to extract the features of the edge image, so its processing time is about 0.8 second less than that of the invariant moment calculation for the whole gray image. Though it needs to deal with the edge image by Canny operator, this procedure costs very little time about 0.2431 second. Taking the time into account, the calculation of invariant moments for the edge image is still much faster than that for the whole gray image. Pseudo invariant line moments contain less information than the invariant moments for the whole image, but they help to improve the real-time processing of a recognition system and they are applicable to the target tracking system.

Table 3. The time required for feature extraction

Method	Average runtime/s
Hu's invariant moment	1.6523
Invariant line moment	1.2314
Polar radius invariant moment	0.9837
Pseudo invariant line moment	0.5211

4.3 Practical Application Experiment for Pseudo Invariant Line Moments

Fig. 4 shows a variety of different types of objects detected in a scene. In this paper, we identify the objects in the monitoring system by calculating the Euclidean distance for the pseudo invariant line moments of moving objects and then track them. The object shown in Fig. 4 (a), Fig. 4 (b) and Fig. 4 (c) is the same vessel. However, the objects shown in Fig. 4 (d) and Fig. 4 (f) are different types of moving vessels.

(a)   (b)   (c)   (d)   (e)

Fig. 4. Video surveillance in various objectives

Table 4. The distance of moment between the objectives

Dis	a	b	c	d	e
a	0	0. 0122	0.1932	0.8566	0.6771
b	0.0122	0	0.2156	0.5811	0.7567
c	0.1932	0.2156	0	0.6736	0.7542
d	0.8566	0.5811	0.6736	0	0.5965
e	0.6771	0.7567	0.7542	0.5965	0

Using the target recognition method based on the pseudo invariant line moment algorithm, whether the performance is good or not depends on the difference between the Euclidean distance for the pseudo invariant line moments of the same object and that for the pseudo invariant line moments of different objects.

Table 4 shows all groups of Euclidean distance for the pseudo invariant line moments of the objects shown as the rectangle boxes in Fig. 4. From the Table 4, we can see that the Euclidean distance between the pseudo invariant line moments of the same object (ship) is less than 0.2, while the Euclidean distance between the pseudo invariant line moments of different objects is relatively large, about 0.6, which has significant difference from the Euclidean distance between the pseudo invariant line moments of the same object. The experimental results show that it is feasible using the Euclidean distance between pseudo invariant line moments to identify the objects.

5   Conclusions

By comparing a variety of methods for the feature extraction of the moving objects' target areas in a scene, we introduced a method based on pseudo invariant line moments to express the regional features of moving objects. By this method, we can detect moving objects and predict their movement directions as well. The experiment shows that the detection method based on pseudo invariant line moments could get the domain of a moving object relatively exactly and rapidly. Moreover, it also can predict the object's movement direction effectively. This method has been applied to the system of dock since Sep, 06. By constant test and modification, until now, the system has good effect in detecting all kinds of events at the dock and has accomplished the expected goal on the whole.

Acknowledgements

References

1. Valera, M., Velastin, S.A.: Intelligent Dstributed Srvuillance Systems A Review. IEEE Proceeding of Vision, Image, and Signal Processing 2005 152(2), 192–204 (2005)
2. Frauel, Y., Quesada, O., Bribiesca, E.: Detection of a Polymorphic Mesoamerican Symbol Using a Rule-based Approach. Pattern Recognition 39(7), 1380–1390 (2006)

3. Hu, M.K.: Visual Pattern Recognition by Moment Invariants. IRE Transactions on Information Theory 8(2), 179–187 (1962)
4. Oliveira, R.J., Ribeiro, P.C., et al.: A Video System for Urban Surveillance: Function Integration and Evaluation. In: International Workshop on Image Analysis for Multimedia Interactive Systems (2004)
5. Alirezaee, S., Aghaeinia, H., Ahmadi, M., Faez, K.: Recognition of Middle Age Persian Characters Using a Set of Invariant Moments. In: Proceedings of the 33rd Applied Imagery Pattern Recognition Workshop (2004)
6. Parodies, T., Ali, F.: Computer Recognition of Handwritten Numerals by Polygonal Approximations. IEEE Transactions on Systems, Man, Cyber SMC26, 610–614 (1975)
7. Shamsuddin, M., Sulaiman, M.N., Darus, M.: Improved Scale-invariant Moments for Deformation Digits. International Journal of Computer Mathematics, 439–447 (2000)
8. Analysis and Invariant Moments. WSEAS Transactions on Information Science and Applications 10, 2066–2077 (2006)
9. Lin, H.B., Si, J., Abousleman, G.P.: Orthogonal Rotation-invariant Moments for Digital Image Processing. IEEE Transactions on Image Processing 3, 272–282 (2008)
10. Dhandra, B.V., Malemath, V.S., Mallikarjun, H., Hegadi, R.: Multi-font English Character Recognition Based on Modified Invariant Moments. Journal of Combinatorial Mathematics and Combinatorial Computing 67, 153–162 (2008)
11. Li, X.M., Shi, Z.Y.: Ellipses and Circles Recognition Based on Invariant Moments. Beijing Gongye Daxue Xuebao / Journal of Beijing University of Technology 11, 1136–1140 (2007)

Building-Based Structural Data for Core Functions of Outdoor Scene Analysis

Hoang-Hon Trinh, Dae-Nyeon Kim, Suk-Ju Kang, and Kang-Hyun Jo

Graduate School of Electrical Engineering, University of Ulsan, Korea
San 29, Mugeo-Dong, Nam-Ku, Ulsan 680 - 749, Korea
{hhtrinh,dnkim2005,sjkang,jkh2009}@islab.ulsan.ac.kr

Abstract. The most important things to realize such an intelligent system are core functions such as landmark detection, recognition and reconstruction. Since where we have core functions, the intelligent system can propagate other procedures like navigation, mapping, localization, etc. Thus, this paper describes an approach to construct a structural data for core functions by using geometrical structure of building. Firstly, line segments are detected. Then several processes such as rejecting noises, calculating dominant vanishing points, filtering the edges of building are used to detect the building surfaces. The criteria are created for decision of building detection function. Secondly, for each surface, a generative model including area, wall histogram and a list of local features are computed for the recognition function. Finally, the geometrical features as windows, doors, floors or rooms are estimated for reconstructing the building. The proposed method has been performed with large databases and sound results of all functions.

Keywords: Landmark detection, building recognition, core function.

1 Introduction

In order to create an intelligent system in the outdoor scene, the designer must carefully consider several functions such as landmark detection, recognition, environment reconstruction, localization and mapping. So it is better if a function can be inherited and helped from the others. In this paper, we base on natural structure of object to build a database with expectation that it can be shared for every function. The building is chosen as an important landmark in urban environment. The proposed method is attempted to apply to detection and recognition functions. Furthermore, the geometrical properties such as doors, windows, columns and wall of building are analyzed. This information can help the system understand environment more by estimated what the size of building is. The information also can be used for reconstruction, localization and mapping functions.

The above functions were usually solved in separateness by several literatures [2,3,4,5,7,8,9,12]. The detection function is approached by [3,6]. Most of methods concerned a dominance of appearance of building like facades, salient or single

building in images. Garcin *et al* [3] detected building by Markov object process for aerial photography and the roof is represented the buildings. Madhavan *et al* [6] used the ground surface as prior knowledge to detect the building with LADAR data. The recognition function is solved with assumption that the image always contains a single building inside [2,8,12]. The works in [2,12] used SIFT descriptors [5] to describe the building. They flexibly chose the keypoints and their descriptors which strongly located and described the objects. Several literatures tried to analyze the geometrical information, for example detecting windows and doors, with hopeful that it makes the available information for reconstruction function and more understanding environment of robot [4,7]. S. Pu *et al* [7] extracted window from terrestrial laser scanning data so that the user easily find building features such as wall, roof, door, etc. The windows are extracted by recovering these holes of wall plane. S.C. Lee *et al* [4] combined information of line segments and calibrated facade to detect and reconstruct the windows in 3D coordinate. The results were sound with centimeter error, but some cases also needs user assistance (semi-automatic).

2   Line Segments for Detecting Building Surface

2.1   Line Segment Processing

Firstly, we detected line segments [10] with assumption that the building is upright positioned. Then, along each segment, two sets ($\Omega_{1,2}$) of pixels are calculated by four sampled segments as in Fig.1(a). A model of line segment (MLS) is constructed by two parameters: difference of average intensities ($\triangle\mathbf{I}$) and minimum of variant of RGB color (σ_m),

$$\triangle\mathbf{I} = \frac{1}{3n} \sum_{R,G,B} \sum_{i=1}^{n} |\mathbf{I}^i_{\Omega_1} - \mathbf{I}^i_{\Omega_2}| \tag{1}$$

$$\sigma_{\Omega_{1,2}} = \frac{1}{3} \sum_{R,G,B} \mathbf{Var}(\mathbf{I}_{\Omega_{1,2}}) \Rightarrow \sigma_m = \min(\sigma_{\Omega_{1,2}}) \tag{2}$$

where n is number pixels of $\Omega_{1,2}$. Thousands of line segments in ZuBud database [9] located on building and non-building patterns, Fig.1(b), are selected for calculating $\triangle\mathbf{I}$ and σ_m. Here, 90.16(%) non-building segments are ruled out while 8.7 (%) building segments are refused if $\triangle\mathbf{I} \geq 20$ and $\sigma_m \leq 20$, Fig.1(c). Thus, these values are chosen as thresholds for refining the noise, Figs.1(d,e).

Dominant vanishing points (DVP) are calculated by MSAC (m- estimator sample consensus, [1]) method with one verticality and maximum five of horizontal DVPs; it is adaptable for multiple faces of building. The priority of horizontal DVPs is arranged from high to low by their number of segments. To enhance the junctions, the vertical segments are extended across the image while the horizontal segments are extended on both of the ends with a delta length (5 pixels). For rejecting sparse segments, a surrounding region comprising a rectangle and two semicircles is created where relation of their sizes is shown in Figs.2(a,b).

(a) $d_{max} = 4, d_{min} = 3$ pixels

(b) Sample lines

(c) Distribution of $\triangle\mathbf{I}$ and σ_m of 200 first sampled segments

(d) Example of non-building image

(e) Example of building image
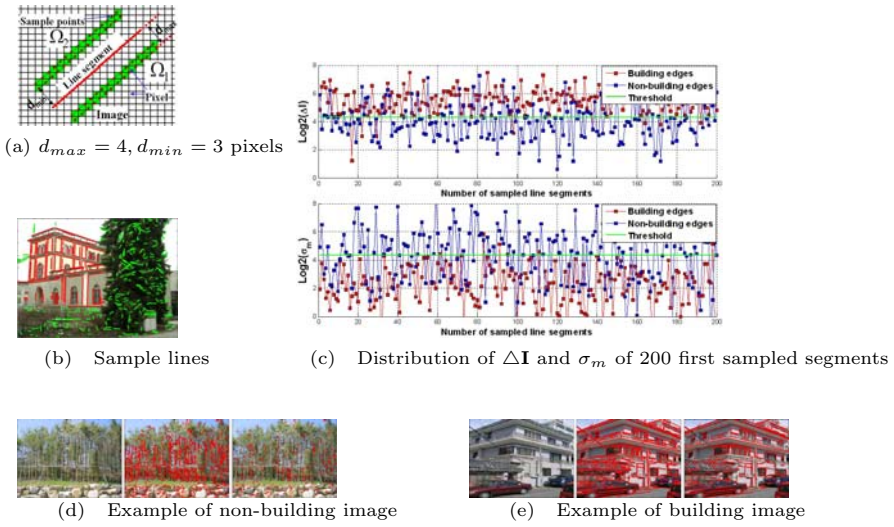
Fig. 1. Illustration of MLS: (d,e) from left to right are the original images, before and after rejecting noise, respectively

A segment whose surrounding region contains less than two ends of other segments is pruned out. Finally, a few of horizontal segments which lie very near to the vanishing line is often misgrouped. Those segments instead of belonging to lower priority, they will go into the higher priority so that they need to be recovered. The recovery process is performed from the lower to higher priority group.

2.2 Empirical Assumptions and Definitions

The vertical lines (VL) are numbered from left to right. If the ith VL belongs a facet, it will intersect horizontal segments with evident y-coordinates and the height, h_i, is defined by $h_i = y_{max} - y_{min}$. Area, $\mathbf{A}$, of candidate facet is estimated by $\mathbf{A} = \alpha\beta\mathbf{L}h_i$, where $\mathbf{L}$ is image width; α is a probability of spread of candidate facet over the image; β is a scalar because of different viewpoint. With 640x480 image size, L equals 480 (or 640). The parameters are empirically chosen that α, β and $\mathbf{A}$ equal 0.5, 1 and 10,000 pixels, respectively; $h_i \geq h_0$ (40 pixels).

Definition 1: Candidate line of a horizontal group is a vertical line whose intersectional number satisfied that $n \geq N_1$ or $n \geq N_2$ and $h_i \geq h_0$, table 1. n and h_i are illustrated by Figs.2(c-e). Figs.2(f,g) show two candidate line groups.

Definition 2: Face pattern is a cluster of continuous candidate lines belonging the same horizontal groups and its width $w \geq w_0$ (30 pixels), Fig.2(h).

Definition 3: Non-face pattern is a space between any two face patterns if its width is larger than w_0.

2.3   Surface Detection

A partial face (PF) is created by concatenating one or more patterns if they have no non-face pattern inside. If two or more horizontal groups are mix-up in one PF then only one group with largest segments is chosen. If more than two horizontal groups distribute along vertical direction then only two PFs with largest segments are kept. The top and bottom boundaries of PF are found out by searching from top down and bottom up. The horizontal boundary is a segment which satisfies one of conditions as enumerating as table 2. Each horizontal segment is replaced by a horizontal line which lies through the middle point and corresponding DVP. N is the number of horizontal co-existent segments on the boundary including itself. α (degrees) is angle of horizontal line and the segment. MD (degrees) is a mean deviation of N co-existent segments. d is ratio of segment length and PF width. N_v is the intersectional number between the horizontal line and vertical segments. Fig.2(i) shows that four partial faces are

(a)   (b)   (c)   (d)   (e)

(f)   (g)   (h)   (i)   (j)
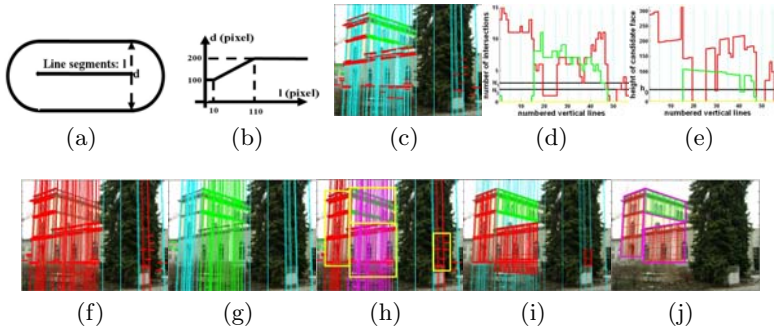
Fig. 2. (a) surrounding region; (b) relation between diameter and length of line segment; (c) line segment processing; (d,e) distribution of n and h_i; (f,g) face patterns of each horizontal group; (h,i) total face patterns and partial faces; (j) surface detection

Table 1. The values of threshold N_1 and N_2

Group orders	$1^{st}, 2^{nd}$	3rd	$4^{th}, 5^{th}$
N_1	3	4	5
N_2	2	3	4

Table 2. The conditions (Cds.) of horizontal boundaries

Cds.	N	α^o	MD^o	d	N_v
1	1	≤ 0.75	-	≥ 0.85	-
2	1	≤ 1	-	≥ 1	-
3	≥ 2	≤ 0.75	≤ 1.5	-	-
4	-	-	-	-	≥ 4

automatically detected. An ambiguous PF will be rejected if it satisfies one of conditions on table 3; where M_v and M_h are the mean lengths of vertical and horizontal boundaries. **A** is the area of PF. Finally, if two PFs are close together with different DVPs then a common vertical boundary is created by maximum number of intersections between horizontal segments of the component PFs (more detail in [10]). If two or more PFs are close together with the same DVP and approximate height then they will be merged into one surface, Fig.2(j).

3 Structural Data Analysis

3.1 Structural Data for Building Detection

It is difficult to clearly discriminate between building and non-building regions in the images. It is affected by the appearance of building and the conditions when the image is being taken. Our goal is to construct a data which can be shared by as much functions as possible. The images are taken from ground views and their appearances are not so far from the camera position. On a certain limitation, the estimation for far or close building appearances is controlled by the threshold for detecting line segment. The hierarchical conditions for detecting building are summarized that there is at least existence of one vertical DVP, one horizontal DVP, and the final area of facet pattern has larger than 10,000 pixels.

3.2 Structural Data for Building Recognition

Each surface is represented by three features such as area, wall color histogram and a list of local descriptors. The wall color histogram is well described in our previous works [11]. The color information of wall is used to calculate a 36-bin histogram. The other visual property of facet is a list of SIFT (scale invariant feature transform, [5]) descriptors. To decrease the noise and increase the effect of match, the surface is transformed into the rectangular shape before calculating SIFT descriptors and only keypoint whose scale is larger than 2 is chosen, Fig.3.

For each building, tens of images are taken under different conditions. The corresponding surfaces are classified by user. Then a generative model (GM) is constructed by using singular value decomposition (SVD) method. Given $n \times 2$ matrix A, we use SVD algorithm to decompose that $A = U \sum V^T$, where $\sum = diag(\lambda_1, \lambda_2)$. Let a_1, a_2 be the columns of A, if $\chi^2_{(a_1,a_2)} \simeq 0$ then $\lambda_1 \gg \lambda_2$ and $\lambda_2 \simeq 0$. Re-calculate A that $A = U \sum' V^T$ where $\sum' = diag(\lambda_1, 0)$. Now, new

Table 3. The conditions (Cds.) of horizontal boundaries

Cds.	1	2	3	4	5	6 (and if $A \leq 8000$ pixels)
M_v (pixels)	< 30	-	< 45	< 120	< 60	< 200
M_h (pixels)	-	< 30	< 120	< 45	< 60	< 200

vector a_1 equals a_2 after normalizing into unit length. Let $a = a_1$, a is called approximate vector. The MG is trained by supervision as follows,

1. *Initial model*: Randomly chose a image in training set.
2. *The first update*: A new facet (F_N) is directly matched to the GM. The correspondences are verified and then used for updating GM by the approximate vector. The F_N is stored at another place as an auxiliary facet (F_A).
3. *The next update*: GM is updated by F_N. Then F_N matches to F_A, the correspondences are added into GM as new features by 2D transformative matrices. The F_A is replaced by F_N. If the number of features of GM increases then some features whose the updated times are smallest will be ruled out.

To overcome the affection of random noise we decompose matrix A with a control factor; $A = [\gamma a_1, a_2]$, where $\gamma(= 2)$ is a control factor. Similarly, the color histogram is updated by the approximate vector.

3.3   Geometrical Surface Analysis

For each surface, a binary image is constructed where the non-wall regions are considered as candidates of windows, doors, columns, etc; the surface is transformed into the rectangular shape, Figs.4(a-d). The ambiguous candidates with irregular shapes, for example, long and thin or very small are coarsely rejected. The geometrical characteristics of each candidate are represented by its coordinates $(x_{max,min}, y_{max,min})$; centroid (x_c, y_c); height $(h = x_{max} - x_{min})$ and width $(w = y_{max} - y_{min})$; bounding box area $(A = wh)$ and aspect ratio, $A_{asp} = \frac{h}{w}$. Let W be a set of windows where R^i, $i = 1, 2, \cdots, M$, $(C^j, j = 1, 2, \cdots, N)$ are the rows (columns), Figs.4(e,f). The windows of building are usually coexisted to each other. Thus, two candidates (a,b) should be satisfied,

$$a \in W \Rightarrow A_{asp}^a \leq \tau_{asp} \tag{3}$$

(a) The first eight of images of one building in our database

(b) Two corresponding surfaces and their wall regions of building

(c) Generative model of surface according to the red surface of Fig.3(b)

Fig. 3. Structural data for an example of one building
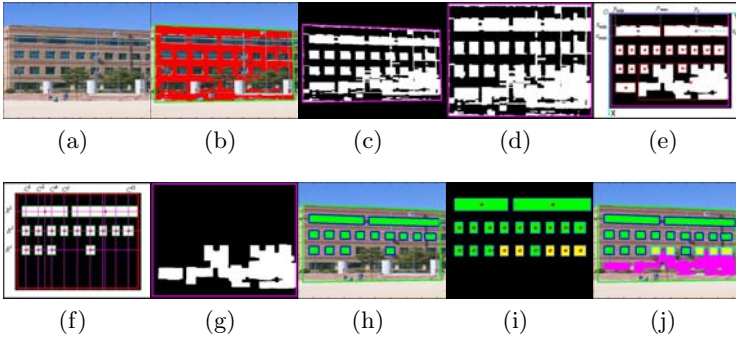
Fig. 4. Step-by-step for detecting window

$$a, b \in R^i (\text{or } C^j) \Rightarrow \frac{|A^a - A^b|}{A^a + A^b} \leq \tau_A \tag{4}$$

$$a, b \in R^i \Rightarrow \frac{|x^a_{max(min,c)} - x^b_{max(min,c)}|}{h^a + h^b} \leq \tau_c(\tau_c, \frac{\tau_c}{2}) \tag{5}$$

$$a, b \in C^j \Rightarrow \frac{|y^a_{max(min,c)} - y^b_{max(min,c)}|}{w^a + w^b} \leq \tau_c(\tau_c, \frac{\tau_c}{2}) \tag{6}$$

where τ_{asp}, τ_A, τ_c are certain thresholds according to 3.5, 0.25, 0.1, respectively. For two or more windows of a row (column), the boundaries are re-calculated by iteratively reweighted least squares of their boundary pixels, Fig.4(h).

The false positive (FP) and negative (FN) are improved by aligned characteristics of windows. From detected windows, the median of area and centroid are calculated. The FP is appeared as small windows at bottom of surface so that it will be pruned out. The FN is appeared when a real window is occluded. Each position of rectangular region where three other windows are located should be a FN. It will be recovered after checking the corresponding position of remained image in Fig.4(g). The final results are shown in Figs.4(i,j). The remained image is considered as the candidates of doors as magenta region in Fig.4(j). Using the alignment of detected windows, the size of building such as the number of floors or the number of rooms (windows) of each floor are easily estimated. For example, the building in Fig.4 has three floors with maximum 9 rooms in each floors. Of course, It will be mis-estimated when the building face is occluded. But this error can be improved when we analyze the image sequence of the movement.

4   Experiments

The proposed method has been performed with three data sets, all images are taken in Ulsan metropolitan city in South Korea.

For building detection, a data set comprising 880 images with 680 building images and 200 non-building ones is used. The building is correctly detected

(a) Results of building de- (b) Results of non-building
tection   detection
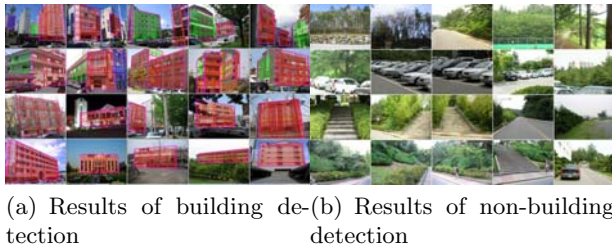
Fig. 5. Examples of building detection function

(a) 30, 52 and 68 correct (b) 42, 72 and 85 correct
matches   matches

(c) 63, 104 and 149 cor- (d) 45, 78 and 100 correct
rect matches   matches
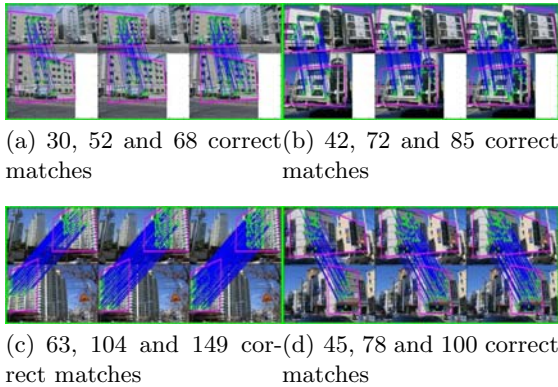
Fig. 6. From left to right, The results of without, 10 and 20 times of update,
respectively

(a)   (b)   (c)   (d)   (e)   (f)
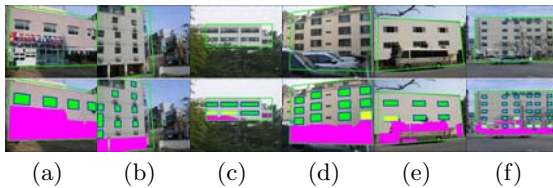
Fig. 7. More examples of results of window detection

when the processes given out a surface which locates on the real one. The re-
sults obtain 660 (184) correct cases for building (non-building) detection. Several
examples are shown in Fig. 5. Here, we obtain 97% and 92% of detection rate of
building and non-building images, respectively.

For building recognition, 50 buildings with their neighbors are considered.
Each building is taken 22 images including one for initial model, 20 images
for training database and remained one for testing the process. 78 surfaces are
detected from 50 images of initial models. Average 369 keypoints (or descriptors)

are found for each each building. The size of database is very smaller than the such works, [12], where around 2500 keypoints are stored for each building. The recognition process has performed with initial database, after 10 and 20 times of updates. The results show that the largest number of matches (before verifying the correspondences) is increase about 10% of each 10 times of update. But the number of correspondences increases about 50% after 10 times of update. Fig.6 shows several examples of our results. In each sub-image, the above image is the test and the bottom one is the initial image after training. We obtain 100% recognition rate for the observing test images with updated database.

For geometrical analysis, 150 evidently appeared surfaces are challenged with total 1607 windows (T). The process given out 1544 extracted windows including 1521 correct cases (TP- true positive), 139 recovered false negative windows, 23 false positive (FP) window extraction. So that 84 windows are omissive extraction (FN-false negative). Fig.7 shows some results where the top picture in each sub-image is the original image with the boundary of detected surface; the under one is result of window detection. We obtained 93.34% detection rate.

5   Conclusions and Future Works

The structural data for core functions has constructed by geometrical characteristics of building. The MLS proves that it is very powerful to distinguish building and non-building patterns by remarkably reducing the line segments of non-building images. So that the building and non-building images are classified. The GM of surface is very effect for the building recognition which automatically verifies the informative descriptors. When the system such as the robot is close the building he needs to more understand the environment. Thus, the window of building is detected which can be used to estimate other information such as the doors, floors, rooms, etc. The structural data has been applied with three large database. We obtain the good performances with high rate of building detection, recognition and window extraction.

We are going to detect the door of building and then apply the method for designing the outdoor robot. Different from the windows, the door usually appears alone in the building image. Therefore, to detect the door, we need more information such as the context and multiple cues from the neighbor regions. For example, the bush or ornament and trained trees appears beside the door; the location is at bottom of building surface, etc.

Acknowledgements

References

1. Fischler, M.A., Bolles, R.C.: Random sample consensus: a paradigm for model fitting with application to image analysis and automated cartography. Communications of the ACM 24(6), 381–395 (1981)
2. Fritz, G., Seifert, C., Paletta, L.: A Mobile Vision System for Urban Detection with Informative Local Descriptors. In: ICVS, vol. 30 (2006)
3. Garcin, L., Descombes, X., Men, H.L., Zerubia, J.: Building Detection by Markov Object Processes. In: IEEE int'l Conf., vol. 2, pp. 565–568 (2001)
4. Lee, S.C., Jung, S.K., Nevatia, R.: Automatic Integration of Facade Textures into 3D Building Models with a Projective Geometry Based Line Clustering. Computer Graphics Forum 21, 511–519 (2002)
5. Lowe, D.G.: Distinctive Image Features from Scale-Invariant Keypoints. IJCV 60(2), 91–110 (2004)
6. Madhavan, R., Hong, T.: Robust Detection and Recognition of Buildings in Urban Environments from LADAR Data. AIPR 39–44 (2004)
7. Pu, S., Vosselman, G.: Extracting windows from terrestrial laser scanning. In: Int'l Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences, 36, part 3/W52, Espoo, Finland, September 12-14, pp. 320–325 (2007)
8. Robertson, D., Cipolla, R.: An Image-based System for Urban Navigation. BMVC (2004)
9. Shao, H., Gool, L.V.: Zubud-zurich Buildings Database for Image based Recognition, Swiss FI of Tech., Tech. report. 260 (2003)
10. Trinh, H.H., Jo, K.H.: Image-based Structural Analysis of Building using Line Segments and their Geometrical Vanishing Points. In: SICE-ICASE, October 18–21 (2006)
11. Trinh, H.H., Kim, D.N., Jo, K.H.: Cross Ratio-Based Refinement of Local Features for Building Recognition. Journal of LNCS (2008)
12. Zhang, W., Kosecka, J.: Hierarchical Building Recognition. IVC 25, 704–716 (2007)

Appearance Feature Based Human Correspondence under Non-overlapping Views

Hyun-Uk Chae and Kang-Hyun Jo

School of Electrical Engineering, University of Ulsan
Bldg. 7-413, Daehak-ro 102, Nam-gu, Ulsan, 680-749, Korea
{hwchae,jkh2009}@islab.ulsan.ac.kr

Abstract. In this paper, a method is proposed, to solve correspondence problem under structured space which is installed multiple cameras. The correspondence between different cameras is an important task to use the multiple camera system. For solving this problem, the proposed method is consists of three steps which are detection of moving object, feature extraction and correspondence among different cameras. First step is to detect moving people by background subtraction from multiple background model. The temporal difference is used jointly to remove noise occurred from temporary change. The detected regions are divided using labeling as individual person. The second step is to segment the each person by a criterion with appearance and context information. The segmented regions in a person are estimated as Gaussian mixture model (GMM) for correspondence. The final step is process of correspondence between different cameras. A GMM from a camera is matched with another GMM from other cameras. A ratio of those GMMs is used as a criteria to identify same person. The experiment was performed with the specific scenarios in quantitative results.

Keywords: Multiple Camera System, Background Subtraction, Correspondence, Gaussian mixture model.

1   Introduction

The automatic monitoring systems using vision sensor improve the necessity itself in the structured space. Those systems are investigated continuously in the field of computer vision such as intelligent space, robot town, and automatic surveillance system [1-3]. The multiple cameras are generally used to overcome limitation of field of view (FOV) when a single camera is used. There are several problems to use the multiple camera systems, i.e. camera calibration, camera switching, synchronization and correspondence among multiple views, and so on. Among those problems, this paper proposes a method to obtain correspondence between different views.

The multiple camera systems can be classified as two situations are overlapping and non-overlapping views. The related works for corresponding objects in multiple camera system are investigated continuously with several approaches. In the geometrical approach, Hu et al. [4] use the principal axis of objects as feature. The correspondence is obtained by finding intersection of axes extracted from each object in

different views using homography. In another approach, the global color model from color histogram is used to match objects under different views. The correspondence is obtained by measuring the sum of distance between global color model and local color model [5]. Porikli [6] proposes inter-camera transfer function from cross-correlation matrix. The color calibration is executed in advance to calibrate images from non-overlapping cameras which have different lighting conditions. Javed et al. [7] also propose the method to identify same person in non-overlapping views by using color transfer function and relationships of inter-camera. However, those methods usually need a training process in advance such as generating color model, color calibration, and learning a relationship of cameras. Also, if the systems need to install additional cameras, they have to execute re-training for whole cameras. To avoid that problem, we propose a method to identify a same person in the non-overlapping cameras without a learning process in advance.
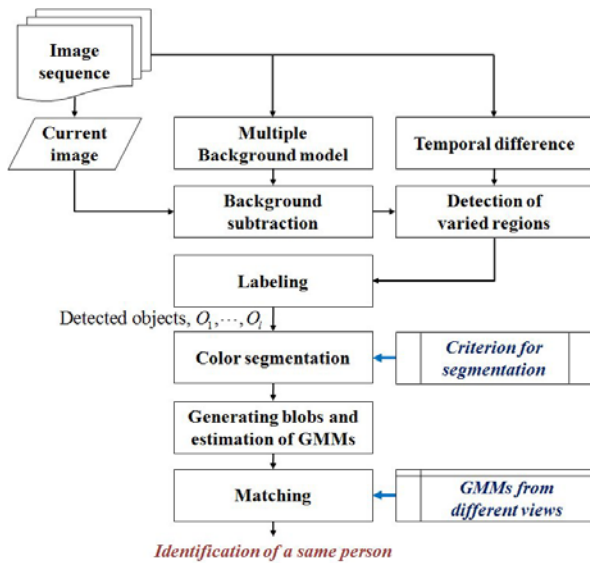
Fig. 1. The process diagram for proposed method

The proposed method is consists of three steps which are motion detection, feature extraction and correspondence as shown in Fig. 1. The first step is a detection of moving objects. The objects are detected by background subtraction from multiple background model. The temporal difference of consecutive images is jointly used to remove noise from temporary change. Then the detected objects are divided by labeling as individual person. In this paper, we make assumptions that moving object is only human and human is walking straightly. The second step is to extract features from detected person in each camera. A person is segmented as blobs by using a criterion with color range and context information of human body. Each blob is generated as Gaussian probability distribution in each color channel. Then, the set of blobs is represented as Gaussian mixture model (GMM). In the final step, a GMM from a camera is matched with another GMM from other cameras by maximum likelihood estimation.

This paper consists of five sections as follows: The method for detection of moving objects is introduced in section 2. The process of generating blobs and estimating probability model is described in section 3. The matching and experiment is presented in section 4 and 5, respectively. Finally, the conclusion is presented in section 6.

2   Detection of Moving Objects

The detection of moving object is to segment foreground which is interesting regions from background scene. The several methods are proposed to detect moving objects such as background subtraction, optical flow, statistical method, and so on. Among them, we use background subtraction which is more simple and efficient method in a fixed camera system. In here, generating background model is important task for accurate detection. The background model is generated as two types which are multiple background model and temporal difference. The background subtraction is executed between multiple background model and current image. Then temporal difference is jointly used to remove noise from temporary change in the result of background subtraction. When more than two objects are detected, labeling is executed to select an individual person.

Table 1. Notations for the process of generating multiple background model

I_t	Input Image sequence, $t = 1, 2, 3,…, N_t$
p	Index of pixel intensity, $p = 1, 2, 3,…,$ (width×height)
C_i	Mean value of ith cluster
M_i	The indexed number in i^{th} cluster
W_i	Weight value of ith cluster
B_i	i^{th} background model, i = 1, 2, 3,…, $n(p)$
t	Index of frame
I	Index of cluster number
δ_{MBM}	Threshold value for separation of each pixel among image sequence
σ	Threshold value for removing clusters with small weight

2.1   Generating Background Models

The background model is generated as two types which are multiple background model and temporal difference. Firstly, the process of generating multiple background model is described with three steps which are pixel-based online clustering, removing clusters by comparing weight value and generating multiple background model as shown in Fig. 2 [8-9]. For explaining this process, the denotation is given in Table 1.

The input sequence is converted RGB color space into gray image in advance. Whole process is executed based on pixel intensity. Selected N frames among total number of sequence N_t are used to generate multiple background model ($N = 100$). Each pixel has several clusters ($C_1,…,C_i$) and i^{th} cluster has mean value M_i and weight

value W_i. The first step is online clustering based on pixel. Every pixel is classified as a cluster by comparing distance with threshold value δ_{MBM}. If the distance is smaller than δ_{MBM}, the pixel is classified as the same cluster. Otherwise, the pixel is updated as new cluster. The process of clustering is executed until no more patterns remained. The second step is to remove clusters by comparing threshold value σ and weight value of each cluster. The weight value of cluster is calculated from the ratio between the number of referred pixels in a cluster and total number of pixel. The final step is to generate multiple background model. Among remained clusters from the second step, the background models are generated. In here, we set the number of background models that is four. If the multiple background model is generated at once, we just need to calculate new frame over time. Therefore, it is more simple and useful than other approach, i.e., mixture of Gaussian, temporal difference, and so on.
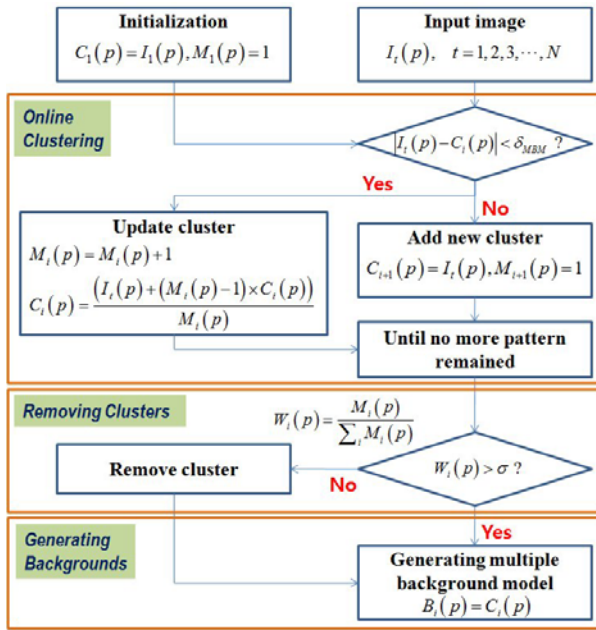
Fig. 2. The process for generating multiple background model: It is composed of three steps which are pixel based online clustering, removing clusters and generating backgrounds

Secondly, the temporal difference of consecutive two images is obtained by (1) and it updates itself in each frame.

$$TD_N = \begin{cases} 0, & |I_N - I_{N-1}| \ge \delta_{TD} \\ 255, & \text{otherwise} \end{cases} \tag{1}$$

where TD_N and I_N are a temporal difference and input image in N frame, respectively. δ_{TD} is threshold value to select motion pixels or regions.

2.2 Background Subtraction and Labeling

Firstly, the moving objects are detected by background subtraction from multiple background model. If the subtraction between i^{th} background $MBM_{N,i}$ and input image including person is larger than δ_{BS}, those pixels are segmented as candidate of moving person by (2). The intersection regions of four results from background subtraction are regarded as moving person BS_N.

$$BS_{N,i} = \begin{cases} 0, & \left| MBM_{N,i} - I_N \right| \geq \delta_{BS}, \quad i = 1, \cdots, 4 \\ 255, & \text{otherwise} \end{cases} \tag{2}$$

The result of temporal difference is jointly used to compensate a limitation of multiple background model, i.e., the system recognize objects as background when the objects are staying in some interval which is smaller than N frames. In the section 2, there are several threshold values δ_{MBM}, σ, δ_{TD} and δ_{BS}. In this paper, those values are selected empirically.
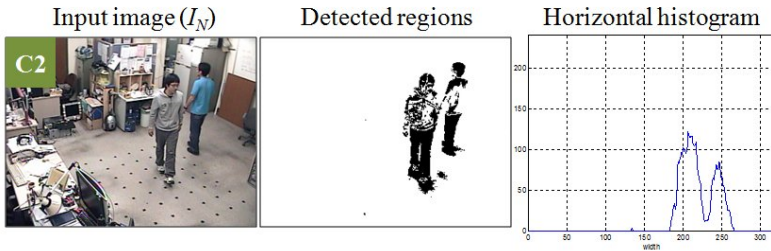
Fig. 3. The example of occlusion

When more than two people are detected, the labeling is used to select each individual person. The detail process of background subtraction and labeling is described in Fig. 3. In the case of occlusion, the number of label in each camera is compared with other cameras'. The labeled regions with smallest label are divided by its histogram of horizontal and vertical. The example of occlusion is shown in Fig. 4. The number of label in second column (C2) is smaller than others. The regions are divided to make same number of label in others. The detail segmentation process of occluded people is explained in [4]. From above process, we detect moving objects in each camera.

3   Feature Extraction and Correspondence

In this section, the detected person is segmented as feature which is a Gaussian mixture model (GMM). Then, the GMM of a person in a camera is compared with another GMM of person from different views to identify a same human.

The RGB color space of detected region is converted into HSI (hue-saturation-intensity) color space. A criterion is generated with color range and context information of human body. After segmentation by the criterion, remain regions are

segmented and grouped using color labeling. The segmented region is represented as Gaussian probability distribution. The person is composed of a set of blobs; therefore they are generated as a model by estimation of Gaussian mixture model (GMM). Each GMM is used as feature to identify a same human.

3.1   Color Segmentation

The detected regions in RGB color space is converted into HSI color space. The range in each channel is normalized in 0~255. The criterion is generated with color range and context information. The height of detected region is used as context information. The criterion is given in Table 2 [10]. The person regions are divided by the criterion. Then, color labeling is used for segmentation of remain regions. Those are segmented and grouped as groups.

Table 2. The criterion for color segmentation

Group	Color range	Context Information
Hair	0 < Intensity < 45	Location > 1/4 of height
Face		
Skin in arm or hand	0 < Hue < 30 or 240 < Hue < 255	1/4 of height ≤ Location < 3/4 of height
Skin in leg		Location ≤ 3/4 of height

3.2   Estimation of Gaussian Mixture Model

The segmented regions are represented as n^{th} blob, $O_{l,1},\cdots,O_{l,n}$. The blobs have Gaussian probability distribution which is given in (3).

$$p\left(O_{l,n}\right) = A_b \cdot \exp\left(-\frac{\left(c_{i,j}-\mu_c\right)^2}{2\sigma_c^2}\right)$$

(3)

where $p(O)$ is the probability distribution of object and A is sum of each blob. μ and σ are the mean and variance according as color value c of the blobs coordinates, respectively. The person has a set of blobs and is represented as a model by estimating Gaussian mixture model in (4) from distributions of blobs.

$$p\left(c \mid \Theta\right) = \sum_{j=1}^{b} p\left(c \mid \omega_j, \theta_j\right) p\left(\omega_j\right)$$

(4)

where c is a vector which has color values as components and $\Theta = \{\theta_1,\cdots,\theta_b\}$ are unknown parameters for each Gaussian distribution of blobs $p(c|\omega_j, \theta_j)$, $j = 1,\ldots,b$. ω_j is the j-th labeled blob among known the number of blobs. $\theta = \{\mu, \sigma^2\}$ is statistical values of the mean (μ) and variance (σ^2) of the j-th blob. It is able to estimation statistical values as GMM using an expectation-maximization (EM) [11].

3.3   Correspondence between Different Views

A GMM in a camera is matched with other GMM from different views to identify a same human by maximum likelihood. The ratio of GMMs is used as a criterion to decide correspondence.

From the results of previous sections, an object has three GMM in each channel of HSI color space. The object is represented as a model in (5).

$$M_l^{(k)} = \log \left\{ \alpha_H p_{l,H}^{(k)} + \alpha_S p_{l,S}^{(k)} + \alpha_I p_{l,I}^{(k)} \right\} \tag{5}$$

where M is a feature model of each object l in the camera k. p and α are GMM and weight value in each channel, respectively. The weight values are as follows: $\alpha_H=0.4$, $\alpha_S=0.4$, $\alpha_I=0.2$. The intensity value is frequently changed from illumination variance. Therefore we set the different weight value in each channel.

The generated models are stored during 10 frames to confirm correspondence when the objects are reappeared in different space.

4   Experiments

The system configuration for experiment is as shown in Fig. 4. The experimental spaces are laboratory (S_{lab}) and corridor (S_{cor}) which is connected with laboratory. The five cameras (C1~C5) and two cameras (C-A1, C-A2) are installed in the S_{lab} and S_{cor}, respectively.

All set of cameras are used with same-typed CCD cameras. The frame ratio is 15 fps (frame per seconds) and 1300 frames are taken by each camera. The system is experimented under off-line.
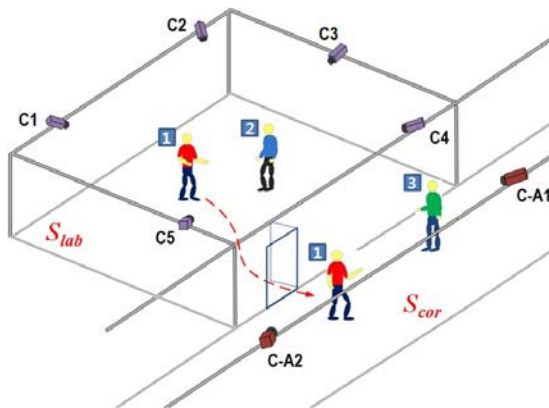
Fig. 4. The system configuration for experiment. The five cameras (C1~C5) and additional two cameras (C-A1, C-A2) are installed in the laboratory (S_{lab}) and the corridor (S_{cor}), respectively.
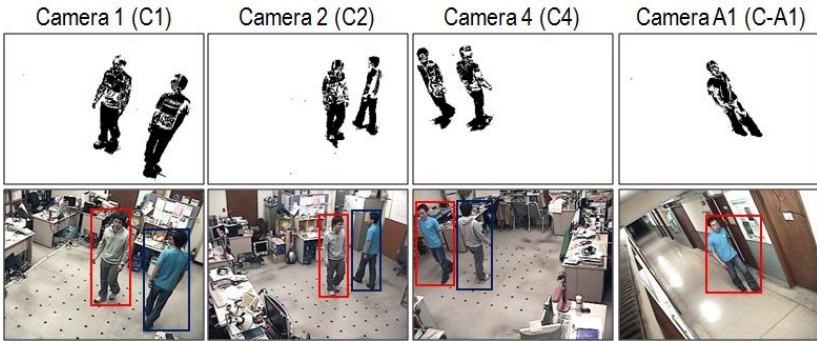
Fig. 5. The result of motion detection
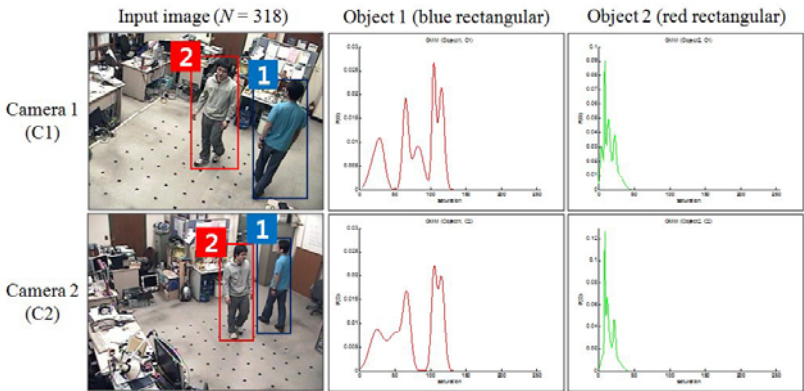
Fig. 6. Comparison between GMMs of objects from different views in same spaces (S_{lab})

The detected result is shown in Fig. 5. The first row is the results of background subtraction from multiple background model and the second row is the results of labeled person in each camera. The experiment for correspondence is executed with two scenarios as follows:

- **Case 1:** The correspondence of objects among the different cameras in the laboratory space (S_{lab}). We need to identify a person in the overlapping views obtained from multiple cameras
- **Case 2:** The correspondence of objects among the different cameras in the laboratory (S_{lab}) and the corridor space (S_{cor}), respectively. The person in a space (S_{lab}) of overlapping views move to another space (S_{cor}) which is connected with (S_{lab}). This correspondence is a necessity when we need to track continuously even though the person is stayed in the non-overlapping views.

Fig. 6 and Fig. 7 show the result of comparison between GMM of objects under above scenarios, respectively. Firstly, the GMM of each person is compared in the laboratory as shown in Fig. 6. From comparing each distribution, we are able to identify that GMM of each person is obviously different. Secondly, the GMM of each person in

(a) Image of detected person

(b) Comparison of GMM (saturation)
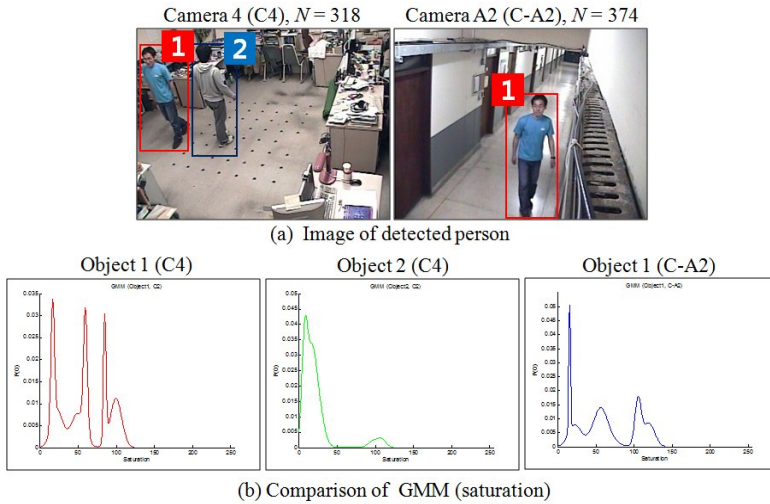
Fig. 7. Comparison between GMMs of objects from different views in different spaces

the laboratory and corridor is described as shown in Fig. 7. In the second case, the distributions of same object are more different than the first case from illumination difference. However, the same person is identified by selecting minimum of the ratio of GMM. Therefore the correspondence is also obtained in case 2. The ratios of correspondence are 94.88% and 86.65% in the case 1 and case 2, respectively. The 500 images among accurate detected images in each camera are used to obtain the ratio of correspondence.

5   Conclusions

This paper proposes a kind of method to solve correspondence problem in the different views when the system need to additional space. The moving objects are detected by background subtraction from multiple background model and temporal difference. The color distribution of detected objects is used as feature for obtaining the correspondence. For extracting the color features, we make the criterion with color range and context information. Each object is represented as estimated GMM. The GMM of a detected person from a camera is matched with another GMM of another people from the different cameras by maximum likelihood estimation. That is, the ratio of GMM is a criterion to identify same person. The experiment with several sequences and three scenarios was performed in the quantitative results. We are able to obtain correspondence of objects in different views without a learning process in advance.

References

1. Murakami, K., Hasegawa, T., Kurazume, R., Kimuro, Y.: A Structured Environment with Sensor Networks for Intelligent Robots. In: IEEE Conference on Sensors, pp. 705–708 (2008)
2. Hashimoto, H.: "Intelligent Space: a Commentary on Research Trends. In: International Symposium on Industrial Electronics, vol. 1, pp. 11–16 (2002)
3. Collins, R.T., Lipton, T., Kanade, T., Fujiyoshi, H., Duggens, D., Tsin, Y., Tolliver, D., Enomoto, N., Hasegawa, O., Burt, P., Wixson, L.: A System for Video Surveillance and Monitoring, Technical Report, Carnegie Mellon University, Pittsburgh, PA, CMU-RI-TR-00-12 (2000)
4. Hu, W., Hu, M., Zhou, X., Tan, T., Lou, J., Maybank, S.: Principal Axis-Based Correspondence between Multiple Cameras for People Tracking. IEEE Transactions on Pattern Analysis and Machine Intelligence 28(4), 663–671 (2006)
5. Morioka, K., Mao, X., Hashimoto, H.: Global Color Model Based Object Matching in the Multi-Camera Environment. In: Proceedings of the 2006 IEEE/RSJ International Conference on Intelligent Robots and Systems, pp. 2644–2649 (2006)
6. Porikli, F.M.: Inter-Camera Color Calibration by Cross-Correlation Model Function. In: IEEE International Conference on Image Processing, vol. 2, pp. 133–136 (2003)
7. Javed, O., Shafique, Z., Rasheed, Z., Shah, M.: Modeling Inter-camera Space-time and Appearance Relationships for Tracking across Non-overlapping Views. Computer Vision and Image Understanding 109, 146–162 (2008)
8. Seo, D.W., Chae, H.U., Jo, K.H.: Multiple Camera-Based Correspondence of Ground Foot for Human Motion Tracking. Journal of Institute Control, Robotics and Systems 14(8), 848–855 (2008)
9. Kim, T.H., Jo, K.H.: Robust Background Acquisition from Dynamic Scene caused by a Moving Camera. International Journal of Information Acquisition 4(3), 227–236 (2007)
10. Chae, H.U., Kang, S.J., Jo, K.H.: Identification of a Human using Accorded Blobs on the Varied Region from Image Sequence by Multiple Cameras. In: International Conference on Control, Automation and Systems, pp. 1887–1891 (2008)
11. Dempster, A.P., Laird, N.M., Rubin, D.B.: Maximum Likelihood from Incomplete Data via the EM Algorithm. Journal of the Royal Statistical Society, Series B 39(1), 1–38 (1997)

Web-Based Unified-Directory Service for Social Networking Services and Ubiquitous Sensor Network Services

Yung Bok Kim

Department of Computer Engineering/Venture Biz Center
Sejong University, Seoul, Korea, 143-747
`yungbkim@sejong.ac.kr`

Abstract. For integrated social networking and sensor networking services, a unified approach using a unified directory service based on web-based directory was studied. As a convenient and usable mobile web service for unified social/sensor networking service, the multi-lingual single-character domain names as mobile user interface for accessing the metadata of social/sensor information in unified directory are convenient, efficient and consistent. For searching for social/sensor information as well as registering metadata of sensor/social information, we introduce the web-based unified-directory service with the requirements, performance metrics for QoS, resource utilization and real-time estimation of the performance metrics.

1 Introduction

Mobile social networking services, based on mobile interaction, are emerging as a result of the evolution of mobile communication technology and multimedia mobile devices in the ubiquitous information society. Various sensor networks composed of nodes are being deployed, and we found an analogy between human-centric social networking and highly evolved sensor networking. We attempted to find a unified approach for integration and found an analogy between a social and sensor network on the basis of a unified directory service.

A social network is a social structure made of nodes that are generally individuals or organizations. The maximum size of a social network tends to be approximately 150 people and the average size is about 124 [1]. To better understand the concept of culture, and how it is related to human-computer interaction, Ford and Kotze [2] state that culture, as patterns of thinking, feeling and acting, influences the way in which people communicate amongst themselves and with computers. We studied interaction with mobile devices as well as with sensors beyond 'computers' because of the proliferation of mobile devices and applications with intelligence based on context-awareness.

The complexity involved with providing consistent information access from portals or community sites has been increasing and the inconvenience of user interaction for information access has become confusing for even skilled users. For sensor networking, the complexity of accessing to the metadata information in distributed directories will be increased significantly. We suggest and introduce a ubiquitous and unified directory service with the convenient interaction in social/sensor networking.

The mobile phone can be the sink node or the sensor node in a ubiquitous networking environment because the mobile service has been available worldwide even before the proliferation of sensor networks. The mobile device has several features that support the use in the middle of sensor networks; beside local connectivity with stand-alone sensor motes, connecting sensor networks to the Internet on a global scale creates endless opportunities for applications and services, as well as new emerging models of operation [3]. We studied the mobile interaction between a user and the unified USN (ubiquitous sensor network) directory for highly evolved sensor networking applications.

In the following sections, we discuss requirements for social/sensor networking service considering social networking types. We also discuss a unified and web-based directory service for the social/sensor networking types. We then discuss the performance metrics for QoS in accessing a unified directory. Resource utilization and real-time estimation are presented. Mobile interaction for a unified directory service is discussed on the basis of the implementation of a unified web site. Finally, we conclude our study with a consideration of a unified social/sensor directory service including the metadata of sensors in u-City, a highly evolved sensor networking environment.
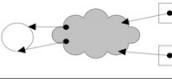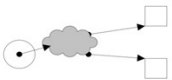
2   Requirements for Social/Sensor Network Services

We characterized the types of social networking. With e-mail or an SMS/MMS, we can communicate with one or several persons via sending a message (i.e. for out-going) as well as by receiving a message from them (i.e. for in-coming). With an instant messenger, we can send a message to a person or to a group of people; therefore we can communicate in characteristics of one-to-one or one-to-many. In a café or on a bulletin board, we can communicate with shared information or message in the form of many-to-many (notated as n in the Fig.1). Similarly, on a blog or mini home page (i.e. the personal home page), each person can send shared information or messages to many persons (i.e. in the form of one to many).

The evolution sequence is as follows. The individual user has his/her own ID or number, e-mail, mobile phone, bulletin board, home page, café, blog, mini home page, mobile blog, instant messenger and so forth. Each has its own advantages and disadvantages (i.e. pros and cons). A large portion of e-mails being received everyday is spam, and with white-lists such as the buddy-list in instant messaging people often communicate within closed circles [4]. If the originator in social networking tools is easily traceable, then the amount of spam is relatively small, however even SMS spam and IP-phone junk calls are commonplace, unless we initiate appropriate safeguards.

Most sensor network researchers would probably agree that we have placed too much attention on the networking of distributed sensing and too little on the tools necessary to manage, analyze, and understand the data [5]. However, with standards in place for networking and other infrastructure-related issues, or at least an emerging consensus as to how to solve these issues, the demands for a sensor web revolve around questions of information management. In terms of service discovery, service discovery protocols are designed to minimize administrative overhead and increase usability [6].

Social Networking Types

Fig. 1. Social Networking Types and Characteristics
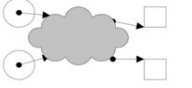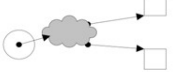
Design goals for a unified web service must be considered on the basis of a real service from a user's perspective. We tried to satisfy the design concept for the implementation of a unified and web-based directory for mobile interaction between human beings and social groups as well as for interaction between sensor nodes and sensor network. We tried to improve the design goals. We introduce the current status of our research to apply the insight of social networking to the advanced sensor networking for the u-City of the future, especially for the integration of two heterogeneous types of networking.

3   Unified Directory Service for Social/Sensor Network

Ubiquitous applications need a middleware platform to interface between many different sensors/transducers and applications. Adding context awareness to middleware increases device usability and allows better user interaction [7]. We found that the ontology for semantics is heavily related to the word or letter of a multilingual language such as Korean. We found also the first letter of a person's name in social networking as well as a name of a sensor network/node in sensor networking could be used as an index key in a directory service. Therefore, if we use the first letter of names for social/sensor networking service, integrated user interface will be quite convenient. We studied using the consonants and vowels of the first multi-lingual letter as index keys and domain names for real-time access of information with mobile devices.

We characterized the sensor networking types to integrate with a social networking area for the unified directory service. In the sensor networking type in Fig.2, there is no posting type as in the social networking type. We can use the metadata directory for posting the sensed information.

Sensor Networking Types

Fig. 2. Sensor Networking Types

For consistency of information when using real-time social/sensor networking services, in addition to usable user interfaces, we need a unified and a ubiquitous web service based on the wired and mobile Internet. Fast and convenient access to metadata information as well as registration is required for the unified social/sensor networking service. We need to write the information in the ubiquitous and unified directory and make it accessible with single-character multilingual domain names instead of long URL strings for convenient web-based information management.

For ubiquitous information portals of social/sensor networking services, the unified information service is indispensable in social/sensor networking even before the full-browsing capability of mobile phone. The full-browsing phone becomes popular. The ubiquitous web server must have the capability of showing unified contents, i.e. the HTML contents for wired Internet as well as the mobile contents for many different kinds of mobile Internet and various contents using different mobile markup languages, e.g. WML (Wireless Markup Language) and mHTML (mobile Hypertext Markup Language). Easy typing of the URL is also important especially for usable mobile Internet interaction. The single-character multilingual domain names, which could become the indexing key character for information retrieval from ubiquitous and unified directory as well as for information registration into unified and web-based directories, have been considered a convenient user interface for unified social/sensor networking services.

4   Performance Metrics for QoS

We need to consider the accessibility to ubiquitous web-based directory sites as well as the performance of the mobile interaction for a unified web service. We used a single web server for metadata information access as a unified service for the simplicity of

information management and for cost-effectiveness in this study. This method gives effectiveness and efficiency for the access of information and the utilization of resources, in terms of the bandwidth for communication and the size of disk storage to build ubiquitous web-based directory sites for mobile social networking service as well as the metadata USN directory service.

Understanding user experiences taking place in different social contexts and physical and cultural environments is not an easy task for designers [8] or for performance analysts. We studied the important performance metric, e.g. delay from the user's perspective as well as the performance metric, e.g. delay, not only with the time in the network and server, but also with the spent time by the user and the input interaction time with the keypads for URLs or metadata information for the notification/registration interaction with the web site for mobile social community services as well as for the metadata USN directory service.

As performance metrics for the end-to-end QoS, we assume that the random variables, the round-trip response time for a user's single interaction in a session, from user to the contents in a unified directory through wired/mobile Internet before the next interaction with a mobile phone is R. The preparation time for any user/agent in online social/sensor networking to get devices for interaction is U. The time spent by the user/agent with mobile devices to do an appropriate interaction for a service is D. The aggregate interaction time to the web-based and unified directory server after the mobile device through networks for service is S. The interaction time depending upon social/sensor information in the metadata directory DB is C.

For the pulling service, we may order the dominating factors in the overall performance of mobile interaction from the user/agent's perspective as follows. In general, the relationship between the delay metrics for interaction with a mobile device could be $U>D>S>C$. Here, we need to decrease the major interaction times U and D for the pulling service with usable mobile interaction, as well as the network and server interaction time S (or interaction time to contents or metadata in a unified DB, C).

To be time-deterministic for an application with QoS (Quality of Service) in the unified directory service, e.g. the searching of required metadata in one minute, the interaction with the device must be usable, the unified directory server must be efficient and must have a high interaction performance, and the metadata of social/sensor information for interaction must be simplified in an efficient format. The user/agent's preparation time for metadata access, or the ubiquity metric, $U,$ will decrease (i.e. improve) depending upon the proliferation of ubiquitous computing and networking environments. The average keypad-press number with mobile phones is related to the interaction time related to the mobile UI agent.

The aggregate interaction time is related to the performance metric, which can be improved with a simple user/agent interface and a web-based metadata USN directory for the sensor web, especially for usable interaction in a unified and web-based directory service.

5   Resource Utilization and Real-Time Estimation

We need a social/sensor network (SSN) directory for a unified service for social networking as well as for the sensor networking service. For this unified service, we need

a unified directory service including the integrated metadata information for social networking as well as sensor networking service. We used a 'HandBoard' (i.e. online white*board* service using *hand*-sized information in PC or *hand*held phone) as a common posting way for both the metadata information for sensor networks and the person/group information for the social networking service.

Fig.3 shows the integrated social-sensor networking service with 'HandBoard' for web-based unified-directory service. For general application as well as a specific application (e.g. web name card service) with the social-sensor network (SSN), directories are shown on the basis of out-going and in-coming characteristics.

We used a single web-based server for the simplicity of management and the cost-effectiveness of implementation for the text-based web information portals for the mobile social networking services. This helps with the reliability of duplicated web-based directory servers, which will be implemented. We considered the effectiveness and efficiency of the management of web-based information and the utilization of resources for social networking in terms of the bandwidth for communication and the size of Web DB for usable mobile interaction in a ubiquitous web service.

We need to form an efficient way for resources to interact with the unified and web-based directory using pulling interaction schemes instead of the push-style interaction scheme in the real-time social networking service as well as in the metadata information management for the sensor networking service. This concept will be important for the unified directory services in the ubiquitous social/sensor networking services because of the requirement for real-time characteristics and QoS.

Fig. 3. Social-Sensor Networking Service with HandBoard

Fig. 4. Resource Utilization vs Ubiquitous Accessibility for Social Net

The adaptability to sensor networking will increase according to the proliferation of the ubiquitous accessibility by individual users. Depending upon the application, the integration of social networking and sensor networking may or may not be possible, adequate, or valuable. We compared the pulling interaction schemes, e.g. mini home-pages, blogs, mobile blogs, UCCs, mobile UCCs, and 'HandBoard' services, in terms of the resource consumption for interaction, i.e. telecommunication bandwidth and disk space. As a scheme for '1-to-many' interaction service (i.e. one information provider notifies social information to many participants or customers) the above pulling interaction scheme seems to be satisfactory, even though the levels of efficiency and effectiveness are very different.

Resource utilization for interaction is also different from one interaction scheme to another. We do not need to analyze each interaction scheme in detail; instead we can differentiate the above services into two interaction schemes as follows. One scheme, based on multimedia contents, i.e. mini homepages, blogs, mobile blogs, UCCs, mobile UCC services, and another interaction schemes, is using only text-based contents with a limited size, i.e. our proposed **'HandBoard'** interaction service. We chose the **'Hand-Board'** interaction service because of the efficiency and effectiveness of resource utilization and TTS (text to speech) functionality for the mobile social networking service with hand-held mobile devices, especially for usable mobile interaction. This concept is applicable to the unified directory service including the metadata USN directory for static or dynamic metadata information management.

We considered resource utilization for sensor networking as follows. With a different DB table for the metadata USN directory service, the manager or user components for the web-based USN directory service can efficiently manage the metadata

information. The resource utilization for sensor networking is very important because the QoS performance for the unified directory service is heavily dependent upon the scheme of the real-time metadata USN directory service.

Important issues such as data quality of sensor nodes, *real-time aggregation and estimation of random variables* (e.g. temperature, humidity, blood pressure, beat rate, data-update frequency, etc) for the social/sensor networking service have been considered. For **real-time estimation**, we used an exponentially weighted moving average model with the appropriate smoothing parameter α to get the mean value of the random variable x_k. To find the statistical outlier we need to get deviation, and we used the mean absolute deviation (*MAD*) model to reduce the complexity of calculation.

$$\overline{x_k} = \alpha x_k + (1-\alpha)\overline{x_{k-1}}, \tag{1}$$

where $0 < \alpha < 1$. The mean absolute deviation (*MAD*) is defined by

$$MAD = \tfrac{1}{N} \sum_{i=k-N+1}^{k} \left| x_i - \overline{x_k} \right|, \tag{2}$$

where the chosen sample size for statistics is *N,* that is the number of samples to be stored in memory for estimation, the samples have values x_i, the mean is $\overline{x_k}$ (*k > N-1*) and $N \approx \dfrac{1}{\alpha}$. If the smoothing parameter α is too small, *N* becomes larger. For sensor networks, if N_s is the number of sensors, then the memory size for real-time estimation becomes $N_s * N = N_s * \dfrac{1}{\alpha}$. The right place for real-time estimation may be chosen on the basis of the appropriate requirement, for example in a sensor/sink node, in a gateway node, in a middleware, or in an application program.

For mobile interaction in a web service, the implemented system is not only based on wired or mobile Internet, but also many multilingual single-character domains for fast access to the unified directory, the mobile social networking service as well as the sensor networking service. We implemented a web service for a lecture group, composed of approximately 120 students, and an alumni association, composed of about 60 members, and various social networking services. We focused on the social networking, and we applied sensor networking assuming the number of interacting nodes to be between 70 and 160, which is the similar number of sensor nodes in an automobile.

In the deployment of real sensor networks, the appropriate group size of sensor nodes for each sensor network as well as the appropriate group size of sensor networks for the unified directory service will be realized depending upon the application service including the u-City (ubiquitous city).

The pulling of metadata in the directory can be considered similarly to the sampling theorem, the pulling (i.e. sampling) frequency f_s, which is the frequency of pulling the ubiquitous record in the metadata DB table, i.e. a searched web site for mobile social

networking service, and f_u is the frequency of notification (i.e. updating) in 'Hand-Board' by the association for mobile social networking service. Then, $f_s \geq 2f_u$.

If T_u is the mean of updating period, and t_u is the random variable of updating time interval, then

$$T_u = \overline{t_u} = at_k + (1-a)\overline{t_{k-1}}. \tag{3}$$

The sampling frequency f_s may be different for each random variable, and

$$f_s \geq 2f_u = \frac{2}{T_u} = \frac{2}{\overline{t_u}} = \frac{2}{at_k + (1-a)\overline{t_{k-1}}}. \tag{4}$$

(f_u is the updating or changing frequency of a random variable).

With the pulling (i.e. sampling) frequency f_s, the member in the social network and the monitoring agent in the sensor network need to pull the metadata information in the unified directory. In the directory service, the pulling rate by manager or components for the updated metadata information of a sensor network or a sensor node will be very similar to the above concept. It will be also helpful for the queuing analysis within the above arrival rate, i.e. as a practical pulling rate, for the USN directory service. With any mobile device, the directory service for social interaction as well as the sensor networking will be feasible in ubiquitous computing and networking environments with the real-time information management. We implemented the web-based uni-fied-directory service with a single-character multilingual alphabet as well as with the English alphabet as shown in Fig.5.
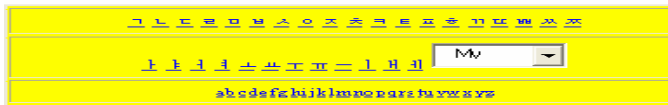
Fig. 5. Web-based Unified Directory with Multilingual Alphabet and English Alphabet

6 Conclusions

A unified approach for an integrated social/sensor networking service was studied. As a convenient and usable mobile interface for mobile web service for online social/sensor networks, the multi-lingual single-character domain names as indexing keys to so-cial/sensor information in ubiquitous web service are convenient mobile user inter-faces. We studied the application of the metadata directory service because of the analogy. The convenience of multilingual single-character domain-names and the tiny 'HandBoard' for usable mobile interaction was studied for a unified and web-based directory service. We implemented a unified directory service for the social/sensor networking services. The unified directory service including applications for the u-City will be studied further on the basis of this unified directory.

References

1. Hill, R., Dunbar, R.: Social Network Size in Humans. Human Nature 14(1), 53–72 (2002)
2. Ford, G., Kotzé, P.: Designing usable interfaces with cultural dimensions. In: Costabile, M.F., Paternó, F. (eds.) INTERACT 2005. LNCS, vol. 3585, pp. 713–726. Springer, Heidelberg (2005)
3. Boda, P.P.: SensorPlanet: An Open Global Research Framework for Mobile Device Centric Wireless Sensor Networks. MobiSensors, NSF Workshop on Data Management for Mobile Sensor Networks, Pitts-burgh, 16–17 (2007)
4. Douglis, F.: On Social Networking and Communication Paradigms. IEEE Internet Computing, 4–6 (2008)
5. Balazinska, M., Deshpande, A., Franklin, M.J., Gibbons, P.B., Gray, J., Nath, S., Hansen, M., Liebhold, M., Szalay, A., Tao, V.: Data Management in the Worldwide Sensor Web. IEEE Pervasive Computing, 30–40 (2007)
6. Zhu, F., Mutka, M.W., Ni, L.M.: Service Discovery in Pervasive Computing Environments. IEEE Pervasive Computing, 81–90 (2005)
7. Costa, C.A., Yamin, A.C., Geyer, C.F.R.: Toward a General Software Infrastructure for Ubiquitous Computting. IEEE Pervasive Computing, 64–73 (2008)
8. Roibas, A.C., Geerts, D., Furtado, E., Calvi, L.: Implications of the Socio-physical Contexts when Interacting with Mobile Media. Pers Ubiquit Comput., 279–280 (2008)

Bilateral Negotiation in a Multi-Agent Energy Market

Fernando Lopes[1], A.Q. Novais[1], and Helder Coelho[2]

[1] LNEG − National Research Institute, Dept. of Modelling and Simulation,
Estrada do Paço do Lumiar 22, 1649-038 Lisbon, Portugal
{fernando.lopes,augusto.novais}@ineti.pt
[2] University of Lisbon, Dept. of Computer Science,
Bloco C6, Piso 3, Campo Grande, 1749-016 Lisbon, Portugal
hcoelho@di.fc.ul.pt

Abstract. Energy markets are systems for effecting the purchase and sale of electricity using supply and demand to set the price. A typical energy market involves a wholesale market for electricity generation, when competing generators offer their electricity output to retailers, and a retail market for electricity retailing, when end-use customers choose their supplier from competing electricity retailers. This paper addresses the challenges created by competitive energy markets towards ensuring the full benefits of deregulation. It presents a multi-agent energy market composed of multiple autonomous computational agents, each responsible for one or more market functions, and each interacting with other agents in the execution of their responsibilities. Additionally, the paper presents a negotiation model for autonomous agents. The model handles bilateral multi-issue negotiation and formalizes a set of negotiation strategies studied in the social sciences and frequently used by human negotiators.

1 Introduction

Multi-agent systems (MAS) have generated lots of excitement in recent years because of their promise as a new paradigm for conceptualizing, designing, and implementing complex software systems. The major motivations for the increasing interest in MAS research include the ability to solve problems in which data, expertise, or control is distributed, the ability to allow inter-operation of existing legacy systems, and the ability to enhance performance along the dimensions of computational efficiency, reliability, and robustness. Agent technology has been used to solve real-world problems in a range of industrial and commercial applications, including manufacturing, process control, telecommunications, air traffic control, information management, electronic commerce, and business process management (see, e.g., [3, 12, 13]).

The electrical power industry provides the production and delivery of electricity to businesses and households through a grid. This industry is commonly split into four processes, namely electricity generation, electric power transmission, electricity distribution, and electricity retailing. Electricity is most often produced at power stations, transmitted at high-voltages to multiple substations near populated areas, and distributed at medium and low-voltages to consumers. Electricity retailing has not changed much over time − customers are normally charged based on the amount of energy consumed.

The electrical power industry was traditionally heavily regulated with extensive public ownership, federalized organisational structures, and a lack of market-price mechanisms. Its deregulation basically separates the contestable functions of electricity generation and retail from the natural monopoly functions of transmission and distribution. This, in turn, leads to the establishment of a wholesale market for electricity generation, when competing generators offer their electricity output to retailers, and a retail market for electricity retailing, when end-use customers choose their supplier from competing electricity retailers. Clearly, opening up electricity production to competition is an important tool to improve the efficiency of the electricity production industry and therefore to benefit all electricity consumers. Competitive forces can drive producers to innovate and operate in more efficient and economic ways. Innovation can lead to lower prices and a better use of energy resources.

Energy markets in general and multi-agent energy markets in particular have received some attention lately (see, e.g., [2, 14, 20]). However, despite these and other relevant pieces of work, most challenges created by deregulation are still waiting to be addressed more thoroughly. At present, there is a need to develop computational tools to help manage the complexity of energy markets towards ensuring long-term capacity sustainability. Against this background, the purpose of this paper is twofold:

1. *to present a multi-agent energy market* − the market is composed of a collection of autonomous computational agents, each responsible for one or more market functions, and each interacting with other agents in the execution of their responsibilities;
2. *to present a negotiation model for autonomous agents* − the model handles bilateral multi-issue negotiation and formalizes a set of negotiation strategies studied in the social sciences and frequently used by human negotiators.

This paper builds on our previous work in the area of automated negotiation [7, 8, 9]. In particular, it considers a number of strategies based on rules-of-thumb distilled from behavioral negotiation theory. It also lays the foundation for performing an experiment to investigate the performance of agents operating in the energy market and equipped with the negotiation model.

The remainder of the paper is structured as follows. Section 2 describes a multi-agent energy market. Section 3 presents a negotiation model for autonomous agents, focusing on the operational and strategic process of preparing and planning for negotiation (usually referred to as pre-negotiation), and the central process of moving toward agreement (usually referred to as actual negotiation, or simply negotiation). Finally, related work and concluding remarks are presented in sections 4 and 5 respectively.

2   Multi-Agent Energy Market

Multi-agent systems are ideally suited to represent problems that have multiple problem solving entities and multiple problem solving methods. Central to the design and effective operation of a multi-agent system are a core set of problems and research questions, notably:

1. *the design problem* − how to formulate, describe, decompose, and allocate different problems and synthesize results among a group of intelligent agents?
2. *the coordination problem* − how to ensure that agents act coherently, accommodating the local decisions or non-local effects and avoiding harmful interactions?

The design problem is focused on the domain the system is intended to solve in a distributed manner, *i.e.*, a deregulated energy market involving a wholesale market and a retail market. Practically speaking, the role of the wholesale market is to allow trading between generators and retailers both for short-term delivery of electricity and for future delivery periods − competing generators can offer their electricity output to retailers. The role of the retail market is to allow trading between energy consumers and retailers − end-use customers can choose their supplier from competing electricity retailers. Accordingly, we consider the following types of agents:

1. generators or producers, who in aggregate sell to the wholesale market;
2. retailers or suppliers, who in aggregate buy from the wholesale market and sell to the retail market;
3. customers or consumers, who in aggregate buy from the retail market.

The agents are computer systems capable of flexible autonomous action in order to meet their design objectives. They can to respond in a timely fashion to changes that occur in the environment, exhibit goal-directed behavior, and interact with other agents in order to reach their design objectives. In particular, competing generators can interact with various retailers to offer their electricity output and, mainly, end-use customers can interact with competing electricity retailers to choose their supplier.

The coordination problem is focussed on ensuring that autonomous agents act in a tightly coordinated manner in order to effectively achieve their goals. This problem is addressed, at least in part, by designing agents that are able to coordinate their activities through negotiation. Specifically, for the case of a deregulated market, the agents are charged with executing actions towards the achievement of their private goals and, thus, conflicts inevitably occur among them. Negotiation is the predominant process for resolving conflicts. Accordingly, the agents are equipped with a negotiation model enabling them to: (i) prepare and plan for negotiation, (ii) generate, evaluate and exchange offers, (iii) come to agreements acceptable to all parties, and (iv) implement final agreements, *i.e.*, determine who needs to do what once contracts are signed.

Now, in order to move towards the full benefits of deregulation, we put forward the following requirement for market design: the arrangement of customers' electricity supply should be achieved by having contracts that specify the provision of an amount of energy for a certain period of time (e.g., one hour). At present, the overall energy consumption in several countries seems to be increasing at a faster pace than energy production. The consequences of not being able to support the demand of energy in the near future are seriously worrying: brownouts or blackouts and the subsequent economical loses. Furthermore, peak loads can approach or even go beyond existing supply capabilities. An appealing method to prevent this worst-case scenario consists of reducing the energy demand and, if this is not possible, executing energy intensive processing tasks whenever the demand of energy is lower.

To this end, there have been a number of initiatives to distribute energy demand over time to avoid peak loads. In particular, many companies have already presented a two-rate tariff to smooth the daily demand profile (cheaper night tariff). This dual model can easily be refined if, instead of two rates, companies offer three rates or even an hour-wise tariff. For instance, a company can present a three-rate tariff by considering that from 7 to 9 am and from 7 to 9 pm are peak hours, from 12 am to 3 pm and from 10 pm to 11 pm is a medium-load period and, finally, the rest is low-demand time.

Furthermore, we put forward another requirement for market design: suppliers and consumers should be able to enter into contracts to protect themselves from volume and price risks. Volume risk is a term used to denote the phenomenon whereby the market participants have uncertain quantities of consumption or production. Price risk is a term used to denote extremely high price volatilities at times of peak demand and supply shortages.

3   Multi-Agent Negotiation

Negotiation is an important and pervasive form of social interaction − it may involve two parties (bilateral negotiation) or more than two parties (multilateral negotiation), and one issue (single-issue negotiation) or many issues multi-issue negotiation). This section briefly introduces a model for autonomous agents that handles two-party multi-issue negotiation (see [7, 8, 9] for an in-depth discussion).

3.1   Pre-negotiation

Pre-negotiation is the process of preparing and planning for negotiation and involves mainly the creation of a well-laid plan specifying the activities that negotiators should attend to before actually starting to negotiate [6]. Accordingly, we describe below various activities that negotiators make efforts to perform in order to carefully prepare and plan for negotiation.

Let $Ag = \{ag_1, ag_2\}$ be the set of autonomous negotiating agents. Let $Agenda = \{is_1, \ldots, is_n\}$ be the negotiating agenda − the set of issues to be deliberated. The issues are quantitative variables, defined over continuous intervals. Effective pre-negotiation requires that negotiators prioritize the issues, define the limits, and specify the targets. Priorities are set by rank-ordering the issues, $i.e.$, by defining the most important, the second most important, and so on. The priority pr_{il} of an agent $ag_i \in Ag$ for each issue $is_l \in Agenda$ is a number that represents its order of preference. The weight w_{il} of is_l is a number that represents its relative importance. The limit lim_{il} or resistance point is the point where ag_i decides that it should stop to negotiate, because any settlement beyond this point is not minimally acceptable. The level of aspiration or target point trg_{il} is the point where ag_i realistically expects to achieve a settlement.

Additionally, effective pre-negotiation requires that negotiators agree on an appropriate protocol that defines the rules governing the interaction. The protocol can be simple, allowing agents to exchange only proposals. Alternatively, the protocol can be complex, allowing agents to provide arguments to support their negotiation stance. However, most sophisticated protocols make considerable demands on any

implementation, mainly because they appeal to very rich representations of the agents and their environments (see, e.g., [4, 10]). Therefore, we consider an alternating offers protocol [11]. Two agents or players bargain over the division of the surplus of $n \geq 2$ issues (goods or pies) by alternately proposing offers at times in $T = \{1, 2, \ldots\}$. The negotiation procedure, labelled the "joint-offer procedure", involves bargaining over the allocation of the entire endowment stream at once. An offer is a vector $(x_1, \ldots, x_n)$ specifying a division of the n goods. Once an agreement is reached, the agreed-upon allocations of the goods are implemented.

The players' preferences are modelled by assuming that each player ag_i discounts future payoffs at some given rate δ_i^t, $0 < \delta_i^t < 1$, (δ_i^t is referred to as the discount factor). The cost of bargaining derives from the delay in consumption implied by a rejection of an offer. Practically speaking, the justification for this form of preferences takes into account the fact that money today can be used to make money tomorrow. Let U_i be the payoff function of ag_i. For simplicity and tractability, we assume that U_i is separable in all their arguments and that the per-period delay costs are the same for all issues:

$$U_i(x_1, \ldots, x_n, t) = \delta_i^{(t-1)} \sum_{l=1}^n w_{il}\ u_{il}(x_l)$$

where w_{il} is the weight of is_{il} and x_l denotes the share of ag_i for is_{il}. The component payoff function u_{il} for is_{il} is a continuous, strictly monotonic, and linear function. The distinguish feature of time preferences with a constant discount rate is the linearity of the function u_{il} [11]. The payoff of disagreement is normalized at 0 for both players.

3.2   Actual Negotiation

Actual negotiation is the process of moving toward agreement (usually by an iterative exchange of offers and counter-offers). The negotiation protocol defines the states (e.g., accepting a proposal), the valid actions of the agents in particular states (e.g., which messages can be sent by whom, to whom, at what stage), and the events that cause states to change (e.g., proposal accepted). It marks branching points at which negotiators have to make decisions according to their strategies. Hence, this section formalizes a set of strategies studied in the social sciences and frequently used by human negotiators.

Negotiation strategies can reflect a variety of behaviours and lead to strikingly different outcomes. However, the following two fundamental groups of strategies are commonly discussed in the behavioral negotiation literature [15, 19]:

1. *concession making* − negotiators who employ strategies in this group reduce their aspirations to accommodate the opponent;
2. *problem solving* − negotiators maintain their aspirations and try to find ways of reconciling them with the aspirations of the opponent.

Two explanatory and cautionary notes are in order here. First, most strategies are implemented through a variety of tactics. The line between strategies and tactics often seems indistinct, but one major difference is that of scope. Tactics are short-term moves designed to enact or pursue broad (high-level) strategies [6]. Second, most strategies are only informally discussed in the behavioral literature − they are not formalized, as typically happens in the game-theoretic literature.

Concession making behaviour aims at partially or totally accommodating the other party. Consider two incompletely informed agents bargaining over n distinct issues $\{is_1, \ldots, is_n\}$. For convenience, each issue is_l is modelled as an interval $[min_l, max_l]$. The opening stance and the pattern of concessions are two central elements of negotiation. Three different opening positions (extreme, reasonable and modest) and three levels of concession magnitude (large, moderate and small) are commonly discussed in the behavioral literature [6]. They can lead to a number of concession strategies, notably:

1. *starting high and conceding slowly* − negotiators adopt an optimistic opening attitude and make successive small concessions;
2. *starting reasonable and conceding moderately* − negotiators adopt a realistic opening attitude and make successive moderate concessions;
3. *starting low and conceding rapidly* − negotiators adopt a pessimistic opening attitude and make successive large concessions.

Let $p_{j \to i}^{t-1}$ be the offer that ag_j has proposed to ag_i in period $t-1$. Likewise, let $p_{i \to j}^{t}$ be the offer that ag_i is ready to propose in the next time period t. The formal definition of a generic concession strategy follows.

Definition 1. *Let $ag_i \in Ag$ be a negotiating agent. A concession strategy for ag_i is a function that specifies either the tactic to apply at the beginning of negotiation or the tactic that defines the concessions to be made during the course of negotiation:*

$$conc \stackrel{def}{=} \begin{cases} apply\ tact_i^1 & if\ ag_i\text{'s turn and } t=1 \\ apply\ tact_i^t & if\ ag_i\text{'s turn and } t>1 \\ if\ U_i(p_{j \to i}^{t-1}) \geq U_i(p_{i \to j}^{t}) \quad accept\ else\ reject & if\ ag_j\text{'s turn} \end{cases}$$

where $tact_i^1$ is an opening negotiation tactic and $tact_i^t$ is a concession tactic.

The two aforementioned concession strategies are defined by considering different tactics. For instance, the "starting reasonable and conceding moderately" strategy is defined by: "$tact_i^1 = starting\_realistic$" and "$tact_i^t = moderate$" (but see below).

Problem solving behaviour aims at finding agreements that appeal to all sides, both individually and collectively. This behaviour can take several forms, notably logrolling − negotiators agree to trade-off among the issues under consideration so that each party concedes on issues that are of low priority to itself and high priority to the other party [15, 19]. Effective logrolling requires information about the two parties' priorities so that concessions can be matched up. This information is not always easy to get. The main reason for this is that negotiators often try to conceal their priorities for fear that they will be forced to concede on issues of lesser importance to themselves without receiving any repayment [15]. Despite this, research evidence indicates that it is often not detrimental for negotiators to disclose information that can reveal their priorities − a simple rank order of the issues does not put negotiators at a strategic disadvantage [19]. Hence, we consider that negotiators willingly disclose information that can help to identify their priorities (e.g., their interests). The formal definition of a generic logrolling strategy follows.

Definition 2. *Let $ag_i \in Ag$ be a negotiating agent and $ag_j \in Ag$ be its opponent. Let Agenda denote the negotiating agenda, $Agenda^\oplus$ the subset of the agenda containing the issues of high priority for ag_i (and low priority for ag_j), and $Agenda^\ominus$ the subset of the agenda containing the issues of low priority for ag_i (and high priority for ag_j). A logrolling strategy for ag_i is a function that specifies either the tactic to apply at the beginning of negotiation or the tactics to make trade-offs during the course of negotiation:*

$$log \overset{def}{=} \begin{cases} apply\ tact_i^1 & if\ ag_i\text{'s turn and } t=1 \\ apply\ tact_i^{t\oplus}\ and\ tact_i^{t\ominus} & if\ ag_i\text{'s turn and } t>1 \\ if\ U_i(p_{j\rightarrow i}^{t-1}) \geq U_i(p_{i\rightarrow j}^t) \quad accept\ else\ reject & if\ ag_j\text{'s turn} \end{cases}$$

where $tact_i^1$ is an opening negotiation tactic, $tact_i^{t\oplus}$ is a concession tactic (to apply to the issues on $Agenda^\oplus$), and $tact_i^{t\ominus}$ is another concession tactic (to apply to the issues on $Agenda^\ominus$).

A number of logrolling strategies can be defined simply by considering different tactics. For instance, a strategy that specifies an optimistic opening attitude followed by null concessions on issues on $Agenda^\oplus$ and small concessions on issues on $Agenda^\ominus$ is defined by: "$tact_i^1 = starting\_optimistic$", "$tact_i^{t\oplus} = stalemate$", and "$tact_i^{t\ominus} = tough$". Similarly, a strategy that specifies a realistic opening attitude followed by null concessions on issues on $Agenda^\oplus$ and large concessions on issues on $Agenda^\ominus$ is defined by: "$tact_i^1 = starting\_realistic$", "$tact_i^{t\oplus} = stalemate$", and "$tact_i^{t\ominus} = soft$" (but see below).

At this stage, it is worth making the point that logrolling is a major route − though not the only route − to the development of mutually superior solutions (*i.e.*, solutions that are better for all parties). In fact, the host of existing problem solving strategies includes expanding the "pie", nonspecific compensation, cost cutting, and bridging. These strategies are implemented by different sets of tactics and require progressively more information about the other parties (see, e.g., [15]).

Opening negotiation tactics are functions that specify the initial values for each issue is_l at stake. The following three tactics are commonly discussed in the behavioral literature [6]:

1. *starting optimistic* − specifies a value far from the target point;
2. *starting realistic* − specifies a value close to the target point;
3. *starting pessimistic* − specifies a value close to the limit.

The definition of the tactic "starting realistic" follows (the definition of the other two tactics is essentially identical, and is omitted).

Definition 3. *Let $ag_i \in Ag$ be a negotiating agent and $is_l \in Agenda$ a negotiation issue. Let trg_{il} be the target point of ag_i for is_l. The tactic starting realistic for ag_i is a function that takes is_l and trg_{il} as input and returns the initial value $v[is_l]_i^1$ of is_l:*

$$starting\_realistic(is_l, trg_{il}) = v[is_l]_i^1$$

where $v[is_l]_i^1 \in [\,trg_{il} - \epsilon,\ trg_{il} + \epsilon\,]$ and $\epsilon > 0$ is small.

Concession tactics are functions that compute new values for each issue is_l. The following five tactics are commonly discussed in the literature [6]:

1. *stalemate* − models a null concession on is_l;
2. *tough* − models a small concession on is_l;
3. *moderate* − models a moderate concession on is_l;
4. *soft* − models a large concession on is_l;
5. *accommodate* − models a complete concession on is_l.

The definition of a generic concession tactic follows (without loss of generality, we consider that ag_i wants to maximize is_l).

Definition 4. *Let $ag_i \in Ag$ be a negotiating agent, $is_l \in Agenda$ a negotiation issue, and lim_{il} the limit of is_l. Let $v[is_l]_i^t$ be the value of is_l offered by ag_i at period t. A concession tactic for ag_i is a function that takes $v[is_l]_i^t$, lim_{il} and the concession factor $Cf \in [0, 1]$ as input and returns the new value $v[is_l]_i^{t+2}$ of is_l:*

$$concession\_tactic(v[is_l]_i^t, lim_{il}, Cf) = v[is_l]_i^{t+2}$$

where $v[is_l]_i^{t+2} = v[is_l]_i^t - Cf\ (v[is_l]_i^t - lim_{il})$.

The five tactics are defined by considering different values for Cf. In particular, the stalemate tactic by $Cf = 0$, the accommodate tactic by $Cf = 1$, and the other three tactics by different ranges of values for Cf (e.g., the tough tactic by $Cf \in\ ]0.00, 0.05]$, the moderate tactic by $Cf \in\ ]0.05, 0.10]$, and the soft tactic by $Cf \in\ ]0.10, 0.15]$).

4   Related Work

Artificial intelligence (AI) researchers have investigated the design of agents with negotiation competence from two main perspectives: a theoretical or formal mathematical perspective and a practical or system-building perspective. Researchers following the theoretical perspective have drawn heavily from game-theoretic and economic methods (see, e.g., [1, 5, 17]). Most researchers have primarily focused on formal bargaining, auctions, market-oriented programming, contracting, and coalition formation. On the other hand, researchers following the practical perspective have drawn heavily on social sciences techniques for understanding interaction and negotiation (see, e.g., [4, 10, 16]. Most researchers have primarily focused on the central process of moving toward agreement, notably the design and evaluation of negotiation protocols and negotiation strategies.

The theoretical models have some highly desirable properties such as Pareto efficiency, stability, and the ability to guarantee convergence. However, most models work with abstract problems and often fail to capture the richness of detail that would be necessary to successfully apply them in realistic domains. Furthermore, most models make the following restrictive assumptions: (i) the agents are rational, (ii) the set of candidate solutions is fixed and known by all the agents, (iii) each agent knows either the other agents' potential payoffs for all candidate solutions or the other agents' potential attitudes toward risk and expected-utility calculations.

Most computational models are being used successfully in a wide variety of real-world domains. These models exhibit the following desirable features: (i) they are based on realistic assumptions, and (iii) they make use of moderate computational resources to find acceptable solutions (according to the principles of bounded rationality [18]). However, most models lack a rigorous theoretical underpinning − they are essentially ad hoc in nature. Also, they often lead to outcomes that are sub-optimal. Finally, there is often no precise understanding of how and why they behave the way they do. Consequently, they need extensive evaluation.

Nevertheless, the class of models referred to as computational models is gaining increasing popularity within the mainstream AI community and therefore has received our attention in this paper. Furthermore, despite the aforementioned pieces of work and other relevant models, we are aware of no similar efforts to define strategies as functions that specify the tactics to be used at every period of negotiation. Tactics, in turn, are defined as functions that specify the short-term moves to be made throughout negotiation. Also, our interest lies in formalizing important strategies studied in the social sciences and frequently used by human negotiators, and in evaluating the effectiveness of these strategies in different situations.

5   Conclusion

This article has presented a simplified multi-agent energy market composed of a collection of autonomous computational agents, each responsible for one or more market functions, and each interacting with other agents in the execution of their responsibilities. Additionally, the article has presented a negotiation model for autonomous agents that handles bilateral multi-issue negotiation and formalizes a set of negotiation strategies studied in the social sciences and frequently used by human negotiators. The strategies are based on rules-of-thumb distilled from behavioral negotiation theory.

Autonomous agents equipped with the negotiation model are currently being developed using the Jade framework and the Java programming language. Our aim for the future is to perform a set of inter-related experiments to empirically evaluate the key components of the agents operating in the energy market. Each experiment will lay the foundation for subsequent experimental work. Also, the current work forms a basis for the development of more sophisticated agents that are able to negotiate under both complete and incomplete information. In particular, we intend to study the bargaining game of alternating offers in order to define equilibrium strategies for two incompletely informed players.

References

1. Fatima, S., Wooldridge, M., Jennings, N.: A Comparative Study of Game Theoretic and Evolutionary Models of Bargaining for Software Agents. Artificial Intelligence Review 23, 185–203 (2005)
2. Gustavsson, R.: Agents with Power. Communications of the ACM 42, 41–47 (1999)
3. Jennings, N., Sycara, K., Wooldridge, M.: A Roadmap of Agent Research and Development. Autonomous Agents and Multi-Agent Systems 1, 7–38 (1998)

4. Jennings, N., Faratin, P., Lomuscio, A., Parsons, S., Wooldridge, M., Sierra, C.: Automated Negotiation: Prospects, Methods and Challenges. Group Decision and Negotiation 10, 199–215 (2001)
5. Kraus, S.: Strategic Negotiation in Multi-Agent Environments. MIT Press, Cambridge (2001)
6. Lewicki, R., Barry, B., Saunders, D., Minton, J.: Negotiation. McGraw Hill, New York (2003)
7. Lopes, F., Mamede, N., Novais, A.Q., Coelho, H.: A Negotiation Model for Autonomous Computational Agents: Formal Description and Empirical Evaluation. Journal of Intelligent&Fuzzy Systems 12, 195–212 (2002)
8. Lopes, F., Mamede, N., Novais, A.Q., Coelho, H.: Negotiation Strategies for Autonomous Computational Agents. In: ECAI 2004, pp. 38–42. IOS Press, Amsterdam (2004)
9. Lopes, F., Novais, A.Q., Coelho, H.: Towards an Interdisciplinary Framework for Automated Negotiation. In: E-Commerce and Web Technologies. LNCS, vol. 5138, pp. 81–91. Springer, Heidelberg (2008)
10. Lopes, F., Wooldridge, M., Novais, A.Q.: Negotiation Among Autonomous Computational Agents: Principles, Analysis and Challenges. Artificial Intelligence Review (in press, 2009)
11. Osborne, M., et al.: Bargaining and Markets. Academic Press, London (1990)
12. Parunak, H.: A Practitioners Review of Industrial Agent Applications. Autonomous Agents and Multi-Agent Systems 3, 389–407 (2000)
13. Pechoucek, M., Vladimir, M.V.: Industrial Deployment of Multi-agent Technologies: Review and Selected Case Studies. Autonomous Agents and Multi-Agent Systems 17, 397–431 (2008)
14. Penya, Y.: Optimal allocation and Scheduling of Demand in Deregulated Energy Markets. In: Dissertation, Vienna University of Technology, Austria (2006)
15. Pruitt, D., Kim, S.: Social Conflict: Escalation, Stalemate, and Settlement. McGraw Hill, New York (2004)
16. Rahwan, I., Ramchurn, S., Jennings, N., McBurney, P., Parsons, S., Sonenberg, L.: Argumentation-based Negotiation. The Knowledge Engineering Review 18, 343–375 (2004)
17. Sandholm, T.: Distributed Rational Decision Making. In: MultiAgent Systems – A Modern Approach to Distributed Artificial Intelligence, pp. 201–259. MIT Press, Cambridge (1999)
18. Simon, H.: The Sciences of the Artificial. MIT Press, Cambridge (1981)
19. Thompson, L.: The Mind and Heart of the Negotiator. Prentice-Hall, Englewood Cliffs (2005)
20. Ygge, F.: Market-Oriented Programming and its Application to Power Load management. In: Dissertation, Lund University, Sweden (1998)

An Approach to Automated User Interest Matching in Online Classified Advertising Systems

Valeriya Gribova and Pavel Kachanov

Institute of Automation and Control Processes
Far Eastern Branch of Russian Academy of Sciences
Radio 5, 690041 Vladivostok, Russia
gribova@iacp.dvo.ru, pavel.kachanov@gmail.com

Abstract. The paper presents an approach to automated user interest matching in online classified advertising systems which is based on the analysis of the structure and semantics of classified ads. A classified advertisement structure and classified ad types along with the examples are described in the paper.

Keywords: Internet, Online classified systems, Intelligent systems, User interest matching.

1 Introduction

Nowadays the online classified advertising systems are becoming one of the most popular and rapidly growing classes of Internet resources. The goal of a user of such web resources is to find a user, a contact and interaction with whom may lead to the matching of their mutual interests.

This goal can also be achieved by using other classes of the Internet resources - general and vertical search engines, but certain differences make online classified advertising systems a separate class. General search engines have the widest search range, but the lowest search depth. Vertical search engines are intended for information retrieval in specific domains, therefore they can perform the deepest search, but only in a narrow field. The class of online classified advertising systems is intermediate between the abovementioned classes.

Statistic data indicates high traffic to online classified advertising systems and proves their relevance and importance. According to comScore Media Metrix research, traffic in July 2006 has grown by 47 percent compared to 2005 and made up 37.4 million visitors, which is about 22% of total amount of Internet users in USA [1].

The most recent research (April 1, 2009) in this area by Hitwise® shows that the amount of traffic to online classified advertising systems continues to grow. The traffic increased 84 percent in February 2009 compared with last year [2].

Today there are a lot of software tools for development of online classified advertising systems, such as Esvon Classifieds, e-Classifieds, PHP Classifieds Script, 68 Classifieds and others. There are also standalone systems that were built without using any public software tools. The most popular online classified advertising systems belong to the latter. Examples of such systems are: Craigslist, Backpage, Oodle, Geebo, Kijiji, Facebook Marketplace etc.

The existing online classified advertising systems use the following methods to match user interests:

1.   Hierarchical classification of the classified ads [3, 4]. A system administrator sets up a set of hierarchical categories, and each classified in a system belongs to a certain category. Such classified structuring makes the active interaction with system easier: users have the ability to look through the subset of all classifieds in a system, not the full list of them.

2.   Either structural or natural language input, or both combined. A standard classified structure for all of the online classified advertising systems is a set of two fields: "Title" and "Description". Simple systems allow users only to fill these two fields when posting a classified. More complicated systems allow their administrators to describe the classified structure as a set of fields of different types. These fields may vary for different classified ad categories. Thus the classified ad author can use these fields when posting a classified, and this makes the description of his/her interest more formal and convenient to search.

3.   Either full-text search or the parameter search based on the classified ad structure [5]. The existing online classified advertising systems allow searching for classified ads according to the input method used in a specific system. Again, a standard for the existing systems is the full-text search. It allows users to find classified ads that contain specific keywords in their descriptions and titles. If a system supports complex classified ad structure, it can also use the additional classified ad fields as the search parameters. Both search types allow a user to choose the category of a desired classified ad. Some systems also allow a user to save the search query and perform the search repeatedly as new classifieds are posted to the system. If there are classifieds that match the specified search conditions, the system sends a notification email to the user.

The analysis of existing systems has shown that these systems and the methods they use to match user interests have serious drawbacks.

These drawbacks can be divided in two groups. The first is the functionality that is present in standalone online classified advertising systems but is lacking in the systems implemented with existing software tools. The second group contains of the common drawbacks of both standalone systems and the systems implemented with software tools.

1.   The drawbacks of the systems implemented with the software tools compared to the standalone systems:

a.   Rigid classified ad structure that does not allow to describe a user's interest in a flexible way. This includes the possibility to extend the classified ad structure by allowing a user to form a specific set of structure fields he/she needs. This also includes the possibility to extend data directories by allowing a user to add the necessary data to these directories if such data does not exist there yet.

b.   The methods of user interests matching do not utilize the interest types (demand or supply).

c.   The methods of user interests matching do not support partial interest matching. The existing systems also lack the ability to calculate the relevance of each matching result to the user's interest.

2.   The common drawbacks of both standalone systems and the systems implemented with software tools:

a.   An action is an essential part of a classified ad structure, but in the existing systems actions are used as parts of hierarchical classifier categories and not in a classified ad structure. This leads to the data duplication and also makes it impossible to use semantics of the actions in user interest matching process.

b.   A classified can describe only a single object of interest while there are classes of classified ads that require the description of several objects.

c.   The existing systems are supply-oriented and do not allow a user to describe an interest that implies demand. The possibility to describe both the desired object and the one being offered (i.e., in classified ads of "exchange A for B" type) is not supported in existing systems.

d.   The lack of automated user interest matching after a new classified has been posted to the system.

e.   The auto-notification features of existing systems are one-way. The saved search query used for the notifications is actually a description of an interest of a user, and it may be useful for those users who have matching interests. But the existing systems make these descriptions internal and not available to other users.

f.   The drawbacks of the administrative tools used to describe domain ontologies. This includes the lack of inheritance support for the ontology classes and the lack of a possibility to use the classes that are already defined as the properties of the new classes.

g.   The drawbacks of the classified ad description methods. This includes support of synonymic values and values with spelling errors.

To eliminate the drawbacks of existing tools a new approach to development of tools for design of online classified advertising systems has been proposed by the authors of the report [6]. One of the main ideas of the approach is to automate the process of user interest matching by applying both active and passive interaction methods. By using the combination of these methods the system automatically chooses classified ads which are potentially interesting for a user.

The aim of this report is to describe the basis for the approach to the automated user interest matching.

2   Basic Ideas of the Approach to Automated User Interest Matching

Automated user interest matching is based on the structure and semantics of classifieds. To provide it we propose the following: to define classified advertisement types in accordance with their structure and semantics; to develop a domain independent ontology of classified ads for matching user interest; to develop extendable ontologies for different domains for forming classified ads using these ontologies and user's information about classified details.

Thus the approach utilizes two basic structures: the interest matching ontology and domain ontologies. The interest matching ontology describes structural and semantic elements of a classified such as actions, objects, links between them etc. The type of a

classified ad is determined in the beginning of the user interest matching process, and the information about the type determines the method of user interest matching.

Domain ontologies contain descriptions of domain objects and relations between them along with the additional information which is useful for interest matching. Such information includes a set of synonyms for an object name, a set of alternative names et al.

3   Structural and Semantic Classified Ad Types for Automated User Interest Matching

The analysis of Internet classified ads has allowed us to reveal the structural and semantic elements of classified ads. The main classified parts are *actions* and *objects*. The complementary elements are *locations* and *additional information*.

The commitment of an *action* with an object (or objects) meets user interest. An object consists of a name and a set of properties and their values representing user interest. *Locations* are descriptions of object positions or places related to the user interest. *Additional information* is the textual description with information that cannot be described using the abovementioned structural elements.

An action may have an additional attribute called an opposite action. This attribute means an action that satisfies the interest of another user. Also, according to semantics, a classified ad has the following elements: *links* and *interest fields*. *Links* represent relations between the actions and objects described in the classified. A link consists of four attributes: an action and an object described in the classified ad; a link type ("demand" or "supply"); interest matching conditions that specify what is necessary for the interests to be matched. The interest matching condition has an additional attribute called an obligation attribute and indicating whether the condition *must* be implemented for the interest matching. An *interest field* is a short textual description of an area related to the classified author interest.

Using this structure we have defined three main classified ad types: classified ads with single-type links, classified ads with multi-type links, classified ads without links.

1.   Classified ads with single-type links. All links in classified ads of this type have the same type. The type consists of subtypes: action – link type; action – opposite action, link type; object – link type.

1.1.   Action – link type. Classified ads of this type are characterized by an action name, which determines the type of the link related to this action unambiguously. A classified ad of this type is illustrated by Example 1.

Note: This example (and all the further examples) does not imply any natural language processing of the classified. The certain structure presented in an example is just a way of representing a classified in the framework of the structural and semantic classified model.

Example 1. An action – link type classified.
I am looking for a big (42" or bigger) lcd/plasma/projection TV. I can pick it up if it is till 100$.

Here the *action* name is "looking for" and the *object* is "TV". We cannot determine the opposite action exactly (it can be "sell" or "give") but the type of *link* between the "looking for" action and the object is "demand". The "TV" object will have such properties as "screen size", "type" and "price". The ability of a customer to pick up a TV will be the content of *additional information* element.

   1.2.   Action – opposite action, link type. Classified ads of this type are characterized by an action name, which determines the opposite action and the type of the link related to this action unambiguously. A classified ad of this type is illustrated by Example 2.

Example 2. An action – opposite action, link type classified ad.
Selling Glass Mosaic Tiles: Copper Link mosaics 13"x13" $8.99 per pc, buy 3 get 1 free. Each sheet is about 13" x 13" (1.15 Sq. ft each sheet) Each small tile is 3/4" x 3/4" x 5 mm thick. The sheets are back mounted on mesh with a joint of 1/8"; it is much easier to be installed than paper mount. This is great for kitchen back splash, wall, floor, bathroom/shower, and swimming pool. We have everything in stock.

Here the *action* name is "selling" and the *object* is "tile". The opposite action for the "selling" action is "buying", and the type of *link* between the action and the object is "supply". The *interest fields* of the author may be, for example, "construction" or "interior design". The *additional information* element will contain the "The sheets are back mounted on mesh with a joint of 1/8"; it is much easier to be installed than paper mount. This is great for kitchen back splash, wall, floor, bathroom/shower, and swimming pool. We have everything in stock" part.

   1.3.   Object – link type. In classified ads of this type the action is not defined, but the type of link related to the object can be determined unambiguously. A classified of this type is illustrated by Example 3.

Example 3. An object – link type classified ad.
Established and well known barber shop, in great condition, and location. It is fully equipped and ready for the right person to take over. This barber shop has been in business since 2000. Has 5 stations, overhead spot lights, etc. Starting price is $57,000 or best offer, owner retiring. Must see, best looking shop in town.

Here the *object* is "barber shop", with properties "condition", "number of stations", "price" etc. The *action* is not specified but we can determine the *link* type related to the object, which is "supply". Such characteristics as "It is fully equipped and ready for the right person to take over" and "best looking shop in town" will be placed in the *additional information*.

   2.   Classified advertisements with multi-type links. Classified ads of this type contain both links of "demand" and "supply" types. This type consists of two subtypes: *action – link type* and *action – opposite action*.

   2.1.   Action – link type. Classified ads of this type contain an action name, which determines the type of link related to this action unambiguously. A classified of this type is illustrated by Example 4.

Example 4. An action – link type classified with multi-type links.
22 years old student is looking for restaurant work in the North Bay preferably in Petaluma or Novato. I would prefer to bartend, prep cook, or line cook. I have about 3 years total in the restaurant industry. I have gone to the Santa Rosa Bartending School

and I am in the process of applying to culinary school. I am honest, trustworthy, quick learner, and a critical thinker when problems arise. The thing I want the most out of a job in the culinary field is to learn as much as I can and continue training.

Here the *action* name is "looking for", and the *objects* are "work" and "student". The "looking for" action is connected to these objects with the *links* of different types: the link for the "work" object has the "demand" type, and the link for the "student" object has the "supply" type. The *opposite action* cannot be determined. The *interest matching conditions*, and *interest fields* are not specified for this classified. The *additional information* contains "The thing I want the most out of a job in the culinary field is to learn as much as I can and continue training". The *locations* in this classified will be the following: "North Bay", "Petaluma", "Novato".

2.2. Action – opposite action. Classified ads of this type contain an action name, which determines the opposite action unambiguously. A classified of this type is illustrated by Example 5.

Example 5. An action – opposite action classified with multi-type links.
Cute smart girl (19) wants to meet funny intelligent boy not older than 23. I'm not a country girl. Sorry country boys. Go find someone else who enjoys that stuff. I don't really know what to say about myself except I like some sports, watching nascar, playing games on the computer, riding horses. I like to try new things. Would like to find someone to do these things with.

Here the *action* name is "meet", and the *objects* are "girl" and "boy". The action is connected to these objects with the *links* of different types, but unlike the previous example the *opposite action* can be determined. The *additional information* is "I'm not a country girl. Sorry country boys. Go find someone else who enjoys that stuff".

3.   Classified advertisements without links. This type contains classified ads consisting of the object description and nothing more, therefore it is impossible to determine actions and links in it. A classified of this type is illustrated by Example 6.

Example 6. A classified without links.
I've launched my brand new 3D city website. Please visit http://ggeometr.narod.ru. Only a web-browser is required.

Here the *object* is "website" with the "URL" and "Requirements" *properties*. This classified ad does not contain any *actions* and *links*.

4   A Classified Ontology for Automated User Interest Matching

The results of structural and semantic analysis of classified ads published in existing online classified advertising systems were used to develop an ontology for user interest matching. It describes all classified ad components, relations between them and defines a possible set of values and (or) data types for these components. Using this ontology a user of the system can form a classified ad according to his/her real knowledge about it. We also use this ontology as a basis for development of methods for automated user interest matching. This report describes the ontology using set-theoretical notation and includes four high-level elements – data types, concepts, domain ontologies and classified ads: *Model = <DataTypes, Concepts, DomainOntologies, Classifieds>*, where *DataTypes* is the data types set, *Concepts* is the

concepts set, *DomainOntologies* is the set of domain ontologies and *Classifieds* is the classified advertisements set. *DataTypes* and *Concepts* sets are not described in this paper due to large size.

4.1 Domain Ontologies

The *DomainOntologies* set consists of the domain ontologies used to describe the objects in a classified and to match user interests at the object level: *DomainOntologies = {DomainOntology$_d$}$_{d=0}^{do\_count}$* , where *do_count* is the number of domain ontologies and *DomainOntology$_d$* is the ontology of a specific domain.

A domain ontology consists of a domain name, domain objects and relations between them: *DomainOntology$_d$ = <DomainName$_d$, DomainObjects$_d$, DomainRelations$_d$>*, where *DomainName$_d$* is the domain name, *DomainObjects$_d$* is a set of domain objects, *DomainRelations$_d$* is a set of relations between domain objects. Domain ontologies are described with OWL (Web Ontology Language), and the domain ontology structure is not examined in this paper.

4.2 Classified Advertisements

A classified ad consists of actions, objects, links between them, interest fields, additional information and location information: Classified = <Actions, Objects, Links, InterestFields, AdditionalInfo, Locations>, where Actions is a set of classified ad actions; Objects is a set of classified ad objects (objects are described by a specific domain ontology); Links is a set of links between actions and objects; InterestFields is a set of interest fields related to the classified; AdditionalInfo is the extra information for the classified; Locations is the set of locations related to the classified.

The set of all classified ads contains all classified ads defined in the system: *Classifieds = {Classified$_i$}$_{i=1}^{advcount}$* , where *advcount* is the number of all classified ads in the system, and *Classified$_i$* is a single classified ad.

4.3 Actions

The classified ad actions set has the following structure: *Actions = {Action$_i$}$_{i=1}^{actioncount}$* , where *actioncount* is the number of actions in the classified, and *Action$_i$* is one of the actions in the classified ad. An action *Action$_i$ = <ActionName$_i$, OppositeAction$_i$>*, where *ActionName$_i$* is an action name, *OppositeAction$_i$* is an opposite action. The action name is a concept: *ActionName$_i$ ∈ Concepts*, where *Concepts* is a set of all concepts. The opposite action is one of the actions defined in the system: *OppositeAction$_i$ ∈ AllActions*, where *AllActions* is a set of all actions defined in the system.

4.4 Objects

Here we describe a simplified structure of classified ad objects for interest matching. Objects themselves are described in specific domain ontologies. The set of classified ad objects has the following structure: *Objects = {Object$_i$}$_{i=1}^{objectcount}$* , where

objectcount is the number of objects in the classified ad, and *Object$_i$* is one of the objects in the classified. An object *Object$_i$* = *<ObjectName$_i$, PropertiesAndValues$_i$>*, where *ObjectName$_i$* is an object name, *PropertiesAndValues$_i$* is a set of object properties and their values. The object name is a concept: *ObjectName$_i$* ∈ *Concepts*, where *Concepts* is a set of all concepts. The set of object properties and their values consists of "property-value" pairs: *PropertiesAndValues$_i$* = *{<ObjectProperty$_{ij}$, ObjectPropertyValue$_{ij}$>}$_{j=1}^{propertycount}$* , where *ObjectProperty$_{ij}$* is a object property, and *ObjectPropertyValue$_{ij}$* is the value for this property.

The object property *ObjectProperty$_{ij}$* consists of a property name and a property data type: *ObjectProperty$_{ij}$* = *<PropertyName$_{ij}$, PropertyDataType$_{ij}$>*, where *PropertyName$_{ij}$* is the property name, and *PropertyDataType$_{ij}$* is the property data type. The property name is a concept: *PropertyName$_{ij}$* ∈ *Concepts,* where *Concepts* is a set of all concepts. The property data type belongs to a set of all data types defined in the system: *PropertyDataType$_{ij}$* ∈ *DataTypes*, where *DataTypes* is a set of all data types defined in the system. The property value belongs to a range defined by a property data type: *ObjectPropertyValue$_{ij}$,* ∈ *PropertyDataType*, where *DataType*s is a range of values defined by the property data type.

4.5 Links

The set of classified ad links has the following structure: *Links* = *{Link$_i$}$_{i=1}^{linkcount}$* , where *linkcount* is the number of links in the classified, and *Link$_i$* is a link between an object and an action described in the classified. The link *Link$_i$* = *<ClassifiedAction$_i$, ClassifiedObject$_i$, LinkType$_i$, MatchConditions$_i$>,* where *ClassifiedAction$_i$* is one of the actions defined in the classified ad, *ClassifiedObject$_i$* is one of the objects defined in the classified ad, *LinkType$_i$* is the link type, *MatchConditions$_i$* corresponds to interest matching conditions. *ClassifiedAction$_i$* ∈ *Actions*, where *Actions* is a set of classified actions; *ClassifiedObject$_i$* ∈ *Objects*, where *Objects* is a set of classified objects; *LinkType$_i$* ∈{*"Demand", "Supply"*}, where *"Demand"* corresponds to "demand" link type, and *"Supply"* corresponds to "supply" link type.

Interest matching conditions have the following structure: *MatchConditions$_i$* = *{MatchCondition$_{ij}$}$_{j=1}^{mccount}$* , where *mccount* is a number of interest matching conditions for the link, and *MatchCondition$_{ij}$* is an interest matching condition. This condition has the following structure: *MatchCondition$_{ij}$* = *<ConditionName$_{ij}$, Obligatory$_{ij}$>*, where *ConditionName$_{ij}$* is the condition content, and *Obligatory$_{ij}$* is the obligation attribute. Condition content is a concept: *ConditionName$_{ij}$* ∈ *Concepts*, where *Concepts* is a set of all concepts. Obligation attribute is a logical value: *Obligatory$_{ij}$* ∈ *Boolean*.

4.6 Interest Fields

The set of interest fields has the following structure: InterestFields = *{InterestField$_i$}$_{i=1}^{interestcount}$* , where *interestcount* is the number of interest fields in the classified ad and *InterestField$_i$* is one of interest fields for the classified ad. Interest field is a concept: *InterestField$_i$* ∈ *Concepts*, where *Concepts* is a set of all concepts.

4.7 Additional Information

The additional information of a classified ad is a string: *AdditionalInfo $\in$ String*.

4.8 Locations

The set of locations in a classified ad has the following structure: *Locations = {Location$_i$}$_{i=1}^{locationcount}$* , where *locationcount* is the number of locations for a classified ad, and *Location$_i$* is one of locations defined in a classified ad. *Location$_i$* has a value of type «Address»: *Location$_i$ $\in$ Address.*

5   Conclusions

The analysis of existing online classified advertising systems and tools for their development has shown that they have various drawbacks. In the Intelligent Systems Laboratory of the Institute of Automation and Control Processes of Far Eastern Branch of Russian Academy of Sciences the new approach to design of an intelligent tool for automated user interest matching was proposed. The approach is based on the structural and semantic classified ad model and intended for eliminating the drawbacks of existing tools.

References

1. ComScore, Inc.: Online Classifieds Increase in Popularity; Category Visitation Surges 47 Percent in the Past Year,
   `http://www.comscore.com/press/release.asp?press=991`
2. Hitwise Pty. Ltd.: Classifieds traffic soars in recession,
   `http://www.hitwise.com/press-center/hitwiseHS2004/`
   `classifieds-traffic-soars.php`
3. Wikipedia: Hierarchical classifier,
   `http://en.wikipedia.org/wiki/Hierarchical_classifier`
4. Vityazev, Y.: Typical web directory problems: problems of classification and automation,
   `http://webdirbasics.contentoffice.ru/article/9/`
5. Wikipedia: Full text search,
   `http://en.wikipedia.org/wiki/Full_text_search`
6. Gribova, V.V., Kachanov, P.S.: Structure-Semantic Classifieds Model for Automated Matching of User Interests in Online Classified Systems. The Bulletin of Computer and Information Technologies 4 (2009)

Buyer Coalitions with Bundles of Items by Using Genetic Algorithm

Laor Boongasame[1] and Anon Sukstrienwong[2]

[1] Department of Computer Engineering,
Bangkok University,
Bangkok, Thailand
`laor.b@bu.ac.th`
[2] Department of Science and Technology,
Bangkok University,
Bangkok, Thailand
`anon.su@bu.ac.th`

Abstract. There are several existing buyer coalition schemes. These schemes do not consider forming a buyer coalition with bundles of items. There is only one scheme that forms a buyer coalition with bundles of items. Nevertheless, the scheme suffers from computational complexity. In this paper, we have applied genetic algorithms (GA) to form buyer coalitions with bundles of items, called the GroupPackageString scheme. The fitness function is defined by total discount of the buyer coalitions over the GA to measures the GroupPackageString scheme. The coalition results show that the total discount of any coalition in this scheme is higher than those in the GroupBuyPackage scheme.

Keywords: group buying, coalition formation, bundles of items, genetic algorithm.

1 Introduction

A buyer coalition is a group of buyers who join together to negotiate with sellers for purchasing identical items at a larger discount [5]. Buyer coalitions have becoming widely important on the internet [5]. One reason is that buyers can also improve their bargaining power and negotiate more advantageously with sellers to purchase goods at a lower price. The other reason, it helps to reduce the cost of communication and makes buyer comfortable in joining a coalition. Buyers can benefit from purchasing the items in large lots or bundles of items through buyer coalitions if the price of the lots or the price of the bundles of items is less than the normal retail price. Additionally, the buyer coalitions may help to reduce cost of stock of the items in case that the items are not produced yet. On the other hand, sellers can benefit from selling the items at large lots or bundles of items[1]

---

[1] Bundles of items refers to the practice of selling two or more goods together in a package at a price that is below the sum of the independent prices [1].

via buyer coalitions if the cost of wholesale marketing (such as the advertising or bidding costs) is less than that of the retail marketing or their market power are preserved or increased.

Many existing buyer coalition schemes [2,3,5,6,8,9,10,11,12,13,14,15] are proposed. Nevertheless, these schemes do not consider forming a buyer coalition with bundles of items. This practice can be frequently noticed in the real world such as durable consumer goods (e.g., personal computer options), and nondurable consumer goods (e.g., dishwasher detergent and rinse aid packages). There is only one scheme, GroupBuyPackage scheme [4], that forms a buyer coalition with bundles of items. However, the scheme meet with computational complexity. Specifically, when the number of buyers increases, the run time of these algorithms increases drastically.

In this paper, we use a genetic algorithm approach for forming a buyer coalition with bundle of items, called the GrouPackageString scheme, which aims at maximizing the total discount. Of course, there is no guarantee for convergence of genetic algorithms but our experimental results have shown that in most cases the result of the algorithm is acceptable. The rest of the paper is organized as follows. Section 2 compares related works. Section 3 is an important part of the full paper, the design of our GroupPackageString scheme. Section 4 presents the simulation and experimental results before we discuss the conclusions with our future works in Section 5.

2   Related Works

Several existing buyer coalition schemes are proposed. The GroupBuyAuction [15] scheme forms a buyer coalition with item categories. The Combinatorial Coalition Formation scheme [11] looks on an e-market where buyers bid on item combinations with reservation prices while sellers offer discounts for each item based on volume. Hyodo, Matsuo, and Ito [6] optimally allocate buyers to several group buying sites that are selling similar (or the same) goods by a genetic algorithm. In Matsuo, Ito, and Shintani [13] scheme, buyers with multi-attribute preferences (utilities) are integrated into a coalition and the system supports buyers' decision-making by using an analytic hierarchy process. In He and Ioerger [2] scheme, individual buyer wants to purchase different goods as bundle, while sellers offer discounts based on the total cost of goods sold in one transaction. They propose some purchasing strategies that give the maximized discounts to buyers. In addition, a buyer coalition scheme [14] considers e-market where each buyer places a bid for all possible sets of items with reservation prices, while sellers offer discounts based on volume of each item. All of the research above do not consider forming buyer coalitions with bundles of items. GroupBuyPackage scheme [4] is only one existing scheme that forms a buyer coalition with bundles of items. Nevertheless, computation of this scheme is still complexity.

3   Forming Buyer Coalitions with Bundles of Items by Using Genetic Algorithm

3.1   Example

The purchasing office of a parent company wants to get a discount on purchasing bundles of items for all subsidiaries in the parent company. Therefore, when any subsidiaries in the parent company orders the items within the bundle of items, the purchasing office will gather these orders to purchase the bundles of items with a larger discount. The price list for these computer packages is shown in Table 1. Additionally, the price of these packages is set by simple rule as following. The average price of any item to a package is high if the package is a single item. For instance, the Package Number 1, 2, 3 and 4, called p_1, p_2, p_3 and p_4, are a single item package. These packages cost \$1000, while the package p_5 consisting of a unit of printer and CPU is sold at the price of \$1950. For this particular package, the average price of each item is $1950/2 = 975$ which is, obviously, cheaper than p_1 and p_2. The subsidiaries or buyers in the company cannot buy the bundles of items by themselves. The maximum price that these buyers are willing to pay for a unit of each item, called reservation price rs, as shown in Table 2. For instance, the buyer A requests to purchase a unit of printer at the maximum price of \$990.

3.2   Problem Formulation

To form the coalition for buyers with bundles of items, the GroupPackageString is defined. The GroupPackageString is a fixed-length character string with the size of L. If a seller has made a set of m packages, $\{p_1, p_2, ..., p_m\}$, then the

Table 1. The price list for particular computer packages

Package Number	Printer	CPU	Monitor	RAM	Unit Price (\$)
p_1	✓	-	-	-	1000
p_2	-	✓	-	-	1000
p_3	-	-	✓	-	1000
p_4	-	-	-	✓	1000
p_5	✓	✓	-	-	1950
p_6	-	✓	-	✓	1850
p_7	✓	-	✓	✓	2750

Table 2. Subsidiaries' required computer equipments and their reservation prices

Subsidiaries	rs of Printer (\$)	rs of CPU (\$)	rs of Monitor (\$)	rs of RAM (\$)
A	990	-	-	-
B	-	920	925	-
C	925	930	-	925
D	930	-	930	930

GroupPackageString $P_{seq} = \{p_j\}^L$ where seq is order of the GroupPackageString generated by GA and $p_j \in \{p_1, p_2, ..., p_m\}$. Any package p_j, where $0 \leq j \leq m$ is comprise of items in the $Gd = \{gd_1, gd_2, ..., gd_q\}$ and its package price $Price_j$. Additionally, given a set of buyers $B = \{b_1, b_2, ..., b_k\}$ in a community, each buyer $b_i \in B$ wants to purchase only one unit of each item within a set of $Gd = \{gd_1, gd_2, ..., gd_q\}$. Each buyer $b_i \in B$ places a set of reservation prices $Rs_i = \{rs_1^i, rs_2^i, ..., rs_q^i\}$. A buyer can place only a set of seal bids to an honest coordinator H. Different buyers generally want different items and have different bids. A coalition C is a subset of buyers $C \subseteq B$ who want to join together to purchase the bundles of items p_j with a larger discount. The coalition C has a set of items Gd_C which is the sum of the items that the coalition buyers require. The problem that we solve is forming buyer coalitions $C \subseteq B$ for purchasing many packages of bundles of items such that $\forall b_i \in B$ can purchase required items and $\sum_i V_i$ is maximal where V_i is the joint utility that members of C can reach by cooperating via coalitional activity for purchasing a specific bundles of items. Formally, the utility V of a coalition C in purchasing item packages p_j is defined as $V_{cl} = \sum rs_h^i$ - $Price_l$, called PackageEanedPrice.

3.3 Genetic Algorithm (GA)

In general, the genetic algorithm begins with generation 0 (Gen := 0) with the population size M which is randomly created. There exist three operations to perform during one generation, reproduction, crossover, and mutaion. During the run, a given individual might be mutated and reproduced and reproduced and crossed with in single generation. GA searches the space of possible character strings in an attempt to find good string based on fitness value of strings. There are three steps in preparing to use the genetic algorithm [7].

1. determining the representation scheme: Selecting a representation that facilitates solution of the problem by means of the genetic algorithm.
2. determining the fitness measure: In general, the fitness measure is inherent in the character string it encounters. The fitness measure assigns a fitness value to each possible member in the population.
3. determining the steps of algorithm: Once representation scheme and fitness function have been competed, the genetic algorithm can be run. Three are three substeps in executing our simplified genetic algorithm operating on fixed-length character strings can be summarized as follows:
   - Randomly create an initial population of individual fixed-length character strings.
   - Create a new population for the next generation based on fitness value by randomly applying two operators, crossover operator and mutation operator together with reproduction operator.
   - The best individual based on the fitness function in any generation is designated as the result of the genetic algorithm. The result may represent approximate solution to the problem.

The entire complexity of this algorithm is $O(n)$ computations, where n is the number of buyers in B.

3.4   Example Revisited

The algorithm in this research will be described through the example as follows:

1. determining the representation scheme
   For this problem, the representation scheme in the algorithm is a mapping that expresses each possible GroupPackageString in the search space of the problem as fixed-length character string. Specification of the representation scheme requires selecting the string length m. In this problem, the set of alphabets of this string is $\{p_0, p_1, p_2, ..., p_m\}$. Each individual in the population belongs to $(m + 1)^L$.
2. determining the fitness measure
   The fitness function is defined by total discount of the buyer coalitions over the GA to measure the profit from the GroupPackageString. Execution of the genetic algorithm begins with an effort to learn the problem by testing a number of randomly selected GroupPackageString in the search space. For this particular problem, the fitness function of GroupPackageString is the PackageEanedPrice.
3. determining the steps of algorithm
   - Randomly create an initial population of individual fixed-length character strings. All initial parents of the population are GroupPackageString randomly selected according to their fitness value. Each individual will be evaluated in order to satisfy all of the buyers' requirements. Suppose four initial random members of population, with the of size $L <= 5$, are as shown below. These GroupPackageString are $P_1 = p_2 p_7 p_7 p_5$, $P_2 = p_1 p_5 p_6 p_7 p_3$, and $P_3 = p_1 p_3 p_7 p_6 p_1 p_2$. The first package of Group-PackageString P_1 is p_2 which has only one unit of CPU. The algorithm will find the buyer with the highest reservation price of CPU from Table 2, which has not been choosen for this item. we can see that the buyer C has made the hight reservation price of \$930 to CPU. If the package is comprise more than one item, the algorithm will map the rest of the item to the rest of buyers in the group who have made best reservation price. The difference of the sum of reservation price and package's price is -70 which will be called PackageEanedPrice. As shown in Table 3, the sum of PackageEanedPrics P_1 is -45. This value will be assigned to be a fitness value of P_1. As the same way, the fitness value of the P_2 and P_3 are -145 and -195, respectively.
   - Create a new population for the next generation based on fitness value by randomly applying two operators, crossover operator and mutation operator together with reproduction operator. Suppose that the simulation randomly chooses crossover operator to perform for creating generation 1. The crossover operation is applied to individual GroupPackageString in the population with a specific control of frequencies, P_c. After that, the individual of population will be copied to next population based on its fitness value. At this point, the adjustment is needed to apply to individual GroupPackageString, if its fitness value is negative. The best-of-generation is P_1 which has the probability of $(-45 + 195)/((-45 + 195) +$

Table 3. Observe values of the fitness of PackagesStrings P_1

GroupPackageString	Package No.	Printer	CPU	Monitor	RAM	PackageEanedPrice Individual package	Sum
P_1	p_2	-	930	-	-	930-1000=-70	-70
	p_7	990	-	930	930	2790-2750=100	30
	p_7	930	-	925	925	2780-2750=30	60
	p_5	925	920	-	-	1840-1950=-105	-45

Table 4. The resulting of the crossover operation at random crossover point $= 2$

Parents	Crossover fragments and remainder flagmeng after spliting	Offspring produced by crossover
1 $p_2p_7p_7p_5$	$p_2p_7 - -, - - p_7p_5$	$p_2p_7p_6p_7p_3$
2 $p_1p_5p_6p_7p_3$	$p_1p_5 - -, - - p_6p_7p_3$	$p_1p_5p_7p_5$

$(-145+195)+(-195+195)) = 0.75$. Thus, it will be appeared 75 percents in the new population. Similarly, the individual P_2 has the probability of 25 percents, and the worst-of-generation of the experiment is P_3 with the 0 percents of the probability of being selected for participating in the next processing. The resulting of population is called the mating pool.

After create mating pool, the crossover operation randomly chooses two existing parents in the mating pool, and also randomly selects the position of a string in the GroupPackageString. Suppose that the position point is 2. Each parent will be split at this position into two substrings, a crossover fragment and the remainder. The two random parents in the mating pool selected in proportion to the fitness value are P_1 and P_2. The offspring produced by the crossover operation are $p_2p_7p_6p_7p_3$ and $p_1p_5p_7p_5$, as shown in the Table 4. These new GroupPackageString can be added into the next generation, if their fitness value is better than the current population of generation. For this paper, the first operations, reproduction operation, is applied to individual GroupPackageString with $P_r = 100$all of the population in the current generation will be automatically transform to the next generation.

For creating generation 2, suppose that the simulation randomly chooses the mutation operator. Originally, the mutation operator is controlled by a parameter called the mutation probability, P_m. This probability is calculated by 1-P_c. The simulation begins by randomly a GroupPackageString from the current population of generation 1, and then randomly selects a number as the mutation point for GroupPackageString. Then, a single character at the position point of the selecting GroupPackageString is changed. The single character can be changed to be any packages in the GroupPackageString, $\{p_1, p_2, ..., p_7\}$. Suppose that the random selecting GroupPackageString in the mating pool is P_2, and the random number is 5. The possible result would have been

the new GroupPackageString of $p_1p_5p_6p_7p_7$. This new offspring is also required to calculate the fitness value.

- The best individual based on the fitness function in any generation is designated as the result of the genetic algorithm. The result may represent approximate solution to the problem. At the end of each generation, the reproduction operator will be applied to all current members and new offspring besed on the fitness value. The genetic algorithm keeps processing in the same steps finding new population for each generation until it gets maximum number of generation to be run. The best Group-PackageString which are possible to yield the highest profit to the group buyer might be found.

4 Simulation

In this section, we have used the simulation to show the performance of the proposed GA-based algorithm.

4.1 GA-Based Program Setup

We have implemented more than 3000 lines of C++ program to simulate all of the proposed algorithms. All simulation experiments are run on a Pentium(R) D CPU 2.80 GHz, 2 GB of RAM, IBM PC. We have tried several of run with different values of the population size (M), mutation probability P_m, and crossover probability P_c, to find which values would steer the search towards the best solution. The default parameters were varied and the results collected from test runs were used to study the effects of changing these parameters. Table 5 summarizes the simulation parameters.

4.2 Experimental Results and Analysis

In order to get best experimental results, we have tried several of run to find the parameter settings of our algorithm. We discover from the experimental results

Table 5. Summarize the control of simulation parameters for the genetic algorithm

Constant	Detail	Range
NumOfBuyer	The size of the group buyer	10-20
NumOfPackage	The total number of the packages made by a seller	15
NumOfItem	The number of items	4
Gen	The Maximum number of generation to be run	50-300
AIS	The Average items per package of the seller	1-2
M	The population size of each generation	300
P_c	The crossover probability	0.95
P_m	The mutation probability	0.05
P_r	The reproduction probability	1.0
GroupPackageStringLength	The fixed length of the GroupPackageString	20

when the crossover probability P_c is 0.95 and mutation probability P_m is 0.05, the convergent rate is fast.

Changing the number of generation (Gen)

Average total discount of any coalition comparison while increasing the number of generation is shown in Figure 1. The horizontal axis of this graph is the number of generations. The vertical axis of this graph is average total discount of any coalition. Brief justifications for the values of these figures used are given below. These values were: M = 300, P_c = 0.95, P_m = 0.05, GroupPackageStringLength = 20, NumOfPackage = 15, NumOfBuyer = 20. From Figure 1, while the number of generations was increased, average total discount was increased. When the number of generation was 300, the curve of the graph of our scheme was higher than the resulting of GroupBuyPackage scheme. It is because the quality of the formed buyer coalition is improved after each generation.
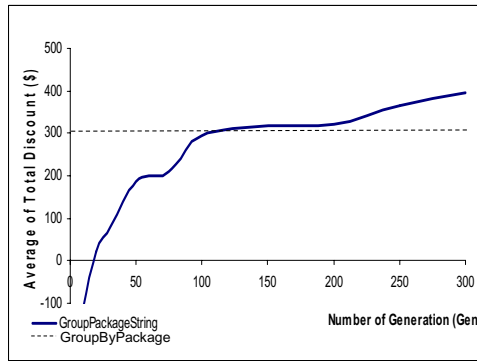
Fig. 1. Average total discount of any coalition while increasing the number of generation

Changing the population size (M)

Average total discount of any coalition comparison while increasing the population size is shown in Figure 2. The horizontal axis of this graph is the population size. The vertical axis of this graph is the average total discount of any coalition. Brief justifications for the values of these figures used are given below. These values were: M = 300, P_c = 0.95, P_m = 0.05, GroupPackageStringLength = 20, NumOfPackage = 15, NumOfBuyer = 20. From Figure 2, while the population size was increased, average total discount was increased. When the number of the population size was set to be over 100, the curve of the graph of our scheme was higher than the resulting of GroupBuyPackage scheme. It is because the number of parents of the population was increased; therefore better offspring of the next generation may be generated. It might be possible to lead to the optimal outcome.
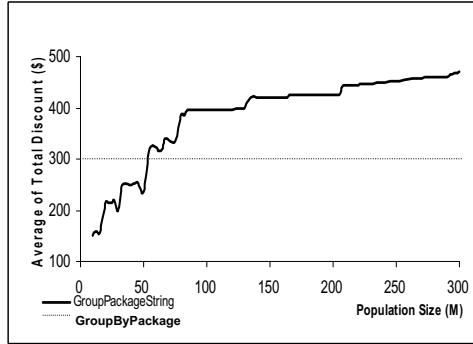
Fig. 2. Average total discount of any coalition comparison while increasing the population size

Changing the buyers' order

In order to evaluate a buyer coalition in this algorithm, we have set three tests of the experiment with different number of buyers. The parameter settings for GA based simulation is shown in Table 6. Each test operates 10 runs with specific packages made by the seller. Moreover, for each test, we have set the buyer's reservation price of each item at the retail price which is \$1000, the highest price of each item. The parameters for our scheme are: Generation (Gen) = 300, crossover probability P_c = 0.95, and mutation probability P_m = 0.05.

From Table 7, the experimental results show that when the the average items per package of seller (AIS) is high, the average total discount will be obviously high. It is because higher AIS might be able to lead less effort for the group buyer to form the coalition. In Figure 3, the horizontal axis of this graph is standard deviation (SD) of buyer's order, which indicates the characteristic of group's orders. If SD is high, it means that most of buyers tend to buy very close

Table 6. Parameter settings for GA based simulation

No. of test	NumOfBuyer	NumOfPackage	NumOfItem	Total number of Items required from buyers
Test 1	10	15	4	20
Test 2	20	12	4	20
Test 3	20	10	4	20

Table 7. Experimental results of four tests for GA based simulation

No. of test	Average of best discount (10 runs)	Average of worst discount	Average total discount	AIS	Average required items per buyer
Test 1	2500	0	1092.89	2.00	2
Test 2	1500	-46.25	756.13	1.92	1
Test 3	1450	-100.00	609.51	1.80	1
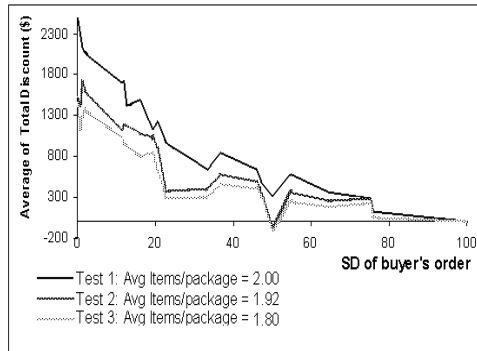
Fig. 3. Average total discount of any coalition comparison while increasing SD
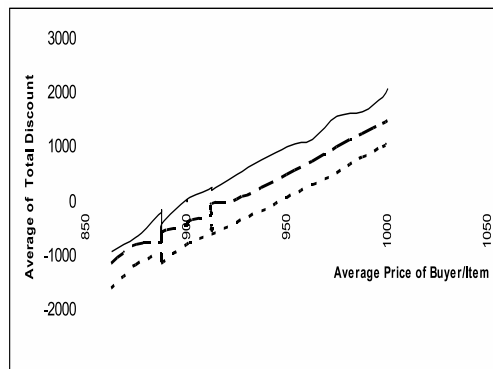
Fig. 4. Average total discount of any coalition comparison while increasing average price of buyer/item

to the few items. For instance, if all buyers of the group want to purchase only printer, the SD of the group is 100. On the other hand, when buyers have made their requests with the low SD, the total numbers of each type of item in the package are very much the same. For instance, if the total numbers of printer, CPU, monitor, and RAM are 5:4:4:7, the SD of group is 2. The vertical axis of this graph is average total discount of any coalition. From Figure 3, we found that all of the tests give the largest total discount when the SD of buyer's order is low. It is because buyer's order is spread equally in the items of the package, which is easily for GA to form the best solutions.

From Figure 3, the low SD of group's order obviously yields the good discount to the group buyer. So, we arbitary choose the SD of group = 2 for Figure 4. In Figure 4, the horizontal axis of this graph is average total discount. The vertical axis of this graph is average price of buyer/item. As the resulting, it shows that when the average price of buyer per item is high, the average total discount is

high. It is true for every value of AIS. Moreover, the GroupPackageString scheme can form the group with some profit at the point below the average price of item of seller.

5   Conclusions and Further Work

This paper presents the GA-based algorithm for forming a buyer coalition with bundles of items. This algorithm is suitable for cases where individual buyer in a coalition cannot purchase the bundles of items by themselves and they want to get maximum total discounts from forming a buyer coalition. Nevertheless, they may not want to form the coalition if the different between discounts from purchasing a bundle of items and those from buying the items individually is low. To evaluate a buyer coalition in this scheme, we compare the results of this scheme with those of the GroupBuyPackages scheme. The complexity of this algorithm ($O(n)$) is less than that of the algorithm in the GroupBuyPackage scheme ($O(2^n)$). From Figure 1 and Figure 2, we found that the total discount of any coalition in this scheme is higher than that in the GroupBuyPackage scheme, especially, in case that the number of generations was increased and the population size was increased average total discount was increased. Additionally, we found the experimental results as follows 1) from Table 7, when the AIS is high, the average total discount will be obviously high 2) from Figure 3, all of the tests give the largest total discount when the SD of buyer's order is low and 3) from Figure 4, when the average price of buyer per item is high, the average total discount is high.

This scheme has restrictions of forming a buyer coalition as follows 1) only the price attribute is considered in forming a buyer coalition 2) each buyer can buy only one unit of each item 3) unlimited units of the item are supplied by one seller 4) the period of time for forming the buyer coalition is not limited and 5) the package in this scheme is mixed bundling in which the bundle of items are also sold separately. These constraints can be extended to investigate in future research.

References

1. Gurler, U., Oztop, S., Sen, A.: Optimal Bundle Formation and Pricing of two Products with Limited Stock. J. International Journal of Production Economics (2008)
2. He, L., Ioerger, T.: Combining Bundle Search with Buyer Coalition Formation in Electronic Markets: A Distributed Approach through Explicit Negotiation. J. Electronic Commerce Research and Applications 4, 329–344 (2005)
3. Anand, K.S., Aron, R.: Group Buying on the Web: A Comparison of Price-Discovery Mechanisms. J. Management Science 49, 1546–1562 (2003)
4. Laor, B., Leung, H.F., Boonjing, V., Dickson, K.W.: Forming Buyer Coalitions with Bundles of Items. In: Nguyen, N.T., Hakansson, A., Hartung, R., Howlett, R., Jain, L.C. (eds.) KES-AMSTA 2009. LNCS (LNAI), vol. 5559, pp. 121–138. Springer, Heidelberg (2009)

5. Tsvetovat, M., Sycara, K., Chen, Y., Ying, J.: Customer coalitions in electronic markets. In: Dignum, F.P.M., Cortés, U. (eds.) AMEC 2000. LNCS (LNAI), vol. 2003, pp. 121–138. Springer, Heidelberg (2001)
6. Hyodo, M., Matsuo, T., Ito, T.: An Optimal Coalition Formation among Buyer Agents based on a Genetic Algorithm. In: Chung, P.W.H., Hinde, C.J., Ali, M. (eds.) IEA/AIE 2003. LNCS (LNAI), vol. 2718, pp. 151–157. Springer, Heidelberg (2003)
7. Koza, J.: Genetic Programming on the Programming of Computers by Means of Natural Selection. MIT press, Cambridge (1992)
8. Indrawan, M., Kijthaweesinpoon, T., Srinivasan, B., Sajeev, A.: Coalition Formation Protocol for E-Commernce. In: Proceedings of the International Conference on Intelligent Sensing and Information Processing, pp. 403–407 (2004)
9. Kraus, S., Shehory, O., Tasse, G.: Coalition Formation with Uncertain Heterogeneous Information. In: Proceedings of the Second international Joint Conference on Autonomous Agents and Multiagent Systems, pp. 1–8 (2003)
10. Kraus, S., Shehory, O., Taase, G.: The Advantages of Compromising in Coalition Formation with Incomplete Information. In: Proceedings of the Third international Joint Conference on Autonomous Agents and Multiagent Systems, New York, pp. 588–595 (2004)
11. Li, C., Sycara, K.: Algorithm for Combinatorial Coalition Formation and Payoff Division in an Electronic Marketplace. In: Proceedings of the First International Joint Conference on Autonomous Agents and Multiagent Systems, pp. 120–127 (2002)
12. Li, C., Rajan, U., Chawla, S., Sycara, K.: Mechanisms for Coalition Formation and Cost Sharing in an Electronic Marketplace. In: Proceedings of the 5th International Conference on Electronic Commerece, pp. 68–77 (2003)
13. Matsuo, T., Ito, T., Shintani, T.: A Buyers Integration Support System in Group Buying. In: Proceedings of the IEEE international Conference on E-Commerce Technology, pp. 111–118 (2004)
14. Matsuo, T., Ito, T., Shintani, T.: A Volume Discount-based Allocation Mechanism in Group Buying. In: Proceedings of the 2005 international workshop on data engineering issues in e-commerce, pp. 59–67 (2005)
15. Yamamoto, J., Sycara, K.: A Stable and Efficient Buyer Coalition Formation Scheme for E-Marketplaces. In: Proceedings of the 5th International Conference on Autonomous Agents, pp. 576–583 (2001)

Energy Efficient MAC Length Determination Method for Statistical En-Route Filtering Using Fuzzy Logic*

Hyeon Myeong Choi and Tae Ho Cho

School of Information and Communication Engineering, Sungkyunkwan University,
Suwon 440-746, Republic of Korea
{hmchoi,taecho}@ece.skku.ac.kr

Abstract. In wireless sensor networks (WSNs) individual sensor nodes are subject to security compromises. An adversary can use compromised sensor nodes to inject false reports into the WSN. If undetected, these false reports are forwarded to the base station. Such attacks by compromised sensor nodes can not only result in false alarms but also depletion of the limited amount of energy in battery powered sensor nodes. The statistical en-routing filtering (SEF) scheme can detect and drop false reports during the forwarding process. In SEF, the number of the message authentication codes (MAC length) is important for detecting false reports and saving energy in network. In this paper, we present a fuzzy-based MAC length determination method for SEF. If there are fewer nodes surrounding the occurred event in the field in the network than the MAC length, the node cannot generate a legitimate report in SEF. The fuzzy-based method can prevent this problem and provide energy savings. We evaluated the proposed method's performance via simulation.

Keywords: Wireless sensor networks, SEF, Injected false report attack, Fuzzy, Filtering scheme.

1 Introduction

Recent advances in micro-electro-mechanical system (MEMS) technology, wireless communications, and digital electronics have enabled the development of low-cost, low-power, multifunctional sensor nodes that are small in size and use untethered communication over short distances [1]. Sensor networks consist of a large number of sensor nodes that monitor the environment, and a few base stations that collect the sensor readings [10]. Sensor networks may be deployed in hostile environments and potential threats can range from accidental node failures to intentional tampering. Due to their relatively small size and unattended operations, sensor nodes are at high risk of being captured and compromised [2]. If sensor nodes are physically captured and compromised, the security information such as the keys used to generate message authentication codes (MACs), will be revealed to the adversary. False sensing reports can be injected into sensor networks by adversaries via compromised nodes. These

---

false reports can not only result in false alarms but also depletion of the finite amount of energy in sensor nodes [3].

To combat false reports injected via compromised nodes, several research have proposed mechanisms [3-9, 11]. Fan Ye et al. [3] proposed statistical en-route filtering (SEF) as a means to detect and drop injected false reports during the forwarding process. SEF carefully limits the amount of security information assigned to each node, in order to prevent any single compromised node from disrupting the entire system. It relies on the collective decisions of multiple sensors for false report detection. When an event occurs in the field, multiple surrounding sensors collectively generate a legitimate report and endorse it by attaching to it their MACs. Thus a report with an inadequate number of MACs, viz., MAC length is dropped during the forwarding process by intermediate nodes. As a report is forwarded via multiple hops toward the base station (BS), each intermediate node verifies the correctness of the MACs carried in the report with a certain probability, and drops the report if an incorrect MAC is detected. Therefore the MAC length is very important in SEF, since it involves a trade-off of security against communication overhead. The probability of detecting incorrect MACs increases according to the number of hops the report travels.

In this paper, we propose an energy efficient MAC length determination method for SEF using fuzzy logic. In the proposed method, the BS determines the MAC length based on the node density, the number of hops traveled by the reports, and the energy of nodes. Thus the proposed method prevents the problem in SEF whereby nodes cannot generate legitimate reports, when there are fewer nodes sensing the same event than the MAC length.

The rest of the paper is organized as follows: Section 2 briefly describes SEF as background knowledge. Section 3 details the proposed method. Section 4 reviews the simulation results. Finally, section 5 concludes the paper.

2   Statistical En-Route Filtering Scheme (SEF)

Fan Ye et al. proposed SEF [3] as a means to detect and drop injected false reports. In SEF, each legitimate report carries multiple MACs generated by different nodes detecting the same event. As a report is forwarded via multiple hops toward the BS, intermediate forwarding nodes detect incorrect MACs and drop the false report en-route. The BS verifies the correctness of each MAC and eliminates any remaining false reports that eluded en-route filtering. Before the nodes are deployed, the BS maintains a global key pool and each sensor node stores a small number of keys that are randomly chosen from the global key pool in order to endorse and verify the reports. When real events occur in the sensing field, multiple surrounding nodes can detect the event and a center of stimulus (CoS) node is elected in order to generate the report. Each detecting sensor node endorses the report by producing a keyed MAC via one of the stored keys and sends a message that includes the event information and key index, and the generated MAC to the CoS node. The CoS collects messages from the nodes and attaches them to the report. However, when no real event occurs, normal sensor nodes do not detect events. Thus they do not send a message to the CoS node. Therefore, compromised nodes are forced to forge MACs in order to generate a false report, which is dropped by en-route filtering or discarded by the BS. The

number of MACs (i.e. the MAC length) attached to the report is important, because if the MAC length is small, an adversary can generate false reports using only few forged MACs. On the other hand, if the MAC length is too big, the overhead of each report is increased.

2.1 Key Assignment and Report Generation

The BS maintains a global key pool, and divides it into n non-overlapping partitions. Each partition has m keys, and each key has a unique key index. Before the nodes are deployed, the user randomly selects one of n partitions, and randomly chooses k keys from it. The chosen keys and the associated key indices are stored in the node(Fig. 1).
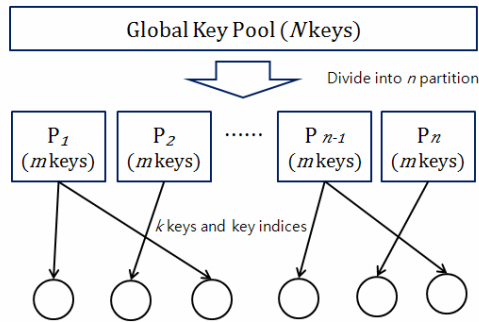
Fig. 1. The global key pool is divided into n partitions, and k keys chosen from a randomly selected partition and the associated key indices are stored in the node

When an event occurs in the sensing field, all surrounding nodes detect the event and one of these nodes is elected to the CoS node. Detecting nodes generate the MAC via one of keys, and send it along with the key index to the CoS node. The CoS node collects and classifies MACs based on the key partitions. Then the CoS generates a report consisting of the event information and the set of key indexes and MACs chosen from distinct partitions. If the MAC length is T, each report carries exactly T sets of key indexes and MACs. A report with fewer T MACs or key indices or more than one key index in the same partition is dropped.

The MAC length involves a trade-off between the detection power and overhead. A larger MAC length makes it difficult to generate false reports, but the communication overhead is increased. In an area with low a node density, there can be fewer distinct partition nodes than the MAC length. In this case the CoS node cannot generate a legitimate report. In the proposed method, we carefully determine the MAC length via fuzzy logic in order to complement this case.

2.2 En-Route Filtering

As a report is forwarded multiple hops, each intermediate forwarding node can verify it(Fig. 2). The reports are verified in intermediate nodes via the following operations:

1. Check that the number of sets of MACs and key indexes are exactly the same as the MAC length. If the report has a different number of sets than the MAC length, the node drops the report.
2. Check that the key indices in the attached report belong to distinct partitions. If the report has more than one key index in the same partition, the node drops the report.
3. If the node has a key that is the same as that of a report, the node generates a MAC with that key. If the generated MAC and the corresponding MAC are different, the node drops the report.
4. If the report does not match any of cases 1-3, the node sends the report to the next hop.

SEF can drop the false report en-route, but if there are no detected false reports arriving at the BS, the BS can check the correctness of every MAC and it can detect a forged MAC, because it has all keys. If there are one or more mismatches, the BS discards the report.
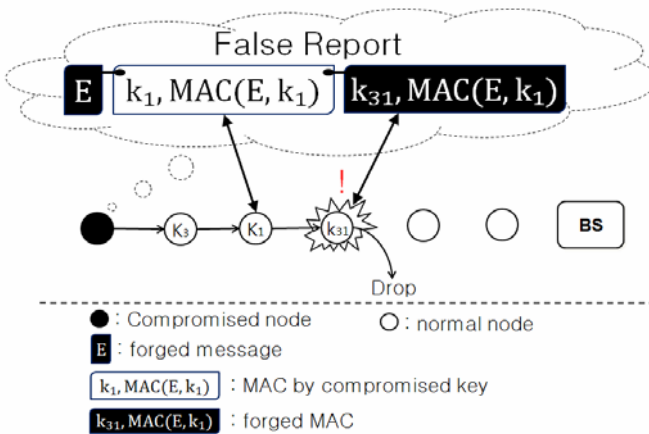
Fig. 2. En-route filtering

3   MAC Length Determination Method Using Fuzzy Logic

We propose the MAC length determination method in order to complement SEF and save energy of sensor nodes via fuzzy logic. In this section, we provide the basic assumptions and motivations of our method and its detailed operation.

3.1   Motivations and Assumptions

In SEF [3], when there are occurred events in the sensing field, they are detected by all surrounding sensor nodes. Then each node generates MAC via one of its keys, and sends the set of MAC and key index that are used to generate MAC to the CoS node. The CoS randomly chooses received sets from distinct partitions. Then it generates a legitimate report consisting of event information and sets of key indexes and MACs.

The MAC length, that is the number of attached sets of key indexes and MACs in legitimate report, is important, because it involves a trade-off of the probability of detecting false reports and the communication overhead. Moreover, if the event occurs in an area of low node density, the CoS node may not be able to collect enough MACs, so it cannot generate a legitimate report of an event. Therefore we carefully determine the MAC length based on the node density and the hop count of the report and the remaining energy of nodes.

We consider a sensor network composed of a large number of small sensor nodes and a BS. We also assume that the reports are forwarded via multiple hops toward the BS and that the BS should know or estimate the density of the nodes, the average number of hops traveled by reports, and the node's energy level. Since SEF relies on the MAC length carried MACs to detect false reports, an attacker that compromises keys in MAC length or more distinct partitions can successfully fabricate reports. In this case, SEF cannot detect or drop such false reports [3], and the proposed method does not address this case.

3.2 Overview

The proposed method is based on SEF. It consists of the same operations involved in partition dividing and key assignment before the nodes are deployed. However the SEF determines the fixed MAC length before deployment, whereas the proposed method determines it after the nodes have been deployed, based on the density of nodes, the number of hops traveled by reports and the nodes' energy level.

Fig. 3 shows the MAC length determination process. The sensor nodes are randomly deployed in the sensing field. The BS estimates the density of nodes, the average number of hops traveled by reports, and the remaining energy level. The estimation of each factor is beyond the scope of this paper. Therefore, we assume that the BS should know the factors. The BS determines the MAC length via fuzzy logic, based on the density, hop count and energy (a). Then the BS broadcasts the MAC length to all nodes (b).

The BS determines the MAC length every time the network status changes, based on the density of nodes, the average number of hops traveled by reports and the energy level, and then rebroadcasts the MAC length. Thus the proposed method is more flexible than the SEF.
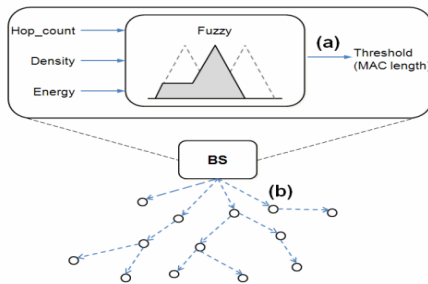
Fig. 3. The MAC length determination process (a) The BS determines the MAC length based on fuzzy logic. (b) The BS broadcasts the MAC length to the nodes.

3.3 Factors

In this section, we discuss the factors that are used for fuzzy inference.

- Hop count (i.e., number of hops traveled by reports): As reports are generated in the CoS node, they are forwarded via multiple hops toward the BS and the communication overhead is increased according to the MAC length. In the case of a large hop count, if the MAC length is large, more energy is consumed in transmitting but false reports are detected in early hops. If the MAC length is small, the overhead is decreased but the probability of detecting false reports is also decreased, thus, the nodes waste energy in transmitting false reports. Therefore, the MAC length is carefully determined based on the hop count.

- The density of nodes: As events occur in the sensing field, the CoS node generates a legitimate report. The legitimate report consists of event information, sets of MAC and key index. The number of sets including the report is exactly same as MAC length. But if the number of nodes detecting the same event is too small to generate a legitimate report, the CoS node cannot generate a legitimate report and the real event cannot be forwarded. Thus, determination of MAC length must be based on the density of nodes.

- The remaining energy of nodes: Generally, sensor nodes expend their limited energy in sensing events, performing computations and transmitting information in a wireless environment [1]. Thus, the energy is the most important resource that should be considered in sensor networks.

3.4 Fuzzy Membership Functions and Rules

Fig. 4(a), (b), (c) illustrate the membership functions of fuzzy logic input parameters. The labels of the fuzzy variables are represented as follows:

- Hop count = { S(Small), M(Medium), L(Large) }
- Density = { VL(Very Low), L(Low), M(Medium), H(High), VH(Very High) }
- Energy = { S(Small), N(Normal), L(Large) }
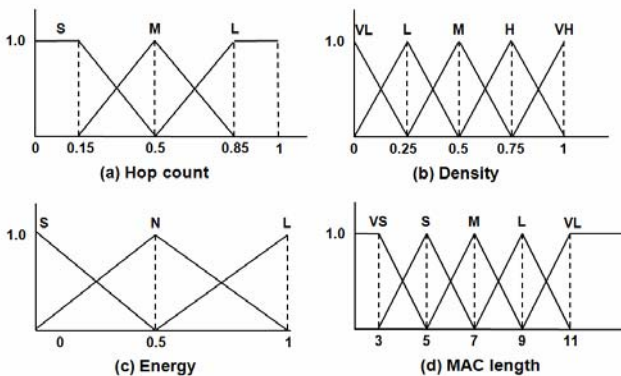
Fig. 4. Fuzzy membership functions

Fig. 4(d) is the membership function of fuzzy logic output parameters. The labels are presented as follows:

- MAC length = {VS(Very Small), S(Small), M(Medium), L(Large), VL(Very Large)

Examples of fuzzy rules are as follows:

- RULE 8: IF (Hop_count IS S) AND (Density IS VL) AND (Energy IS S) THEN (MAC_length IS VS)
- RULE 13: IF (Hop_count IS M) AND (Density IS L) AND (Energy IS M) THEN (MAC_length IS S)
- RULE 19: IF (Hop_count IS L) AND (Density IS M) AND (Energy IS M) THEN (MAC_length IS M)
- RULE 39: IF (Hop_count IS M) AND (Density IS VH) AND (Energy IS L) THEN (MAC_length IS VL)

In our method, we defined 45 fuzzy rules. If Hop count is S, Density is VL, and Energy is S, then we may decrease MAC length to conserve energy (Rule 8). Forwarding nodes do not need to verify such report. They may be detected by the BS immediately. If Hop count is L, Density is M, and Energy is M, then the MAC length is M (Rule 19). In this case, forwarding nodes should detect false reports in earlier stage. Therefore we may increase the MAC length to detect false reports.

4   Simulation Results

To show the effectiveness of proposed method, we compare the proposed method with the original SEF that has a fixed MAC length via simulation. Each node consumes 16.25μJ and 12.5μJ to transmit and receive a byte, respectively. Each MAC generation consumes 15μJ. The size of an original report and a MAC is 24bytes and 1byte, respectively. This data is derived from the simulation results of SEF [3]. There is a global key pool of 1000 keys, number of partitions is 25 and each node possesses 20 keys.

(a) Probability of reports transmission   (b) Energy consumption
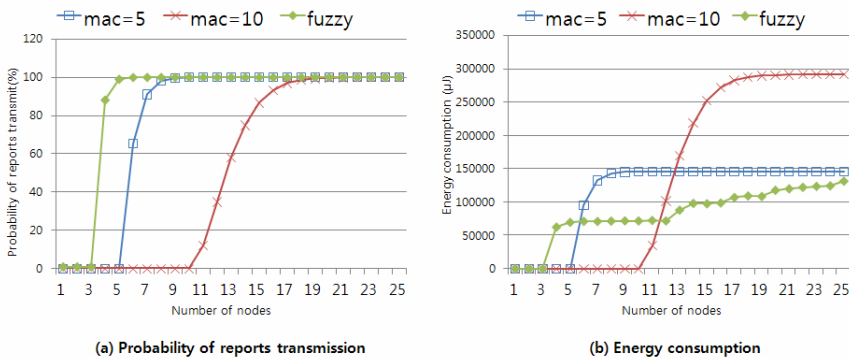
Fig. 5. The probability and energy consumption of transmit normal reports when the hop count and remaining energy are large enough (i.e., hop count=100, energy=100)

Fig. 5 shows the changes in the probability of reports transmit and energy consumption when the number of sensor nodes detecting the same event is increased and hop count is large enough and energy level is high. Fig. 5(a) shows that the proposed fuzzy-based MAC length determination method has higher performance than SEFs with a fixed MAC length (i.e., 5 and 10). Due to the fact that the proposed method can determine the MAC length based on the node density, it can generate legitimate reports when there are a small number of nodes detecting the same event, but the SEF with a fixed MAC length (i.e., 5 and 10) cannot generate legitimate reports until the number of nodes exceeds their MAC length. This is because the CoS nodes need distinct partitions keyed MACs to generate legitimate reports. Fig. 5(b) shows the energy efficiency of the proposed method. The SEFs with a fixed MAC length (i.e., 5 and 10) consume no energy until the number of nodes exceeds their MAC length, because the CoS nodes cannot generate a legitimate report in this case, so real occurred events cannot be forwarded to the BS. On the other hand, the proposed method

(a) Number of Compromised
Partitions = 3

(b) Number of Compromised
Partitions = 4

Fig. 6. Energy consumption for a the false reports ratio of 80%

(a) Number of Compromised
Partitions = 3

(b) Number of Compromised
Partitions = 4
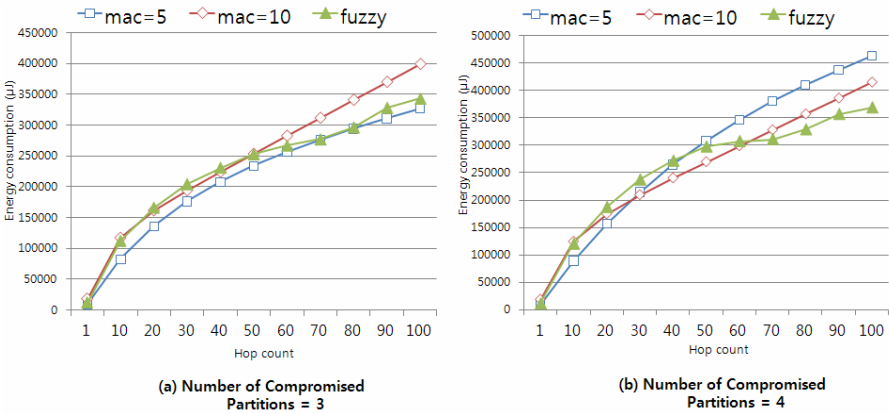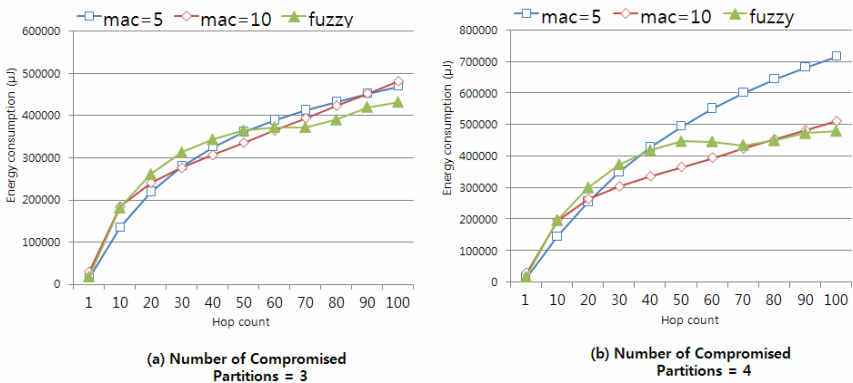
Fig. 7. Energy consumption for a the false reports ratio of 90%

can generate legitimate reports via small number of nodes, and it consumes less energy than SEFs in the case of the reports generated.

Fig. 6 and Fig. 7 show the energy consumption of reports transmission for a false reports ratio of 80% and 90%. The false reports ratio indicates the number of false reports injected into the network. The simulation results show that when the number of compromised partitions is large and the false reports ratio is large, the SEF with the larger MAC length (i.e., 10) consumes less energy and the proposed method consumes less than or same amount of energy than other methods. In addition we know that the SEF with the larger MAC length (i.e., 10) can detect false reports early, but during normal reports forwarding it has more overhead. Thus the proposed fuzzy-based MAC length determination method is more efficient, because it can adapt to the network conditions.

5   Conclusion and Future Work

In this paper, we proposed a fuzzy-based MAC length determination method as a means to conserve energy. In SEF, a small MAC length reduces the overhead but it also reduces the probability of false reports being detected. On the other hand, a large MAC length can detect false reports in early hops, but if the number of nodes detecting the same event is too small to generate legitimate reports, it cannot forward the reports to the BS. Thus we consider the node density, the hop count and the energy in order to determine an efficient MAC length. The simulation results show that the proposed method can provide greater flexibility and higher energy efficiency.

In future works, we will research the mean to improve the accuracy of estimating the fuzzy input parameters affecting the performance of the proposed method. We will also research optimization of fuzzy logic in the proposed method in order to improve the performance.

References

1. Akyildiz, I.F., Su, W., Sankarasubramaniam, Y., Cayirci, E.: A Survey on Sensor Networks. IEEE Commun. Mag. 40(8), 102–114 (2002)
2. Przydatek, B., Song, D., Perrig, A.: SIA: Secure Information Aggregation in Sensor Networks. In: Proc. SenSys, pp. 255–265 (2003)
3. Ye, F., Luo, H., Lu, S.: Statistical En-Route Filtering of Injected False Data in Sensor Networks. IEEE J. Sel. Area Comm. 23(4), 839–850 (2005)
4. Yang, H., Lu, S.: Commutative Cipher Based En-Route Filtering in Wireless Sensor Networks. In: Proc. VTC, pp. 1223–1227 (2003)
5. Al-Karaki, J.N., Kamal, A.E.: Routing techniques in wireless sensor networks: a survey. IEE Wirel. Commun. 11(6), 6–28 (2004)
6. Yu, Z., Guan, Y.: A Dynamic En-route Scheme for Filtering False Data Injection in Wireless Sensor Networks. In: Proc. of SenSys, pp. 294–295 (2005)
7. Zhu, S., Setia, S., Jajodia, S., Ning, P.: An Interleaved Hop-by-Hop Authentication Scheme for Filtering of Injected False Data in Sensor Networks. In: The IEEE Symposium on Security and Privacy, pp. 259–271 (2004)

8. Li, F., Wu, J.: A Probabilistic Voting-Based Filtering Scheme in Wireless Sensor Networks. In: The International Wireless Communications and Mobile Computing Conference, pp. 27–32 (2006)
9. Lee, H.Y., Cho, T.H.: Fuzzy Adaptive Selection of Filtering Schemes for Energy Saving in Sensor Networks. IEICE Trans. Commun. E90–B(12), 3346–3353 (2007)
10. Buttyán, L., Dóra, L., Vajda, I.: Statistical wormhole detection in sensor networks. In: Molva, R., Tsudik, G., Westhoff, D. (eds.) ESAS 2005. LNCS, vol. 3813, pp. 128–141. Springer, Heidelberg (2005)
11. Lee, H.Y., Cho, T.H.: Key inheritance-based false data filtering scheme in wireless sensor networks. In: Madria, S.K., Claypool, K.T., Kannan, R., Uppuluri, P., Gore, M.M. (eds.) ICDCIT 2006. LNCS, vol. 4317, pp. 116–127. Springer, Heidelberg (2006)

A Coverage and Energy Aware Cluster-Head Selection Algorithm in Wireless Sensor Networks

Thao P. Nghiem, Jong Hyun Kim, Sun Ho Lee, and Tae Ho Cho

School of Information and Communication Engineering
Sungkyunkwan University
Suwon 440-746, Republic of Korea
{thaonp,jonghkim,sunholee,taecho}@ece.skku.ac.kr

Abstract. The issue of identifying appropriate cluster-heads has recently been the focus of extensive research and development in wireless sensor networks. Unfortunately, cluster-heads are generally chosen either in a random manner or mainly based on nodes' residual energy. Accordingly, there is no guarantee that network coverage is well-preserved while this QoS is vital in target tracking and surveillance applications. In order to enhance both coverage preservation and energy efficiency, we propose a Coverage and Energy Aware Cluster-Head Selection Algorithm which fully considers three critical factors: the node's energy, location and especially coverage cost metric. Simulation results demonstrate that our algorithm cannot only prolong the network lifetime over 11%, but also substantially enlarge network coverage, from the middle phase of the network lifetime, by over 20% compared to the traditional energy-based selection methods in LEACH and HYENAS system.

Keywords: Wireless sensor network (WSN), Cluster-based protocol, Energy efficiency protocol, Coverage, Cluster-head selection.

1 Introduction

Wireless sensor networks (WSNs) have recently attracted interest for potential application in future ubiquitous computing systems. Due to limited and irreplaceable energy of sensors, the energy efficiency becomes one of the most challenging tasks for WSN design [1,2]. Besides, coverage is another critical measure of network QoS offered by a WSN. It reflects how well an area is monitored or tracked by sensors [3].

For the initial reason of saving energy, a WSN is broken down into several clusters to reduce communication overhead, and then save energy consumption as in Fig. 1. Close nodes group themselves into local clusters with one node acting as cluster-head. The cluster-head collects data from other cluster nodes and send aggregated data to the base station (BS).

Most of existing research on cluster-head selection is only based on energy consumption [4,5,6]. Low-Energy Adaptive Clustering Hierarchy (LEACH) is a cluster-based protocol that uses randomized rotation of cluster-heads to distribute energy dissipation evenly throughout the sensors [4]. Instead of randomly choosing
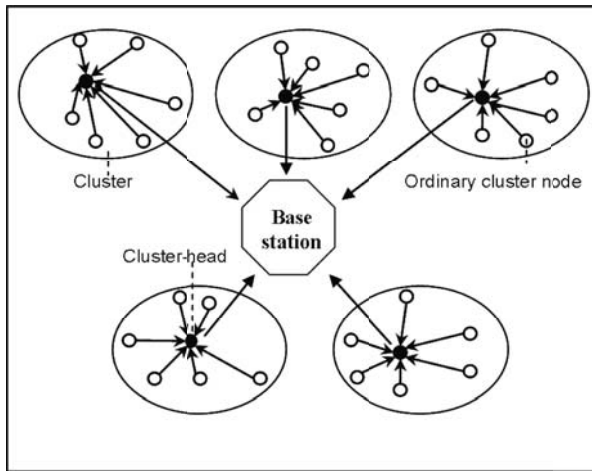
Fig. 1. Black nodes are cluster-heads. White nodes are ordinary cluster nodes.

the cluster-head like in LEACH, the HYENAS system in [5] takes several parameters into consideration including node location and remaining energy. Though these approaches perform much better with regard to reducing energy consumption, they do not ensure to preserve the network coverage. On the other hand, there are also studies on coverage preservation in WSNs [7,8,9]. They show the concept and importance of this metric; however, their cluster-head selection methods are only based on coverage metric rather than integrate this with energy factor.

For integration of coverage preservation and energy efficiency in cluster-based WSNs, we propose a Coverage and Energy Aware Cluster-Head Selection (CEACHS) Algorithm. The main contribution of our paper is suggesting a method to select cluster-heads by taking advantage of nodes' characteristics and especially their coverage cost metric. By applying the proposed CEACHS, we can modify HYENAS system in order to achieve the best sensing coverage and energy efficiency. Simulation results prove that the proposed approach consistently outperforms LEACH and HYENAS in terms of not only extending network lifetime by over 11%, but also enlarging total network coverage by over 20% from the middle phase of the network lifetime.

The rest of the paper is organized as follows. Section 2 is a brief description of the state-of-art work including LEACH, HYENAS and coverage cost metric. In Section 3, we present the overview and details of our algorithm. Section 4 is a discussion of the simulation and its results. Finally, we concludes our work in Section 5.

2   Related Work

2.1   Low-Energy Adaptive Clustering Hierarchy (LEACH)

LEACH is a self-organizing, adaptive clustering protocol. In LEACH, the nodes organize themselves into local clusters with one node acting as the cluster-head.

At each round, every node m is assigned a random number X between 0 and 1. If X is less than a predefined threshold $T(m)$, node m will be selected as a cluster-head node at the current round. The threshold T(m) is set to

$$T(m) = \begin{cases} \frac{p}{1-p[\lambda \mod \frac{1}{p}]}, & \text{if } m \in G; \\ 0, & \text{otherwise} \end{cases} \tag{1}$$

where p is the desired ratio of cluster-heads, λ is the current round and G is the set of nodes each of which has not acted as a cluster-head yet within a period of $1/p$ round.

These cluster-head nodes broadcast their status to the other sensors in the network. Each sensor determines which cluster it wishes to join by choosing the cluster-head that requires the minimum communication energy. Once all are organized into clusters, each cluster-head creates a schedule for the nodes in its cluster. Once the cluster-head has all the data of nodes in its cluster, the cluster-head node aggregates the data and then transmits the compressed data to the BS. Since the cluster-heads spend more energy than other nodes, it is essential to re-select cluster-heads periodically.

2.2   HYENAS System

The HYENAS System selects cluster-heads using a hybrid algorithm which combines model-based processing with a machine learning technique called Case-Based Reasoning (CBR). To appropriately select a cluster-head for each cluster, the BS determines a node metric $CH(m)$ for each node m in the cluster by taking into account each node's remaining energy $Er(m)$, the total sum of squared distance d_{mi} from the concerned node m to other nodes i in the same cluster and distance from node m to the BS d_{BS}. $CH(m)$ is calculated using the formula as below.

$$CH(m) = W_1 Er(m) + W_2(1 - \frac{1}{(\sum_{i=1}^{N} d_{mi}^2) + d_{BS}^2}) \tag{2}$$

where W_1 and W_2 are weights for the node's remaining energy and location respectively. N represents the number of all nodes in the cluster.

2.3   Coverage Cost Metric

Coverage cost is originally introduced in DAPR as a routing metric to avoid routing of data through areas sparely covered by the sensor nodes [8]. Since then, there have been several studies which focus on analyzing the coverage cost metric to prolong network lifetime [9,10].

In the Coverage Preserving Protocol, Y.R. Tsai proposed a cluster-head selection algorithm based on the coverage cost or normalized sensing coverage area of each sensor node m [7]. Accordingly, they assume η_0 is the percent of sensing area only covered by node m and η_i is the percent the percent of sensing area
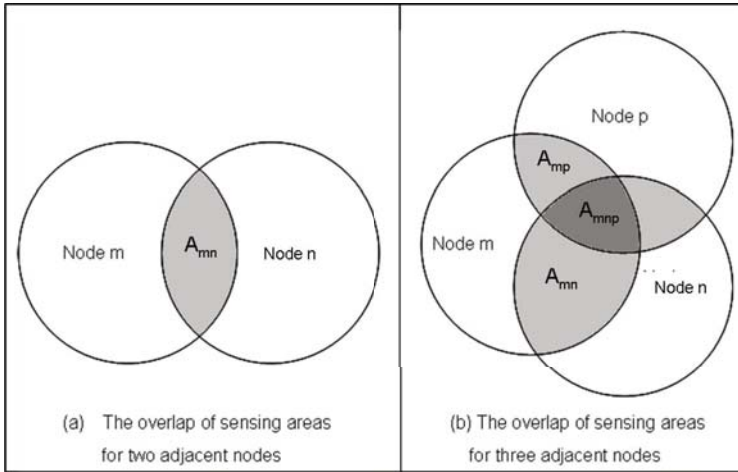
Fig. 2. The overlap of sensing areas for two adjacent nodes

covered by this node and other i neighbor nodes. Coverage cost $\eta(m)$ of node m is defined as below

$$\eta(m) = \eta_0 + \sum_{i=0}^{\infty} \frac{\eta_i}{i+1} \qquad (3)$$

For the simplest example in Fig. 2(a), $\eta(m)$ is calculated as

$$\eta(m) = \eta_0 + \frac{\eta_1}{2} = (1 - \frac{A_{mn}}{\pi R^2}) + \frac{A_{mn}}{2\pi R^2} = 1 - \frac{A_{mn}}{2\pi R^2} \qquad (4)$$

where η_1 is the percent the percent of sensing area covered by node m and node n, A_{mn} is the overlapping are of those two nodes.

In fact, the sensing range of a specific node m likely overlaps with several nodes like Fig. 2(b), thus calculating becomes complicated and needs location information for all nodes. Therefore, they approximate all neighbor nodes as an equivalent node with an equivalent distance to the desired node m. The estimation of this distance is based on energy consumption to transmit and receive beacon messages to all its neighbor nodes.

In addition, four approaches to measure the coverage cost metric are evaluated in [9]. Among them, coverage redundancy cost is similar to the normalized sensing coverage area η in [7].

3   Coverage and Energy Aware Cluster-Head Selection (CEACHS) Algorithm

3.1   Motivation

In HYENAS, the cluster-heads are selected according to their energy and location but not coverage while this is a vital metric, especially in surveillance

applications. If nodes with large coverage area are selected as cluster-heads, they will spend a large amount of energy and die off first. As a result, the total network coverage quickly falls down. Vice versa, if nodes with small coverage area are selected as cluster-heads and run out of their energy first, the reduction of the total network coverage due to these dead nodes is minimized. In this paper, we propose a cluster-head selection algorithm which aims to integrate both coverage preservation and energy efficiency. Apart from those factors in HYENAS, coverage is carefully considered in cluster-head selection procedure.

3.2   Detailed Algorithm

Calculation of Coverage Cost η. At first, every node m calculates its coverage cost or estimated normalized effective sensing area $\eta(m)$ in the initialization and set-up phase. As shown in [7], it is too complicated to calculate the exact value of $\eta(m)$. Therefore, an approximate approach which was based on the amount of energy consumption to transmit and receive beacon messages was proposed.

The specific node m, with radius R, sends beacon messages to neighbors which are in range of $2R$ radius. Transmission energy Er_{trans} (dB) is calculated as

$$Er_{trans} = Er_{sens} + L(2R) \tag{5}$$

where R is the sensing radius of node m. Er_{sens} (dB) represents the sensitivity of radio receiver and $L(2R)$ is denoted as the propagation loss for a distance of $2R$.

It is assumed that there are M neighbor nodes responding to beacon messages. Node m must spend $Er_{receive}$ (dB) to receive these replies

$$Er_{receive} = 10log_{10}\frac{\sum_{i=0}^{M}10^{\frac{P_i}{10}}}{M} \tag{6}$$

where P_i is the received signal energy level, for $i = 1, 2,..., M$.

The equivalent distant R' from the equivalent node m' to node m is approximated as

$$R' = 2R10^{\frac{(Er_{sens}-Er_{receive})}{10\beta}} \tag{7}$$

where β is denoted as the path loss exponent.

With $\rho = R'/2R$, the equivalent normalized overlapping area $\Phi(m)$ of node m is obtained.

$$\Phi(m) = \frac{2[cos^{-1}(\rho) - \rho\sqrt{1 - \rho^2}]}{\pi} \tag{8}$$

According to (3), $\eta(m)$ finally becomes:

$$\eta(m) = 1 - \frac{\Phi(m)}{2} \tag{9}$$

Cluster-Head Selection. For each cluster, the BS calculates the $CH(m)$ value for each node m. The node which possesses the maximum value of $CH(m)$ in each

cluster will become the cluster-head of that cluster. Five input factors related to node m's characteristics are considered to decide $CH(m)$ value. Those are:

- Residual battery power $Er(m)$. The higher $Er(m)$, the higher the probability node m becomes cluster-head.
- Relative distance d_1 from node m to the other nodes in the same cluster.

$$d_1^2 = \sum_{i=1}^{N} d_{mi}^2 \tag{10}$$

where d_{mi} is the distance from node m to other nodes i which is in the same cluster with node m and N is the size of the cluster which contains node m. The lesser d_1, the higher the probability node m becomes cluster-head.
- Coverage cost $\eta(m)$, which has already been calculated. The less $\eta(m)$, the higher the probability node m becomes cluster-head.

All those factors are combined to calculate $CH(m)$ value in the following formula:

$$CH(m) = W_1 Er(m) + W_2(1 - \frac{1}{d_1^2 + d_2^2}) + W_3[1 - \frac{1}{\eta(m)}] \tag{11}$$

where W_1, W_2 and W_3 are weights for the node's remaining energy, location and coverage cost, respectively.

The selection of these weights is adjusted to fix the purpose of WSN applications. For example, coverage is the most critical metric in target tracking and surveillance applications, thus the value of W_3 here should be higher than the others. Otherwise, in networks where full coverage is not an important requirement, W_1 and W_2 should be increased while W_3 should be tuned into a small value.

3.3 Adaption of CEACHS

We apply the proposed algorithm into HYENAS in order to achieve the best improvement in both coverage and energy efficiency. The new protocol is named H-CEACHS. In H-CEACHS, the algorithm is executed in the initialization and set-up phase of each round or when CBR in HYENAS decides to reform clusters. Moreover, it is noted that calculation of coveage cost η should be done whenever the network topology changes. However, the energy consumption for this calculation is insignificant since the amount of energy to transmit and receive beacon messages is too small to compare with that to transmit and receive sensing reports.

4   Performance Evaluation

4.1 Energy and Channel Models

In our proposed algorithm, we use a simple propagation loss model to compute the energy consumption of radio transmission in [7]. According to this model,

the energy dissipation for transmitting a k-bit message to a receiver at a distance d is:

$$E_{T_x}(k, d) = k(E_{elec} + \epsilon_{amp}d^{\beta}) \tag{12}$$

where $E_{elec} = 50\text{nJ/bit}$ is the energy used to operate the transmitter or receiver circuitry and $\epsilon_{amp} = 100\text{pJ/bit/unit}^{\beta}$ is the transmitter power amplifier for transmitting a bit to a receiver with a distance $d = 1$ unit and a path loss exponent β.

To receive a k-bit message, a node spends:

$$E_{R_x}(k) = kE_{elec} \tag{13}$$

4.2 Simulation Environment

We randomly and uniformly deploy 100 nodes with a sensing radius of 7.5 units in a network area of 50x50 using MATLAB programming language. The BS is 50 units away from one side of the sensing area. The path loss exponent $\beta = 2$ for transmission between nodes and $\beta = 2$ for transmission between cluster-heads and the BS. The initial energy of each node is 1J. When a cluster-head node receives data from other non cluster-head nodes in its cluster, it aggregated and compressed data with a compression coefficient . Data sent from a node to its cluster-head node in each round has a length of $k = 2000\text{bit}$. The initial number of cluster is 5, which equals to 5% of the total live nodes. This simulation run for 2000 rounds, which are divided into three phases: the early phase (from the 1st round to the 500th round), middle phase (from the 501st round to the 1300th round), and the final phase (from the 1301st round to the 2000th round).

Since nodes are deployed randomly in the network, the results of each run time are different from the others. We therefore executed this simulation up to 100 times to obtain average results for more accurate and objective data.

4.3 Simulation Results and Discussion

In our simulation, we focus on comparing four metrics of network performance: sensing coverage, number of alive nodes, distribution of dead nodes and network energy consumption among LEACH, HYENAS and H-CEACHS.

In Fig. 3(a), we compare the network sensing coverage C versus the time steps among LEACH, HYENAS and H-CECHS. The network sensing coverage C% between 0% and 100% is defined as the ratio of the network coverage area to the entire desired area. Overall, it is easy to see that H-CEACHS is definitely better in whole network lifetime. Specially, H-CEACHS demonstrates its superior performance with a gain of over 20% from the middle phase onwards. Since our algorithm carefully considers coverage as a factor when selecting cluster-head, it can avoid choosing nodes with high coverage cost as cluster-heads. As a result, a considerable amount of network sensing coverage is preserved in H-CEASHS.

Fig. 3(b) shows number of alive nodes in H-CEACHS is sometimes better and sometimes worse than that in HYENAS before the final phase; however, it is
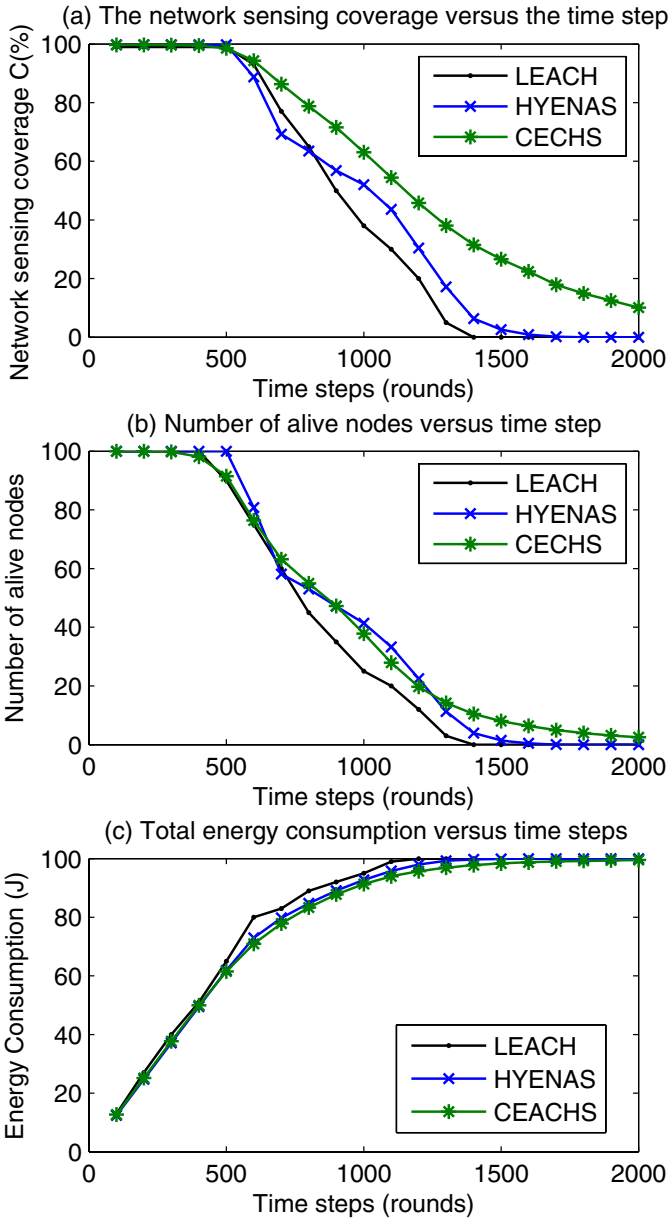
Fig. 3. (a) The network sensing coverage C% versus the time step (b) Number of alive nodes versus time step (c) Total energy consumption versus time steps
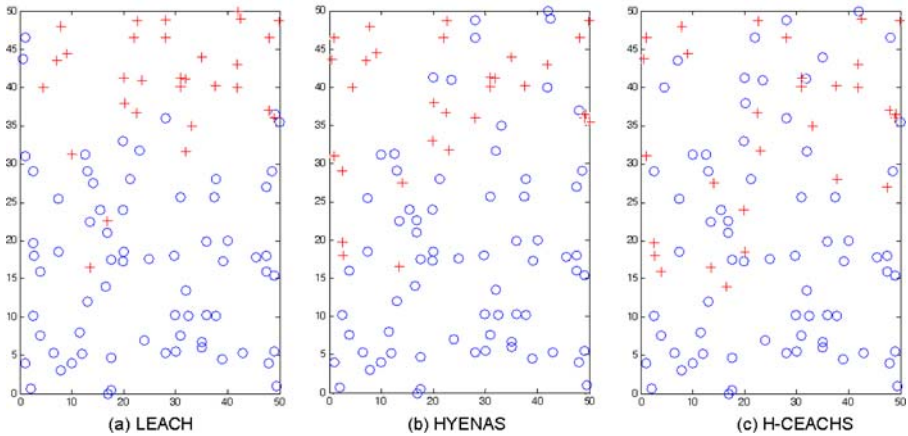
Fig. 4. Location of dead nodes when the number of dead nodes exceeds 30 nodes

absolutely the best at the final phase and always better than that in LEACH. Moreover, to preserver the network coverage, the less number of dead nodes is not as important as the more even distribution of dead nodes. We simulated a case to reflect the location of dead nodes when the number of dead nodes is over 30 in Fig. 4. With the assumption that the BS is far below each sub figure, nodes which is farther from the BS will die off sooner than the other. Compared with LEACH and HYENAS, dead nodes in H-CEACHS spread more widely rather than concentrate in a small area, and thus their invalid coverage area is possibly covered by their neighbors without much reducing the total network coverage.

Fig. 3(c) demonstrate the total energy consumption in the whole system versus time step for LEACH, HYENAS and H-CEACHS. This metric in H-CEACHS is lower and steadier than the other protocols. Furthermore, while other protocols used up their energy before the 1700th round, our protocol still worked after 2000 rounds. As a result, our protocol can extend network life-span by over 11%.

5   Conclusion

In this paper, we proposed CEACHS, a hybrid algorithm to select appropriate cluster-heads. This algorithm considers many critical parameters including nodes' remaining energy, location and, most importantly, coverage cost to achieve optimal energy efficiency and coverage preservation. Our simulation results prove that the proposed algorithm outperforms other ones in baseline protocols such as LEACH and HYENAS in terms of both network sensing coverage and energy efficiency.

References

1. Akyildiz, I.F., Su, W., Sankarasubramanian, Y., Cayirci, E.: A Survey on Sensor Networks. IEEE Communication Magazine 40, 102–114 (2002)
2. Akkaya, K., Younis, M.: A Survey on Routing Protocols for Wireless Sensor Networks. Ad-hoc Networks 3, 325–349 (2005)
3. Huang, C.F., Tseng, Y.C.: The Coverage Problem in Wireless Sensor Network. Mobile Network Applications 10, 519–529 (2005)
4. Heinzelman, W.R., Chandrakasan, A., Balakrishnan, H.: Energy-Efficient Communication Protocol for Wireless Microsensor Networks. In: 33rd Hawaii International Conference on System Sciences, p. 8020. IEEE Computer Society, Washington (2000)
5. Tillapart, P., Thumthawatworn, T., Pakdeepinit, P., Yeophantong, T., Charoenvikrom, S., Daengdej, J.: Method for Cluster-Heads Selection in Wireless Sensor Networks. In: IEEE Aerospace Conference, pp. 3615–3623. IEEE Press, Montana (2004)
6. Younis, O., Fahmy, S.: Heed: A Hybrid, Energy-Efficient, Distributed Clustering Approach for Ad-hoc Sensor Networks. IEEE Transaction on Mobile Computing 3, 366–379 (2004)
7. Tsai, Y.R.: Coverage-Preserving Routing Protocols for Randomly Distributed Wireless Sensor Networks. In: IEEE Global Telecommunications Conference, pp. 1–5. IEEE Press, California (2006)
8. Perillo, M., Heinzelman, W.: DAPR: A Protocol for Wireless Sensor Networks Utilizing an Application-based Routing Cost. In: IEEE Wireless Communications and Networking Conference, pp. 1540–1545. IEEE Press, New York (2004)
9. Soro, S., Heinzelman, W.B.: Cluster-Head Election Techniques for Coverage Preservation in Wireless Sensor Networks. Ad-hoc Network 7, 955–972 (2008)
10. Bae, K., Yoon, H.: Autonomous Clustering Scheme for Wireless Sensor Networks Using Coverage Estimation Self-Pruning. IEICE Transaction on Communications E88-B, 973–980 (2005)

u-Healthcare Service Based on a USN Middleware Platform and Medical Device Communication Framework

Yung Bok Kim

Department of Computer Engineering/Venture Biz Center
Sejong University, Seoul, Korea, 143-747
`yungbkim@sejong.ac.kr`

Abstract. We developed a middleware platform, i.e. COSMOS (Common System for Middleware of Sensor Network), for several types of sensor networks including the Zigbee wireless sensor network, the CDMA cellular phone-based network, the RFID reader-based network and the IP-USN based on 6LowPAN. Development has been focused on interfaces for application programs as well as on sensor network abstractions for various ubiquitous sensor networks (USN). Standard interfaces were defined between the USN middleware and USN networks as well as application services. We studied several USN services including u-Healthcare to examine important issues about middleware platform for integrating with other standardized communication framework, e.g. a medical device communication framework. We introduce application services and the real-time data analysis for QoS in the u-healthcare service.

1 Introduction

The middleware for ubiquitous sensor networks, i.e. USN middleware, was studied to support a variety of applications with the available sensor information and computing resources in standardized way. For USN middleware, we abstracted lower layers related to various sensor networks as well as RFID readers. Various application services can be developed above the USN networks. USN middleware on the USN networks shall support context-aware processing based on qualified data-acquisition for context-aware applications.

The discovery of appropriate sensor/network/service is a key function in the implementation of context-aware applications for mobile users in ubiquitous computing environments. An efficient semantic service discovery scheme called UbiSearch [1] for large-scale ubiquitous computing environments was introduced. We also studied and present a unified USN Directory Service based on the indexing key and metadata for the appropriate discovery of the sensor/network/service.

We introduce the USN software platform in the COSMOS project, a (Common System for Middleware of Sensor Network) middleware platform developed by ETRI in Korea [2]. We developed four kinds of application services operating on the platform, and we introduce one application service among them, i.e. the u-Healthcare service. We focused on the study of an application service, i.e. a u-Healthcare service, of

the various services based on its sensor networks composed of wired sensors or wireless sensors in the USN. We defined the USN as a sensor network composed of ubiquitously prevalent wired-sensors as well as ubiquitously prevalent wireless-sensors, which are connected in the network. Beyond the e-healthcare service, the ubiquitous healthcare (u-healthcare) service will be more evolved in the emerging ubiquitous computing and networking society.

Feki et al. [3] introduced the integration of context-awareness and multimodal functionalities in the smart environment. In the emerging USN, sensor devices could be greatly miniaturized and can harvest energy from the environment and will be able to communicate with other networks and devices integrated in our homes or our cities (e.g. u-city) as smart environments. Recently, the new application of health monitoring and medical care based on USN is gaining popularity among researchers and offers genuine promise for future practical uses. We introduce a manager-agent framework for medical device communication for u-healthcare services with USN in smart environments.

The paper is organized as follows. We discuss the USN middleware platform: COSMOS, a framework for u-healthcare, examples of application service including u-Healthcare and data analysis for QoS. The USN middleware platform must be considered for integration with standardized application platforms such as a medical device communication framework (X73). We present the X73 framework for healthcare service to find some issues for integration with the COSMOS platform. Real-time estimation of sensed parameters has not been studied in previous research. We present an applicable scheme for stochastic parameter estimation. For quality of service (QoS), we introduce an evaluation scheme for the u-healthcare service in smart environments. Finally, we conclude our study.

2   USN Software Platform

We attempted the development of an intelligent and autonomous service platform in the USN software platform, COSMOS. USN middleware covers various heterogeneous sensor networks, RFID readers including the mobile RFID reader in mobile phones, CDMA mobile devices such as sensor/sink/gateway nodes, and IP-USN networks. Additionally, the COSMOS middleware platform supports intelligent processing capabilities such as sensed-data mining, context-awareness, intelligent management and unified metadata/application directory services. The standardization for interoperability is important because of the heterogeneous middleware platform. At the ITU-T SG16/Q22 meeting , the concept of USN middleware was introduced as a new work area for USN standardization, and the discussion with revised contributions for standardization of USN middleware continued at the next meetings, as an officially new work area.

The need for health monitoring in the e-healthcare service is gradually increasing and its application is becoming feasible with sensor network technology. By adopting tiny wireless sensor network devices into the ubiquitous networking environment with some sort of specific health monitoring system, the chronically ill and the elderly can be observed in smart environments that provide an e-healthcare service. For this purpose, wearable vital sign sensors can be attached to the body, allowing continuous

communication transferring the sensed physical status of the wearer. Health monitoring with ubiquitous sensor networks (USN) in the ubiquitous networking environment has some outstanding features compared to traditional medical healthcare systems.

USN middleware is comprised of a Service Layer, Intelligence Layer and Abstraction Layer. The Service Layer is comprised of OAPIC (Open API Component), UMMC (USN Middleware Management Component), USSC (USN Middleware Security Service Component) and UDSC (USN Directory Service Component).

This layer provides an application interface for various services. Among the components in this service layer, we focused on the components for a unified and ubiquitous web service. The UDSC contains various metadata to find appropriate resources such as components or sensor networks. If some services are required to monitor an area, then the USN Directory Service can help to look them up. The USN Directory Service offers all relevant information such as the location, wireless protocol, sensor type, number of sensor nodes, sensor network lifetime, etc. Scalability and modularity are very important factors in distributed environments. For a unified and ubiquitous service in heterogeneous middleware environments, we need to develop this UDSC into a unified-UDSC.

The Sensor Network Intelligent Layer is composed of a SDMC (Sensor Data Mining Component), an IEMC (Intelligent Event Management Component) and a SIIMC (Sensor Information Integration Management Component). This layer provides intelligence for sensed-data processing, event processing and context information processing. It requests data from the wireless information infrastructure, i.e. USN networks, through Sensor Network Abstraction Layer.

The SDMC creates query plans to request data. This component is in charge of simultaneous scheduling for requests as well as processing for responses. This component provides mining functionality for sensed-data. The IEMC models an application specific event and notifies the requesting service of the result. It may use the SIIMC to obtain proper sensed information or use a sensor network monitoring component. The SIIMC models service-specific context information and defines the rules to capture the specific context. This component may use the SDMC to obtain context information. This component enables USN middleware to provide intelligent context-aware processing.

The Sensor Network Abstraction Layer is comprised of the SNCIC (Sensor Network Common Interface Component), the SNAMC (Sensor Network Autonomous Management Component) and adaptor components. This layer provides wireless information infrastructure abstraction and sensor network monitoring functionality. The wireless information infrastructure includes RFID readers, various kinds of sensor networks, CDMA mobile devices as sensor/sink/gateway node and IP-USN router/node, etc. This layer provides an abstract interface to the upper layer to use any kind of USN network. The adaptor is responsible for transforming the data format and the command format between USN middleware and the actual wireless infrastructure protocols. The abstraction layer is realized for various USN networks by various adaptors.

3   Framework for u-Healthcare

Health monitoring based on USN in the ubiquitous networking environment provides a totally different healthcare system scenario. The general features of the sensor

network such as tiny sensor nodes, network construction and self-configuration allows for potential uses in medical care monitoring applications.

The standard for point-of-care connectivity [4] establishes a set of specifications that allow seamless multi-vendor interoperability and communication between point-of-care devices, data concentrators, and clinical information systems. The most influential standards in the vendor community derive from bodies that have achieved international influence and authority beyond the formal International/European standards bodies, e.g. the U.S. based HL7 for messaging, DICOM for imaging and IEEE for medical device communications [5].

ISO11073/IEEE1073 (known as X73), Standard for Medical Device Communications, is a family of documents that defines the entire seven layer communication requirements for the 'Medical Information Bus' (MIB). This is a robust, reliable communication service is designed for intensive care units, operating rooms, and emergency room bedside devices. The goal of X73 is to improve the interoperability and plug-and-play capabilities of the different medical devices and medical information systems. Peripheral Area Network Interface (PAN-IF) connects an application hosting device, such as a personal computer, cell phone, or monitoring hub, to a PAN device, which is either a sensor or an actuator. The PAN-IF upper layers are implemented using the ISO/IEEE 11073-20601 Optimized Exchange Protocol, which leverages work from the ISO/IEEE 11073 Medical Device Communications working group. [6]

The following diagram [7] based on a manager-agent framework of medical device communication shows the upper layer communication stack in X73, or layered set of protocol and service components. The manager/agent (client/server) framework for communication systems is defined in the standard. Application processes in both the manager and the agent provide the (user) functionality of the device. For example, the agent has at least one application process that is able to process signals from an invasive blood pressure transducer. The manager application process is responsible for collecting and archiving vital signs information that is provided by the agent.

The Association Control Service Element (ACSE) is the ISO/OSI standard for association control. The Common Medical Device Information Service Element (CMDISE) is an object management service in principle, a lightweight version of the ISO/OSI Common Management Information Service Element (CMISE). The Remote Operation Service Element (ROSE) provides basic services used by the CMDISE (invoke an Operation, return the result of an operation, return an error, and reject an operation). To comply with the definition of optimized encoding rules, a modified version of the ROSE is needed to work with the CMDISE.

ACSE means Association Control Service Element, MDIB means Medical Data Information Base and MDSE means Medical Device Service Element. The Session Layer and Presentation Layer produce a minimized overhead only.

Fig.1 shows the X73 Framework and the layered COSMOS USN middleware platform. The ISO/IEEE 11073 (known as X73) set of standards for Point of Care Medical Device Communication based on a manager-agent framework of medical device communication is the best-positioned international standard to provide interoperability with different sensors [8]. The X73 standard, currently in the development phase, is a single set of standards for complete connectivity between medical devices that contribute plug-and-play, transparency, and ease of use and configuration. X73 distinguishes four main standards groups: transport (e.g., wireless or cabled), services
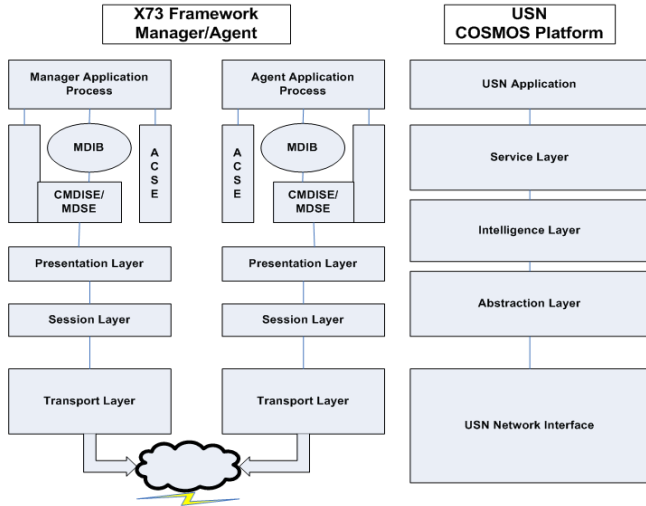
Fig. 1. X73 Framework (Manager/Agent) and USN Middleware Platform

of general applications (e.g., for events or polling), device data (e.g., object-oriented model and representation terminology), and network communication standards (e.g., gateway between the representation of data and ISO/IEEE 11073, DICOM, or HL7 messages). Toledo et al. [8] found a problem with the lack of X73 standardization of wireless connectivity. They plan to fix it when required by using the lower layers of the wireless technology (Bluetooth, Zigbee, etc), then merging them with the upper layers, which actually follow the standard, waiting for the X73 to get updated to these technologies.

In the new and exciting field of Ubiquitous Sensor Networks (USN), e-healthcare services for medical care applications appear as a promising research area with a wide range of possibilities. The following use cases [9] reflect diversity, and there is throughout an assumption of use of existing or supplemented 11073 (X73) standard provisions.

UseCase#1: Mobile Wellness Monitoring of a Single Cardiac Patient;
UseCase#2: Wellness Weight Monitoring of a Single Patient Home;
UseCase#3: Chronic Respiratory Patient Management;
UseCase#4: Elderly patient follow-up;
UseCase#5: Cardiac Parameter Monitoring of a Single Patient Home;
UseCase#6: Secure Registration of Tele-monitoring Device;
UseCase#7: Body Area Network Monitoring;
UseCase#8: Critical cardiac alarms.

The monitoring service keeps tabs on his vital signs once a day and generates a summary of his/her level of activity. Vital signs are recorded using a pressure cuff (blood pressure and heart rate) and a digital thermometer. A wearable 3D accelerometer records patient movements all day and generates parameters that summarize patient activity (% time sitting, % time standing, % time laying, % time sleeping) All this information is sent to a gateway (embedded PC located at the home).

For u-healthcare with the USN, we must consider both the wired sensor/network and the wireless sensor/network. Considering the WSN, we must refer to the afore-mentioned Use-Case#1, #3 and #7 based on X73. We studied the real-time health-monitoring applications, frequently acquiring personal information as well as writing analyzed information as a response to requests from clients or health-monitoring agents in the mobile Internet environment.

4   Application Services and u-Healthcare

We abstracted the sensor network by defining the standard interface between the USN middleware and the sensor network. We decided to use the communication transport protocol with a TCP/IP. The abstraction was accomplished in several steps. First, we standardized the sensing type as temperature, remaining battery lifetime, pressure, gas gauge, bpm (beat per minutes), etc. Second, we defined common message formats used between the USN middleware and the adaptors. We categorize all possible mes-sages into 3 groups. The first group is data request/response. This group contains messages associated with sensor data request/response. The second group is com-mand/notification. This group includes asynchronous sensor data request/notification, activate/deactivate of actuator and error report. The last group is for communication control between the USN middleware and adaptors. This definition of common message format will be refined after further development.

For the consistency of health-monitoring information and for a convenient user interface, we need a unified health-monitoring web server for wired and mobile Inter-net. We need to consider the health-information center accessibility of the doctor or nurse as well as of the patient within the concept of the X73 framework. The health-care web server must have the capacity to show the appropriate health contents, i.e. the HTML contents for the wired Internet as well as the mobile contents for many dif-ferent kinds of mobile devices, e.g. WML, mHTML, etc. For the unified service, there are several constraints compared to the content for the wired Internet. First of all, we must consider the various kinds of mobile devices as well as browsers for the mobile Internet, and each of those devices may have different functionalities. We considered only text-based health-monitoring information from the wrist phone to the health-monitoring Web server and vice versa to be immune to any type of Internet traffic load as well as to minimize the mobile communication cost for cost-effective health-monitoring services based on the X73 framework.

5   Data Analysis for Quality of Service

Important issues such as data quality of sensor nodes, real-time aggregation and esti-mation of random variables for comprehensive accessibility in relation to the u-Healthcare service must be considered. We have implemented health-monitoring sensors in the wrist phone for a patient as a primitive example with single-hop sensor nodes.

We used an exponentially weighted moving average model with appropriate smoothing parameter β to get the mean value of the random variable h_k . To find the

statistical outlier we need to get deviation, and we used the mean absolute deviation (MAD) model to reduce the complexity of calculation.

$$\overline{h_k} = \beta h_k + (1-\beta)\overline{h_{k-1}} \quad \text{(where } 0 < \beta < 1\text{)}. \tag{1}$$

The mean absolute deviation (MAD) is defined by

$$MAD = \frac{1}{N} \sum_{i=k-N+1}^{k} \left| h_i - \overline{h_k} \right|, \tag{2}$$

where the chosen sample size for statistics is N, that is the number of samples to be stored, the samples have values h_i, the mean is $\overline{h_k}$ ($k > N\text{-}1$) and $N \approx \frac{1}{\beta}$. If the smoothing parameter β is too small, N becomes larger. For sensor networks, if N_s is the number of sensors, then the memory size for real-time estimation becomes $N_s * N = N_s * \frac{1}{\beta}$.

If T_u is the mean of updating period, and t_u is the random variable of updating time interval, then

$$T_u = \overline{t_u} = \beta t_k + (1-\beta)\overline{t_{k-1}}. \tag{3}$$

The sampling frequency Fs may be different for each random variable, and

$$F_s \geq 2 F_u = \frac{2}{T_u} = \frac{2}{\overline{t_u}} = \frac{2}{\beta t_k + (1-\beta)\overline{t_{k-1}}}, \tag{4}$$

(F_u is the updating or changing frequency of a random variable).   Equations (1) through (4) will be used for real-time estimation of parameters for u-healthcare service.

We studied real-time health-monitoring applications, frequently gaining access to personal information. Fig.2 shows u-Healthcare service with sensor nodes (e.g. wrist phones) and USN middleware: COSMOS. We will discuss various aspects of the QoS of a ubiquitous application service.

For the real-time health-monitoring network using the mobile Internet in the USN environment, the dominating factor and the standard deviation (or *MAD*) of that random variable must be bound within the deterministic response time.

We studied the important QoS metric, delay, from the user's perspective. The preparation time for a patient to get a service (e.g. medical device, mobile device, etc.) is E (ubiquitous environment time metric); the time spent by the patient with medical devices to make an appropriate action for the service is M (device time metric); the aggregate time to the medical healthcare server after the medical device for medical service is W (service time metric for wired/wireless communication and Web server); the time depending upon medical contents is H (health contents time metric). If iteration of each variables is required, then subscript *i, j, k, l, m* can be used as E_i, M_j, W_k, H_l, P_m. Depending upon ubiquitous environment, device, service and patient, the iteration *i, j, k, l, m* may be different.
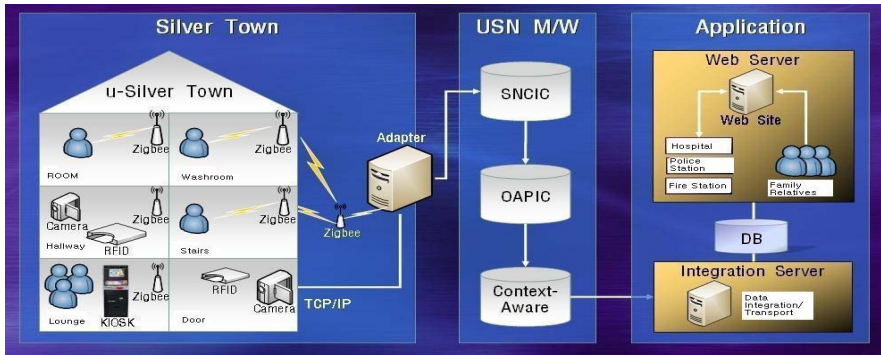
Fig. 2. u-Healthcare Service with Sensor Nodes and USN Middleware: COSMOS

We can represent the statistical values (i.e. mean, deviation or *MAD* using equations (1), (2)) of random variables as performance metrics for QoS (quality of service) of real-time medical service, and the quality of u-healthcare service can be compared with the QoS performance metrics as comprehensive accessibility. Each performance metric can be used to enhance the QoS as follows. The ubiquitous environment time metric *E* will be shortened depending upon the proliferation of the smart environments for u-healthcare. Medical device time metric *M*, i.e. the handling speed of the medical device, is one of the most important performance factors in any medical services for the patient. Wired/wireless Web service time metric *W* can be shortened according to the communication facility and Web server with the DB in the e-hospital. Health contents time metric *H* can be shortened considering accessibility in contents design for the u-healthcare service. Patient time metric *P* may be different from patient to patient depending upon the disability of the patient.

Among various application services, we focused on the u-healthcare service among them, using the USN middleware platform. We considered important issues such as data quality of sensor nodes, real-time aggregation and estimation of random variables for comprehensive accessibility. We considered various aspects of the QoS of a ubiquitous application service.

With the performance metrics (*E, M, W, H, P*), we can estimate other metrics for QoS. For example, total service time for a patient, *T*. We can use the mean values of the random variables of metrics as follows: $\overline{E}$, $\overline{M}$, $\overline{W}$, $\overline{H}$, $\overline{P}$ and

$$\overline{T} = \overline{E} + \overline{M} + \overline{W} + \overline{H} + \overline{P} . \tag{5}$$

With the combination of equations (1), (2), (3), (4) and (5), we can estimate the QoS for each different service requirement based on the random variables. For example, the total service time of u-healthcare for QoS may be more important than each embedded time delay. For improvement of QoS, the statistics and analysis for each random variable may also be important. In the case of Web contents for u-healthcare, the Web-contents accessibility may be more important than the elapsed time to access web contents, if we consider the unified and ubiquitous Web service.

6   Conclusion

We introduced a USN middleware platform, COSMOS, for ubiquitous environments. USN middleware covers various USN networks. We introduced a USN middleware platform defined with standardized interfaces between wireless infrastructure and application services. Of the many services, we studied a u-Healthcare service to consider important issues such as QoS, real-time data analysis of random variables and comprehensive accessibility in the USN environment. A manager-agent framework of medical communication for ubiquitous healthcare service based on X73 with USN was introduced, focusing on the integration of USN middleware platform and medical device communication framework for a u-Healthcare service. We showed important issues of real-time data analysis for QoS metrics for u-healthcare service in the USN environments.

References

1. Kang, S., Kim, D., Lee, Y., Hyun, S.J., Lee, D., Lee, B.: A Semantic Service Discovery Network for-Large-Scale Ubiquitous Computing Environments. ETRI Journal 29(5), 545–558 (2007)
2. Kim, Y.B., Kim, M., Lee, Y.J.: COSMOS: A Middleware Platform for Sensor Networks and a u-Healthcare Service. In: ACM SAC 2008, Fortaleza, Ceará, Brazil, March 16-20, pp. 512–513 (2008)
3. Feki, M.A., Renouard, S., Abdulrazak, B., Chollet, G., Mokhtari, M.: Coupling context awareness and multimodality in smart homes concept. In: Miesenberger, K., Klaus, J., Zagler, W.L., Burger, D. (eds.) ICCHP 2004. LNCS, vol. 3118, pp. 906–913. Springer, Heidelberg (2004)
4. POCT1-A2, Vol. 26 No. 28, Replaces POCT1-A, Vol. 21 No. 24, Point-of-Care Connectivity; Approved, Standard—Second Edition, Clinical and Laboratory Standards Institute, CLSI (2006)
5. IEEE1073 framework Draft V3.1. Report from the CEN/ISSS eHealth Standardization Focus Group. Current and future standardization issues in the e-Health domain: Achieving interoperability. eHealth Standardization Focus Group, July 23 (2004)
6. Acrroll, R., Cnossen, R., Schnell, M., Simons, D.: Continua: An Interoperable Personal Healthcare Eco-system. IEEE Pervasive Computing, 90-94 (October-December 2007)
7. ISO/IEEE P11073-20101/d06f, IEEE 1073.2.1.1^{TM}/d06f, Date: September 2002, Draft Standard for Health informatics – Point-of-care medical device communication - Application profiles – Base standard (2002)
8. Toledo, P., Galarraga, M., Martinez, I., Serrano, L., Fernandez, J.: Pozo: Towards e-Health Device Interoperability: The Spanish Experience in the Telemedicine Research Network. In: IEEE 28th EMBS Conference, New York City, USA, August 30-September 3 (2006)
9. CEN 2006 WG Working document: Use case examples, CEN TC 251/WGIV N 06 xxx. Health informatics: Point-of-care medical device communication -Use cases for mobile, ambulant and home care, July 18 (2006)

Research on Real-Time Software Sensors Based on Aspect Reconstruction and Reentrant

Tao You, Cheng-lie Du, and Yi-an Zhu

School of Computer College
Northwestern Polytechnical University
Xi'an, China
yoytao@163.com

Abstract. In order to effectively monitor the state of software running, it is necessary to embed the software sensor into the program. New software sensors based on aspect oriented technology was designed for overcoming the deficiencies of traditional method in software sensor design and implantation. An inertial navigation system was taken as example and discussed. Practical application shows that the method of software design and implantation of the sensor has guiding significance.

Keywords: Software sensor, aspect-oriented, crosscutting concerns, reconstruction, reentrant.

1 Introduction

With the development of the computer software, the effect to performance and functionality of the entire aviation system is more obvious. Software of its own complexity and scale is also increasing, such as the United States Joint Strike Fighter project in the software required as many as 15 million lines of code, around 35 processes and 400 tools [1]. In china, a number of aviation flight model, the accident more than 60% caused from software. A real-time monitoring technology for aviation software needs to be development urgently [2]. Sensor application software is very flexible and can easily be implanted to the target system, access to the state of procedure, such as the value of a variable, method call, the program's control flow and so on, it can effective to overcome these difficulties.

As the Avionics software systems extremely large and also very high reliability requirements, so it is a very complex task for tracking of its status. Even if the detection logic well designed for sensors, their implantation is a very complicated matter, but reasonably efficient software architecture design can be a very good solution to this problem. Object-oriented technology [3] can effectively achieve software systems, but OOP is only specific object package, its can not good support to the abstract meaning of the act, and sensor information to monitor acts of the majority belong to this scope. Aspect oriented modeling [4] can very well make up for these deficiencies.

This article from the time avionics software features to start. First it extracted the different time points of concern property and formed time aspects, then through software sensors implanted complete the restriction of time aspects, at last sensor along

with the time aspect embedded into the software system. This method avoids the implantation point sensor specific questions; determine the timing of implantation of the sensor and the location of the implants. Finally a certain avionics navigation system was taken as an example.

2   Separation of Cross-Cutting Concerns

AOP is a separation of concerns techniques, namely from the sources to the target under the mapping of the mapping relations, an element of sources scattered in the some of the elements belong to target, and at least there is an element struggle with the certain elements of source in these elements, so we can say that it happened "cross" on the element. Source is the crosscutting concern of the target. Basic idea is to describe the different concerns of the system and their relationship, and achieve a single concern through a loosely coupled way, and then rely on the support mechanism for AOP environment and compile these concerns points to the final run program [5].

As the differences required, it caused the different criteria to achieve the act of crosscutting concerns. we must also consider crosscutting data or property and judge whether they are same. In avionics systems, fast alignment, gyrocompass alignment and calibration alignment complete the basic function of alignment, but they have differences principle, and lead to the differences of established class, so much of their crosscutting concerns should be separately classified as different aspects. During this procedure, this will require strict semantic restrictions to control the balance between the local property and consistency problem.

After the crosscutting concerns summarized the corresponding area, the separate process about the aspects from the OO model of avionics software system is an insinuation extraction and iterative process.

(1)Set up a insinuate set. Class set {C1, C2, …, Cm} and crosscutting concerns set{A1, A2, …, An}.

(2)Sort of concerns set. Accordance with the number of quantity and complexity associated to sort the set {A1, A2, …, An}, and has the A1', A2', …, An'. This will single out the large particle size or scope and will help simplify the separate process.

(3) Insinuate. Pick the first aspect A1', and pick the class which may be include A1 and form a new set {C1, C2,…, Ck}.The new set may be the true subset of {C1, C2, …, Cm}, and also a subset may not be true.

(4) Extraction. Extracting the data and method form the new set one by one in the (3), and making the corresponding connection point, the corresponding Ci recorded as Ci', until the A1 extract from the last Cp. Then bring the new tagged set {C1', C2', …, Cp'} back to the original set {C1, C2, …, Cn}, and separated aspect will cleared from the aspect set.

(5)Repeatedly (3) and (4) steps, until the aspects are null, and then finish the extract job.

It can be seen that, the determine of crosscutting concerns is to reduce the coupling degree of the system process, which improve the system scalability, maintainability; aspects of the separation system enhance the analytical, makes sensors logic behavioral

more easily implant into system through aspect, and pave the way for further aspects of the restructuring.

3   Software Sensor Design and Implantation

After implementation the separation of crosscutting concern, it will combine the concerns points with software sensor techniques, using sensors implanted theology, restructure the crosscutting concerns to aspect oriented. In this way, the software sensors will collect procedures status information, and save to the information database, it extend the assertion system and debugging system, and make it can declare assertion in anywhere, provide the more information; aspect oriented technology separate the time and error-control behavior, and form an aspect to achieve real-time control, such the time of expression system and the error will not be distributed at each module, it easy to implementation and manage system when all the time and wrong behavior centralized managed. Thus the system real-time, low-aspect-oriented coupling, the flexibility of the sensor can be good play.

Software sensor is transparent to users; users only need to know how to specify the necessary tracking information. Under normal circumstance, the sensor will generate a corresponding data structure. Software sensor was inserted into the tracking system, monitoring system status, can be inserted manually or automatically. Software sensor can detect the error state of software, as well as collect the state information of software. In this paper, we will use the software sensor status information, such as variable, method call, control flow information and the sensors of its own static information.

Aspect-oriented restructuring is to have the completed functional design of the sensors implanted into the original components that is the inverse process of crosscutting concerns separation. Analyzing the property of concerns points in concerns set {A1, A2, …, An}, concerns raised here are the smaller particle size of concern, such as variety requirements exist in real-time requirements. The failure time of avionics system, there are 50ms, 2.5min and so many cases, the details of the sensor design in different circumstance should be taken into account.

4   Application Example

In this paper, the sub-model of inertial navigation system of an aircraft avionics integrated fire control software was taken as an example to explain the operating way of aspect-oriented software sensor.

4.1   Brief Introduction of Inertial Navigation System

Inertial/satellite navigation system (INS/GNSS) is an autonomous navigation system, it combine the inertial navigation system with satellite navigation system. The system can measure the plane's attitude, flight path angle and acceleration, mainly to the aircraft navigation state vector and support services. After starting inertial navigation system, the three-axis inertial navigation platform points are arbitrary; therefore we must be inertial navigation platform at the point before the navigation system into

working condition, this process will be known as the initial alignment of inertial navigation system. Alignment model also can be divided into many ways, including: Calibration alignment, gyrocompass alignment and fast alignment, each alignment model has sub model. There are five models in this article; they are calibration coarse alignment, calibration alignment, gyrocompass coarse alignment, gyrocompass alignment and fast alignment.

4.2   Crosscutting Concerns Identified and Separated

Inertial navigation system has a lot of crosscutting concerns. From 3.4 steps, we can see that the separation is the most difficult problem and it should be solved at the first. It is difficult to determine the job cross point, because it is related to the specific system function implementation, and some crosscutting points throughout the whole procedure, so we must select the real-time property of the avionics software as the alignment mode's crosscutting concerns points. When the time is mapped to the components, two simple time operations which are the normal and overtime will across many class, is a crosscutting concerns. Such as Fast Alignment, Compass Alignment, Demarcate Alignment in the process of alignment, three classes have the time crosscutting concerns, they are shown in Figure 1. Separation processes are not discussed here.
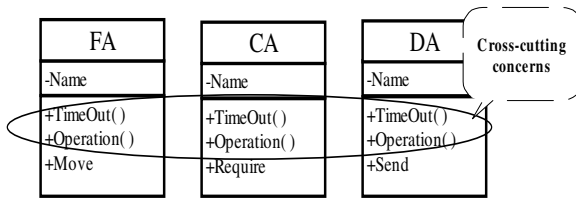
Fig. 1. Alignment mode schematic of cross-cutting concerns

4.3   Time Design

Aspect oriented real-time system modeling take the time as an independence aspect, so it can set up a time model to express the time. The time model in references define a time model to express a unified system time, the specific timing of system model are examples of time mode.

4.4   Time Reorganization

Sensors are embedded into the real-time aspects of alignment system, in order to achieve aspect oriented reorganization. It centralized manage the real-time characteristics and monitor the status of software system at the same time.

1. Software sensor design

This case illustrates the real-time aspects of sub-model in a plane avionics system inertia navigation system, the time required can be divided into overtime 50ms data acquisition and 2.5min sub-state transitions. Software sensors based on time-oriented
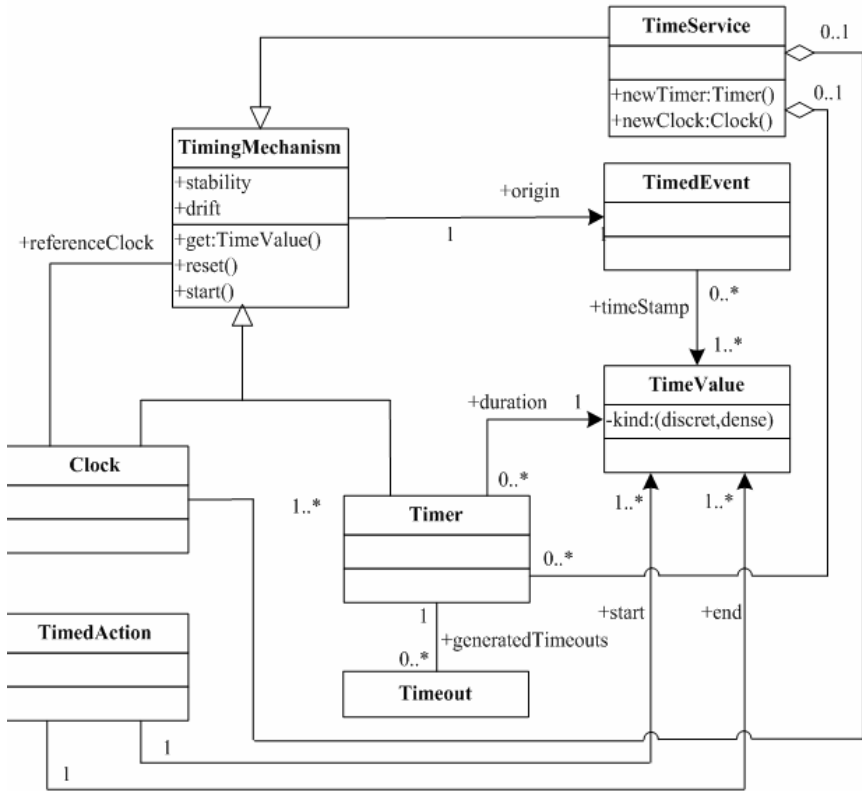
Fig. 2. Time model

are divided into tow categories, SoftwareSensor_Buffer_Timeout(), SoftwareSensor_Transition_Timeout( ), and separate the different concerns points, handle them separately.

2. Software sensor implant

The key of the whole acquisition process is the software sensor must inserted into the location of monitor object in code segment, in order to execute the software sensors Acer operation after the monitored object crate real-time data, so it can reflect monitoring real-time property.

Software sensors spread information into the database, test users can inquiry these data, as the plane's black box, and record the status of the mode transformation and scanning data's validity.

4.5 The Sensor Implant Based on Time Aspect

Real-time requirement is one of the basic requirements in avionics system. Software sensors as a function are implanted into the calibration coarse alignment, calibration alignment, compass coarse alignment, compass alignment, and fast alignment. Aspect-oriented modules will judge the time if it is overtime, and return a result, the

sensor module will execute next step in accordance with the return results. Sensor records the state parameters and save them into database at the same time.

Model time behavior express time relation through adding time tag in the extend sequence diagram, and we put the time as an aspect of separation, and can refine the model through the state diagram. Taking the process of the navigation system as the main clue, we use the sate diagram to express the object relations. But the state diagram does not support aspect-oriented behavior modeling, therefore it is necessary to extend the state diagram and increase the expression mechanism for modeling behavior. In order to clearly express the status of each object, as well as the process of the aspects weaving, we use level state diagrams to describe behavior of crosscutting aspects, and implementation time woven into the core of system function modules.

Time aspects provide initialization time, normal time, overtime and other state. The connection between sub-state transformation and the time is complex in the process of alignment mode, the system's time weaving diagram is shown in Figure 3. Therefore, aspect-oriented time system is divided into two categories. In the chart, time as the current counter value, t as the initial counters value. When the alignment sub-state execute to the determine box, execute different route in accordance with the time systems. Buffer state will continue to deal with if it is overtime, state transforms time is only for overtime operation.
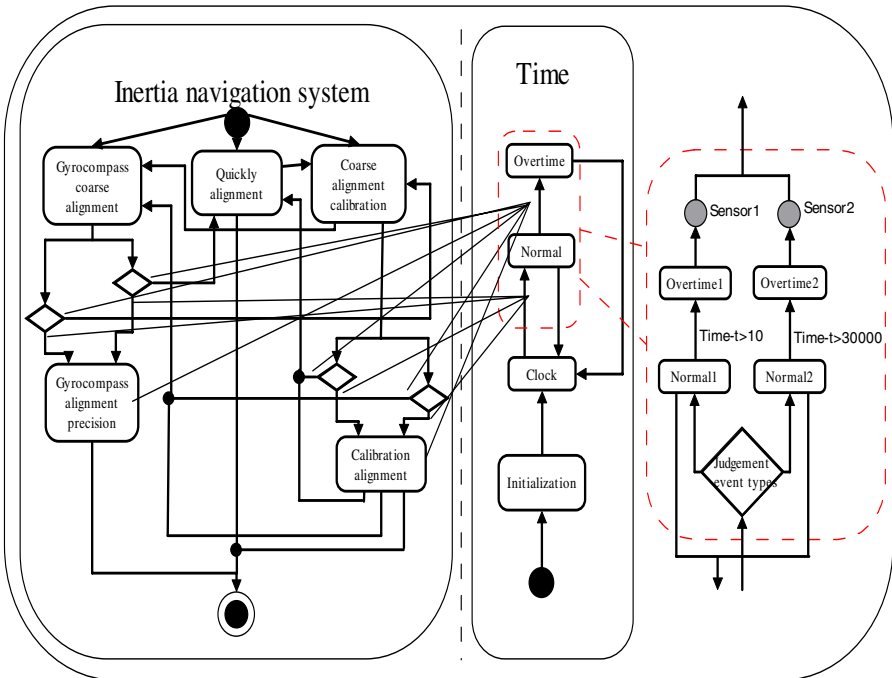
Fig. 3. Alignment sub-mode conversion at the time weaving status

5   Model Validation

Achieve the goal of real-time status monitoring acquisition is a crucial aspect, regardless whether monitoring system-level software status or the unit-level software fault. After obtaining the data of software sensors, each model can analyze the object fault status, and make a conclusion. Under the Rhapsody simulation environment, we simulate the software fault may occurred in INS avionics system, and make real-time monitoring and verification.

Use Rhapsody for software architecture simulation. First of all, it is necessary to known the functional requirements of software architecture, abstract system-level cases and refining a number of modules for unit-level cases. Secondly, according to system-level cases to design the abstract class of software architecture, and create class diagram and state diagram one by one, and define the data structure interaction behavior among the abstract objects. Third, we design a test class for all abstract class object instantiation, run the test behavior, and can obtain C code that can meet the demand of monitoring and software architecture behavior. Finally, Rhapsody will automatically call the integrated operating environment VxWork, and execute C code, and draw a simulation conclusion. Thus, we can build the entire flow simulation, shown in Figure 4.
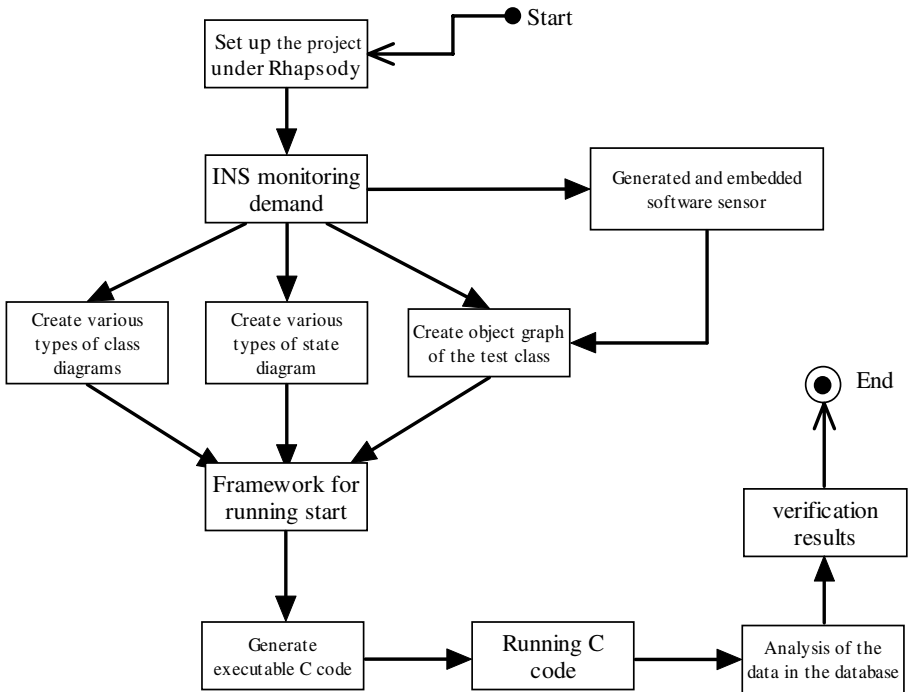
Fig. 4. Simulation verification flow of Inertial navigation system using Rhapsody

5.1   UML Modeling

According to the development vies and behavior view in the UML modeling method, six class in the system set up their own class diagram and state diagram. State diagram is shown in Figure 5.
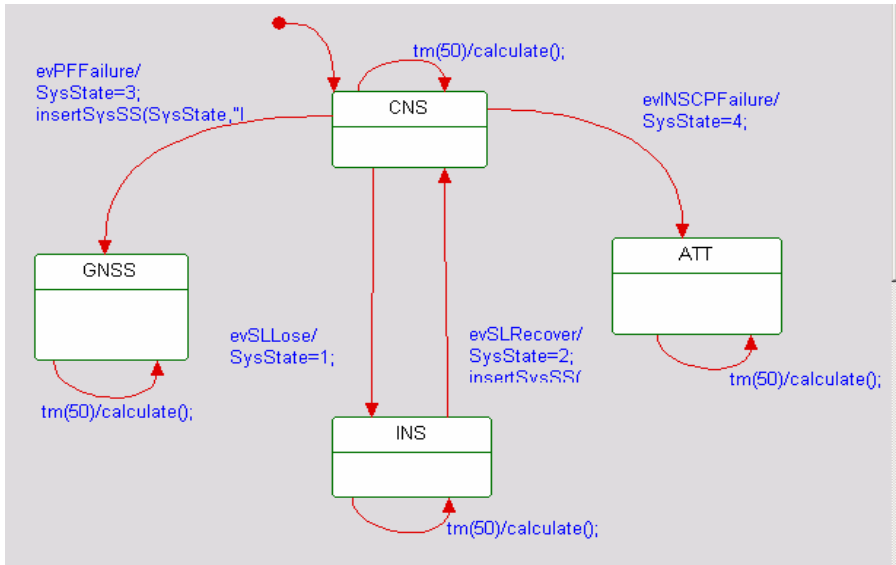
Fig. 5. Status diagram of INS classes in Rhapsody

5.2   Software Sensor Is Designed and Implemented

System inertial/satellite navigation system data acquisition cycle is 50ms, but mode change request 2.5min, so it is can not defined as the fixed way. This system adopts the way of counters; increase one per 5ms in counter, counter occupy two types, the max is 65536; the counter time is about 5.5min. When the counter exceeds the maximum number, data will be cleared. If the starting time is greater than the end of time, it adopts "end time+65536-start time", and obtains time difference. Navigation systems for data collection should not exceed 10; the exchange of alignment sub-mode can not be greater than 30000. Software sensors based on time-oriented are divided into tow categories, SoftwareSensor_Buffer_Timeout(), SoftwareSensor_Transition_Timeout(), and separate the different concerns points, handle them separately. Figure 5 shows Terminal to send data to the cache timeout (category 1).

5.3   Analyze the Results of Software Monitoring

We analyze the database and conclude the monitoring results: I class: normal; II class: Terminal failed to transmit data; III class: Not completed alignment within a specified period of the time. If can data can transmit timely, it has not the problem of real-time

Table 1. Inertial navigation system operating status analysis

NO.	Counter		Time difference	Event name	Event type	Event specification
	Start	End				
1	523	534	55ms	Data overtime	II	Data send not on time
2	42332	42343	55ms	Data overtime	II	Line interruption
3	0	29005	2.4171min	Normal	I	Alignment completed on time
4	0	28694	2.3917min	Normal	I	Alignment completed on time
5	0	30005	2.5004min	Alignment overtime	III	Satellite signal not found
6	0	30004	2.5003min	Alignment overtime	III	Navigation signal abnormal

inadequate. Since this time happed frequently, there is no research necessary. Parts of the running results are shown in Table 1.

6   Concluding

Through simulation Rhapsody and many experimental results show that, INS/GNSS system software condition monitoring and fault detection architecture can be able to detect a complete system of four state job relocation cases and a series of illegal calculated flight path angle case. The real-time software framework based on the aspects successfully solve problem of sensor implanted into the condition monitoring. The next step will focus on the research of complex logic and many software sensors.

References

1. Andy, H.: The Joint Strike Fighter (JSF) Prognostics and Health Management. In: India 4th Annual Systems Engineering Conference, pp. 2799–2813 (2001)
2. Zhang, W.S., Peng, C.N.: Program State Monitoring System. Journal of Tongji University 10, 1385–1390 (2005)
3. Linos, P., Courtois, V.: A Tool for Understanding Object-oriented Program Dependences. In: Proc. 3rd Workshop on Program Comprehension, Washington D C. IEEE Transaction of Software Maintenance, pp. 20–27 (1994)
4. Zakaria, A.A., Hosny, H., Zeid, A.A.: UML Extension for Modeling Aspect-oriented Systems. In: International Workshop on Aspect-Oriented Modeling with UML (2002)
5. Wang, Q.N., Li, Z.L.: Developing Method for HLA Federate Based on Aspect-oriented Technology. Journal of System Simulation 3, 1296–1299 (2007)

Fault Diagnosis of Steam Turbine-Generator Sets Using CMAC Neural Network Approach and Portable Diagnosis Apparatus Implementation

Chin-Pao Hung, Wei-Ging Liu, and Hong-Zhe Su

National Chin-Yi University of Technology, Department of Electrical Engineering,
35, 215 Lane, Sec.1, Chung-Shan Road, Taiping, Taichung, Taiwan
cbhong@ncut.edu.tw, n0915n@yahoo.com.tw,
whitedog175@yahoo.com.tw

Abstract. Based on the vibration spectrum analysis, this paper proposed a CMAC (Cerebellar Model Articulation Controller) neural network diagnosis technique to diagnose the fault type of turbine-generator sets. This novel fault diagnosis methodology contains an input layer, quantization layer, binary coding layer, excited memory addresses coding unit, and an output layer to indicate the fault type possibility. Firstly, we constructed the configuration of diagnosis scheme depending on the vibration fault patterns. Secondly, the known fault patterns were used to train the neural network. Finally, combined with a Visual C++ program the trained neural network can be used to diagnose the possible fault types of turbine-generator sets. Moreover, a PIC microcontroller based portable diagnosis apparatus is developed to implement the diagnosis scheme. All the diagnosis results demonstrate the following merits are obtained at least: 1) High learning and diagnosis speed. 2) High noise rejection ability. 3) Eliminate the weights interference between different fault type patterns. 4) Memory size is reduced by new excited addresses coding technique. 5) Implement easily by chip design technology.

Keywords: Fault diagnosis, Turbine-generator sets, Neural Network, CMAC, PIC, Microcontroller.

1 Introduction

Applications of intelligent fault diagnosis scheme in power equipment have become an important topic. Turbine generator sets are important devices in power systems. There are more than one thousand generator sets operating in China. To improve the reliability of power supply systems, it is of great importance to recognize the incipient failure of generator sets as early as possible, so that the people can maintain them early and avoid the inestimable economic losses.

Due to complexities, coupling and uncertain factors on the structure of turbine generators, the fault diagnosis is difficult to be obtained by theoretic analysis or mathematical model [1]. Over the past decade, intelligent methodology, such as the fuzzy theorem, neural network, and genetic algorithms, had been applied to many fields successfully, including the control, system identification, and pattern recognition. Recently, intelligent theorem had also been applied to the fault diagnosis

of different systems, such as the fault diagnosis of power transformer using fuzzy logic [2-5], the expert system [6-7], and the neural network [8-10]. Also, the fault diagnosis of relay and breaker system and power systems using the neural network theorem had been demonstrated with better performance [11-12]. In the field of turbine generators, the major fault diagnosis scheme is based on the vibration feature of rotary machines [13-14]. Combining the intelligent scheme and vibration feature to improve the diagnosis accuracy of turbine generators, the scholars of mainland China proposed the fuzzy based scheme [1], multiple layer neural network based method [15] and wavelet neural network based technique [16]. Some test results were obtained and outperformed than traditional schemes.

The success of fuzzy rule based schemes strongly depends on the fault patterns and the well if-then rules. To build fuzzy rules needs the expert's specialty or case by case trial and error. Different systems need different experts. The development and maintenance costs are expensive. The multiple layer perceptron (MLP), using error back propagation to train the neural network, can learn the nonlinear mapping relation and achieve the desired diagnosis results. However, the local minimum problem, slower learning speed and the weights interference between different fault patterns are its major drawbacks. Wavelet neural network had been applied on pattern recognition, such as voice and image recognition, in 1992, and obtained successful results [17-18]. Recently, due to quicker convergence speed and higher noise rejection ability, wavelet neural network attracted more attention again. The application of fault diagnosis on turbine generator appeared in [16]. In [16], we see the wavelet neural network must consider many parameters' initialization which is related to the success of wavelet neural network. It is not a simple network and many initial parameters must be decided in advance. Moreover, the iterative training times are more than 500 times for the generator patterns in [16]. The learning time is long and can not be applied to on line training efficiently.

To solve the drawbacks described above, we proposed a CMAC neural network based diagnosis scheme and some results had appeared in [19]. Extending this research, we discussed more topics about the learning speed, memory size, coding scheme and the relation of them in this paper. Depending on the known fault types, the CMAC neural network (CMAC NN) diagnosis structure is built first. By using the known fault patterns as the training data to train CMAC NN, the trained neural network can be used to diagnose the mechanical faults of turbine generators. The proposed excited memory addresses coding scheme simplified the memory size decision and made the diagnosis scheme is easy to be implemented by chip technology. A portable diagnosis apparatus was implemented in this paper also.

2   The CMAC Based Fault Diagnosis System for Turbine-Generator Sets

In 1975, Albus proposed the CMAC neural network and applied it to the robotic manipulator control [20]. Similar to the model of human memory, CMAC performs local reflexive processing. In a table look-up fashion, it produced a vector output in response to state vector input. Local reflexive processing implies only excited

memory cells must be training or tuning. Differ to other neural networks usually must tune all the weights no matter in training mode or diagnosis mode. Therefore, CMAC always has quickly learning speed and can be applied to on line control system with intelligent learning scheme as described above.

Applied it to the fault diagnosis systems such as [21-22], the architecture of separately memory layers avoids the interference between different fault type training and enhances the learning speed also. The successful results demonstrated the quick training speed further. These advantages enhance our confidence to apply it to the diagnosis of turbine generator sets.
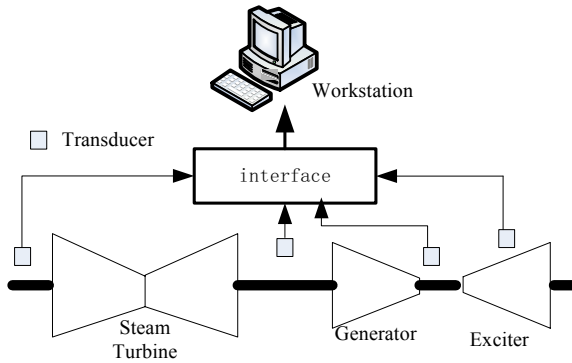
Fig. 1. Schematic diagram of data collection system for steam turbine generator

Table 1. Fault patterns of turbine generator sets

No.	<0.4f	0.4f~0.5f	1f	2f	3f	>3f	RF
1	3.247	46.525	11.908	1.638	2.164	3.247	F1
2	4.348	48.59	10.82	2.715	1.082	2.262	F1
3	3.264	49.772	11.361	1.082	0.541	1.082	F1
4	5.41	45.985	11.912	3.364	1.182	0.552	F1
5	5.951	45.449	15.148	3.246	2.163	2.705	F1
6	1.082	3.345	53.118	6.492	2.263	2.165	F2
7	2.163	1.089	54.16	4.328	0.549	1.623	F2
8	0.542	2.709	48.596	6.498	3.787	1.085	F2
9	0.531	1.633	51.936	5.959	3.246	1.623	F2
10	1.087	1.182	49.772	4.328	0.562	2.168	F2
11	1.623	1.087	22.181	23.804	18.935	8.175	F3
12	0.521	1.282	29.755	21.645	16.23	5.33	F3
13	0.517	1.623	21.64	25.968	17.315	10.829	F3
14	1.085	0.549	24.345	22.181	15.148	15.689	F3
15	0.547	1.096	25.968	25.968	20.478	18.361	F3

The header above spans: Input data (vibration frequency spectrum) over columns <0.4f through >3f.

Note: F1: Oil- membrane oscillation.   F2: Unbalance.
      F3: No orderliness.   RF: Real fault type.
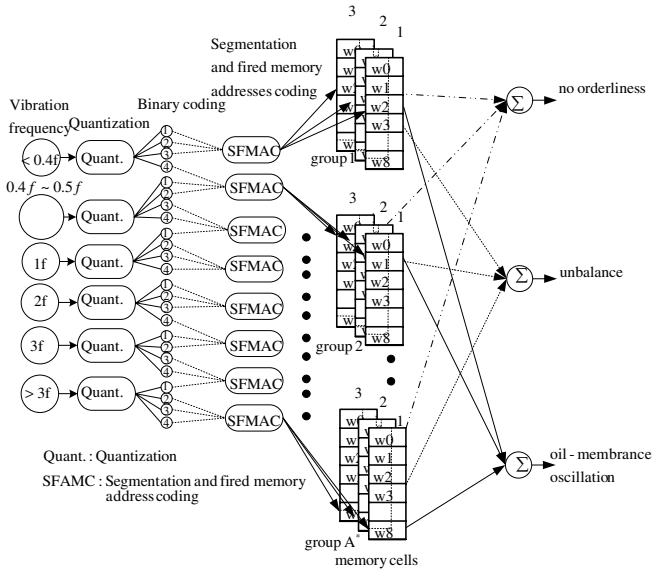
Fig. 2. Turbine generator fault diagnosis system using CMAC neural network

Generally, fault pattern collection is tedious, difficult and arduous work. The researchers must measure, record and analyze the detection signal though a long time. Figure 1 shows the fault collection system for large generator system [23], it concludes a workstation, interface card and vibration transducer. In laboratory, the researchers usually built a simulation model, which must be similar to real system and can reflect the features of real system [24]. And the researchers can focus the research on the data analysis and the development of new diagnosis scheme. In this paper, the necessary fault patterns are list as Table 1 which obtained from [1, 15-16] directly.

In Table 1, the turbine generator sets have three fault types and six vibration frequency detection signals. Symbol f denotes the rotary frequency of the generator. Therefore, the built diagnosis system is shown as Figure 2, including an input layer with six input detection signals, three parallel memory layers, and three output nodes. The operations of the proposed diagnosis scheme include training mode and diagnosis mode. Details are illustrated as follows.

2.1 Training Mode

In training mode, the patterns of fault type k (k=1,2,3) are used to train the memory layer k which memorizes the feature of fault type k only. Inputting the fault patterns to the diagnosis system, via a series of mappings, including quantization, binary coding, segmentation, excited memory addresses coding and summation of the excited memory addresses weights, the CMAC will produce an output value. Assuming the output value is trained equal to 1 to denote a specific fault type, that is the node output 1 confirms what fault type is. Then inputting the diagnosed data to the diagnosis system, the node output can express the possibility of the specific fault type.

2.1.1   Quantization

As shown in Figure 2, the input signals are first put through the quantization mapping to produce a quantization level output. The quantization output can be described as follows [22].

$$q_i = Q(x_i, x_{i\min}, x_{i\max}, q_{i\max}), \quad i = 1, \cdots, n_x, \tag{1}$$

where n_x is the number of input signal. The resolution of this quantization depends on the expected maximum and minimum inputs, $x_{i\max}$ and $x_{i\min}$, and on the number of quantization levels, $q_{i\max}$. To simplify the quantization process, here we consider the $q_{i\max}$ as 15. That is for each input frequency spectrum will be quantized as 0 to 15. In Table 1, the bounds of six frequency spectra, $[x_{i\min}, x_{i\max}]$, are [0.517,5.951], [0.549,49.772], [10.820,54.16], [1.082,25.968], [0.541,20.478], [0.552,18.361], respectively. As shown in the quantization mapping diagram of Figure 3, the quantization level of each input signal can be calculated as

$$q_i(x_i) = ceil((x_i - x_{i\min})/[(x_{i\max} - x_{i\min})/(q_{i\max} - 1)]), \quad i=1,2,\ldots n_x, \tag{2}$$

where ceil(x), the instruction of Matlab, rounds the elements of x to the nearest integers towards infinity.
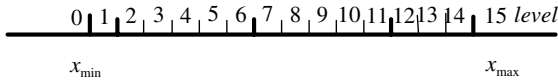
Fig. 3. Quantization mapping diagram

For example, in the first row of Table 1, the quantization level of each input signal is calculated as

$q_1 = ceil((3.247 - 0.517)/[(5.951 - 0.517)/14]) = 8 = 1000\text{B}$

$q_2 = ceil((46.525 - 0.549)/[(49.772 - 0.549)/14]) = 14 = 1110\text{B}$

$q_3 = ceil((11.908 - 10.82)/[(54.16 - 10.82)/14]) = 1 = 0001\text{B}$

$q_4 = ceil((1.638 - 1.082)/[(25.968 - 1.082)/14]) = 1 = 0001\text{B}$

$q_5 = ceil((2.164 - 0.549)/[(20.478 - 0.549)/14]) = 2 = 0010\text{B}$

$q_6 = ceil((3.247 - 0.552)/[(18.361 - 0.552)/14]) = 3 = 0011\text{B}$

2.1.2   Binary Coding

As described above, concatenating the six quantization levels, we have the following binary string.

100011100001000100100011B

The characteristic of CMAC NN is that similar inputs excite similar memory addresses. The excited memory addresses coding must satisfy such condition. Using the binary string will benefit the excited memory addresses coding described below and reduce the memory space. It is different from the traditional coding scheme as described in [25].

2.1.3 Segmentation, Excited Addresses Coding and Output Mapping

Take three bits of the last 24 bits string as a segment (group) and rewrite it as follows.

100011100001000100100**011**B

Then from LSB (least significant bit) to MSB (most significant bit) the excited memory addresses are coded as a1=011B=3, a2=100B=4, a3=100B=4, a4=000B=0, a5=001B=1, a6=100B=4, a7=011B=3, a8=100B=4. It implies that the excited memory addresses number, A^*, is eight. The feature of the specific fault type will be distributed stored on the eight excited memory addresses. To add the weights of excited memory addresses, w_1^3, w_2^4, w_3^4, w_4^0, w_5^1, w_6^4, w_7^3, w_8^4 ,, will produce the CMAC output. The output of CMAC can be expressed as

$$y = \sum_{i=1}^{A^*} w_i^j \ , i=1,...,A^*, \tag{3}$$

where w_i^j denotes the weight of the j-th addresses of group i and j the excited address of group i.

2.1.4 Learning Rule

Assuming the memory layer k (k=1,2,3) output 1 denotes the fault type k is confirmed, then 1 can be thought as the teacher and the supervised learning algorithm can be described as [22,25]

$$w_{i(new)}^j \leftarrow w_{i(old)}^j + \beta \frac{y_d - y}{A^*}, \qquad i = 1,2,..., A^*, \tag{4}$$

where $w_{i(new)}^j$ are the new weight values of group i after the weights tuning, $w_{i(old)}^j$ are the old weight values of group i before weight tuning, and j the excited memory addresses, β the learning gain, $y_d = 1$ the desired output.

2.1.5 Learning Convergence and Performance Evaluation

From [26], the convergence of a supervised learning algorithm can be guaranteed. Assuming the k-th (k=1,2,3) layer output 1 denotes the system has fault type k, and the number of training patterns is n_p. Let the performance index be

$$E = \sum_{n=1}^{n_p} (y_n - 1)^2, n = 1,2,...,n_p , \tag{5}$$

when $E < \varepsilon$ the training process will stop. (ε is a small positive constant).

2.2 Diagnosis Mode

When the training mode is finished, the diagnosis system can be used to diagnose the fault type of a turbine generator. Inputting the diagnosis data to the diagnosis system, the operations of CMAC NN are same as the training mode. But in diagnosis mode, the same excited memory addresses weights of every memory layer are summed up and each layer has one output value. If the input signal is the same as the training

patterns of fault type k, it will excite the same memory addresses of layer k and layer k's output near 1 denotes the exactly fault type. But other layer's output, generally, far away from 1 expresses a low possibility of fault type $j(j \neq k)$. Multiple layers output near 1 value express that multiple fault types happened. Therefore, the proposed diagnosis scheme suits to multiple faults diagnosis naturally. Other features, such as high noise rejection and the memory size reduction results [19].

2.3   Diagnosis Algorithm

In summary, the diagnosis algorithms are described as follows.

2.3.1   Training Mode
step 1. Build the configuration of CMAC fault diagnosis system. It includes 6 input signals, 3 parallel memory layers and 3 output nodes.
step 2. Input the training patterns, through quantization, segmentation, excited memory addresses coding, and summation of excited memory addresses weights to produce the node output.
step 3. Calculate the difference of actual output and the desired output (y_d=1) and using equation (4) to update the weight values.
step 4. Train performance evaluation. If $E < \varepsilon$, the training is finished. Save the memory weights. Otherwise, go to step 2.
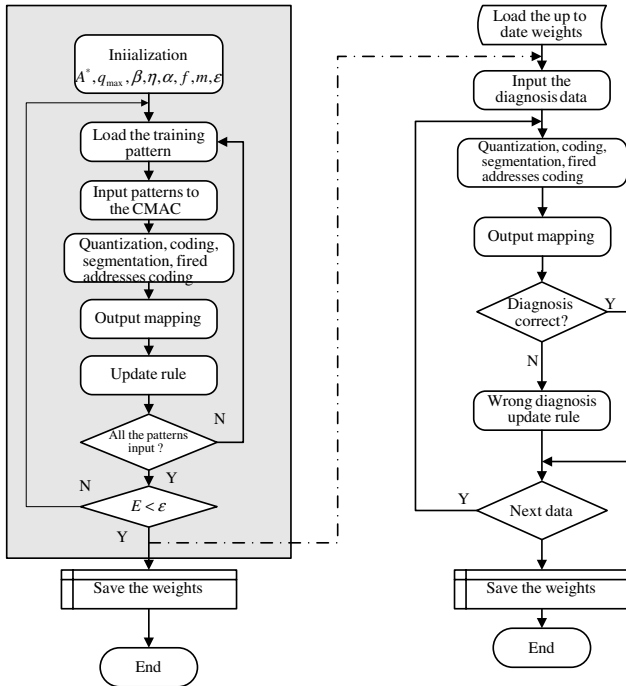
Fig. 4. Flowchart of fault diagnosis system for turbine generator sets

2.3.2 Diagnosis Mode

step 5. Load the up to date memory weights from the saved file.
step 6. Input the diagnosed data.
step 7. Quantization, segmentation, excited memory address coding, and summation of the excited memory weights using equation (3).
step 8. Is the diagnosis correct? Yes, go to step 9. Otherwise, go to step 10.
step 9. Is the next data to be diagnosed? Yes, go to step 6. Otherwise, go to step11.
step 10. Modify the excited memory weights then go to step 9.
step 11. Stop and save the up to date memory weights to file.

In Figure 4, the left hand side represents the off-line mode flowchart and right hand side is the on-line diagnosis mode. The dashed line from left to right denotes the system first time to be started.

3   Case Study and Discussions

To demonstrate the effectiveness of the proposed scheme, the 15 sets patterns of turbine generator sets in [1, 15-16] were tested and list in Table 1. Also, the new excited memory addresses coding scheme with different segmentation bits is tested to make comparison for the memory size, learning speed and correctness. Without loss of generality, we take the bits number m as 3 and 6 respectively and the associated parameters are list in Table 2.

   Using the training patterns of Table 1 to train the CMAC NN (training times is 5), in study 1 and 2. Inputting the training patterns to the CMAC again, Table 3 shows the nodes output of study 1 and 2. As shown in Table 3 output value near 1 denotes the fault type is confirmed. Adding 10% to 50% noise to the original patterns as the diagnosis data, the diagnosis results are shown in Table 4 (partial results). Here, 10% noise means original data multiply (1+10%). In order to obtain larger variation, we didn't take pseudo-random numbers, i.e. (1+10%*rand( )). As shown in Table 4, the node output, near 1, still confirms what the fault type is, and other outputs indicate the possibility of other fault types. The diagnosis results demonstrate the feasibility, high accuracy and high noise rejection ability. Moreover, the learning speed outperforms

Table 2. The CMAC NN parameters of study 1 and 2

Stud	n_x	m	n_b	A^*	n_f	β	q_{max}	ε
1	6	3	24	8	3	0.9	15	0.001
2	6	6	24	4	3	0.9	15	0.001

Table 3. The diagnosis output of training patterns. (partial results)

No	Node output(m=6)			Node output(m=3)			Node output
	F1	F2	F3	F1	F2	F3	
1	**1.0127**	0.3960	0	**0.9858**	0.4211	0.1765	F1
2	**0.9996**	0.3109	0	**0.9940**	0.3472	0.0839	F1
6	0.1874	**1.0009**	0.3450	0.3769	**1.0249**	0.4764	F2
10	0.5645	**1.0000**	0.3450	0.5724	**1.0000**	0.4929	F2
11	0	0.2453	**1.0000**	0.2100	0.4675	**1.0106**	F3
12	0	0.2453	**1.0060**	0.2326	0.4675	**1.0306**	F3

Table 4. The diagnosis results with 10%~50% noise (m=3).

No.	Test data						Node output			AF
	<0.4f	0.4f~0.5f	1f	2f	3f	>3f	F1	F2	F3	
1	*3.247*	46.525	11.908	1.638	2.164	3.247	0.9858	0.4211	0.1765	F1
2	4.348	48.59	10.82	2.715	1.082	2.262	0.9940	0.3472	0.0839	F1
6	1.082	*3.345*	53.118	*6.492*	2.263	*2.165*	0.3001	0.9289	0.5405	F2
10	1.087	1.182	49.772	4.328	0.562	2.168	0.5724	1.0000	0.4929	F2
11	1.623	1.087	*22.181*	23.804	*18.935*	8.175	0.2100	0.4675	1.0106	F3
12	*0.521*	1.282	29.755	*21.645*	16.23	5.33	0.2326	0.4675	1.0840	F3
noise	10%	-20%	30%	+50%	20%	-30%				

Note: Bold type represents with the last row percentage noise.

than traditional methods [1, 15-16]. The learning times of the wavelet neural network based scheme [15] is more than 500 times but CMAC based scheme no more than 5 times under $\varepsilon = 0.001$.

In these study we also found the contradiction between the memory size and learning speed with different group coding bits m and shown as Figure 5. But in a real implementation, it is easy to make a decision between the memory size and learning speed consideration. In an on-line learning operation, the learning speed is the most significant factor and the larger memory size is inevitable; whereas in diagnosis chip design, the memory size affects the possibility of the proposed diagnosis algorithm and the engineer must evaluate the memory size first to match the chip device.

(a)   (b)
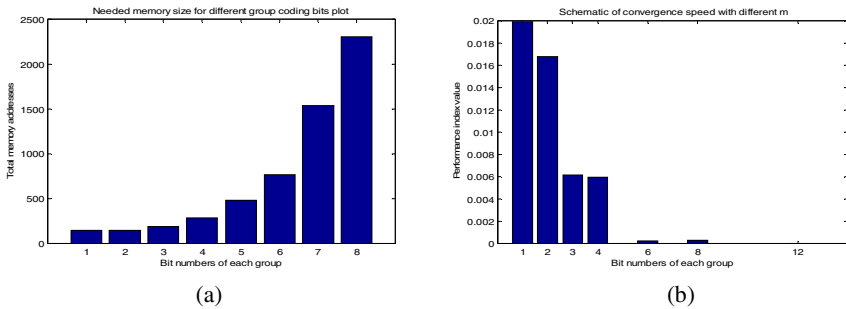
Fig. 5. (a) Memory size and (b) convergence speed with different group coding bit

4   Portable Diagnosis Apparatus Implementation

As shown in Figure 4, the diagnosis program generally is running on a personal computer. Combined with the network communication, it will benefit the remote diagnosis in future. However, a portable diagnosis apparatus is necessary sometimes. Therefore, we also implemented the diagnosis scheme on a microcontroller based diagnosis apparatus. It is composed of a PIC18F8X20 CPU, an LCD display unit, and a 4X4 keyboard input unit. Differ to above, we rewrote the C++ program and used C18 compiler to compile it as the standard PIC18 code. The finished diagnosis apparatus is shown as Figure 6(a). To demonstrate the validity of the portable apparatus, we input the same diagnosis data, as shown in Table 3, into the apparatus and

(a)   (b)   (c)
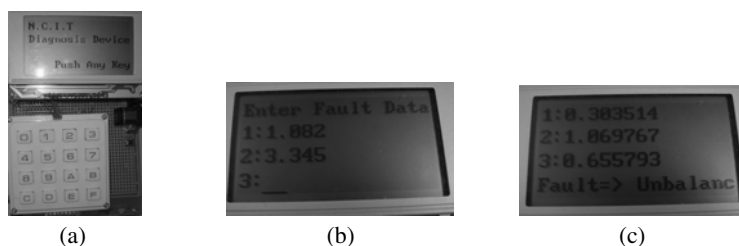
Fig. 6. Photos of diagnosis apparatus

compared its output with the original results. Such as inputting the 6th row data of Table 1 into the apparatus, as shown in Figure 6(b), and the output of Figure 6(c) indeed shows the results are nearly identical to Table 3.

5   Conclusions

This work presents a novel CMAC-based fault diagnosis system for turbine generator systems. Using limited training patterns to train the CMAC neural network, like the brain of human beings, each fault type feature is distributed and memorized on an assigned memory layer. When diagnosed data is input to the CMAC, the diagnosis system will output the possibility of all fault types. It provides useful information to system fault diagnosis and maintenance. As the accumulation of training patterns and learning increases, the diagnosis system will become a more powerful and accurate diagnosis tool. The simulation results demonstrate the proposed scheme with high diagnosis accuracy, high noise rejection ability, and high learning speed. Moreover, the proposed scheme reduced the memory size efficiently and easily implemented by chip technology.

References

1. Li, H., Sun, C.X., Hu, X.S., Yue, G., Wang, K.: The Fuzzy Inputting and Outputting Method in Vibration Fault Diagnosis of Steam Turbine-generator Set. Journal of Chongqing University (China) 22(6), 36–40 (1999)
2. Zhang, G., Yasuoka, K., Ishii, S.: Application of Fuzzy Equivalent Matrix for Fault Diagnosis of Oil-immersed Insulation. In: Proceeding of the 13th International Conference on Dielectric Liquids (ICDL 1999), pp. 400–403 (1999)
3. Su, Q., Mi, C., Lai, L.L., Austin, P.: A Fuzzy Dissolved Gas Analysis Method for the Diagnosis of Multiple Incipient Faults in a Transformer. IEEE Trans. on Power Systems 15(2), 593–598 (2000)
4. Dukarm, J.J.: Transformer Oil Diagnosis Using Fuzzy Logic and Neural Networks. In: Canadian Conference on Electrical and Computer Engineering, pp. 329–332 (1993)
5. Yang, H.T., Liao, C.C., Chou, J.H.: Fuzzy Learning Vector Quantization Networks for Power Transformer Condition Assessment. IEEE Trans. Dielectr. Insulation 8(1) (2001)
6. Wang, Z., Liu, Y., Griffin, P.J.: A Combined ANN and Expert System Tool for Transformer Fault Diagnosis. IEEE Trans. on PWRD 13(4), 1224–1229 (1998)

7. Lin, C.E., Ling, J.M., Huang, C.L.: An Expert System for Transformer Fault Diagnosis Using Dissolved Gas Analysis. IEEE Trans. on PWRD 8(1), 231–238 (1993)
8. Zhang, Y., Ding, X., Liu, Y., Griffin, P.J.: An Artificial Neural Network Approach to Transformer Fault Diagnosis. IEEE Trans. on PWRD 11(4), 1836–1841 (1996)
9. Yanming, T., Zheng, Q.: DGA Based Insulation Diagnosis of Power Transformer via ANN. In: Proceeding of the sixth Conference on Properties and Applications of Dielectric Materials, pp. 133–136 (1999)
10. Thang, K.F., Aggarwal, R.K., Esp, D.G., McGrail, A.J.: Statical and Neural Network Analysis of Dissolved in Power Transformers. In: IEE Dielectric Materials, Measurements and Applications Conference Publication, 473 (2000)
11. Yang, H.T., Chang, W.T., Huang, C.L.: A New Neural Networks Approach to on-line Fault Section Estimation Using Information of Protective Relays and Circuit Breakers. IEEE Trans. on PWRD 9(1), 220–229 (1994)
12. Wang, M.H., Chao, K.H., Hung, F.C.: The Fault Diagnosis of Power Station Using Neural Network. In: Proceeding of 86 Conference of Technology School on Designated Topic, Taiwan, pp. 102–111 (1997)
13. Mitchell, J.S.: Machinery Analysis and Monitoring. Pennwell Publishing Company (1981)
14. Umemara, S.: A New Health Monitoring System for Rotating Machinery. Mitsubishi Juke Giho 18(6), 71–78 (1981)
15. Li, H., Sun, C.X., Liao, R.J., Chen, W.G., Hu, X.S.: Improved BP Algorithm in Vibration Failure Diagnosis of Steam Turbine-generator Set. Journal of Chongqing University (China) 22(5), 47–52 (1999)
16. Li, H., Sun, C.X., Hu, X.S., Yue, G., Tang, N.F., Wang, K.: Study of Method on Adaptive Wavelets for Vibration Fault Diagnosis of Steam Turbine-generator Set. Journal of Electrical Engineering (China) 15(3), 57–60 (2000)
17. Szu, H., Brian, T.: Neural network adaptive wavelets for signal representation and classification. Optical Engineering 31(9), 1907–1916 (1992)
18. Brian, T., Szu, H.: Adaptive Wavelets Classification of Acoustic Backscatter and Imagery. Optical Engineering 30(7), 2192–2202 (1994)
19. Hung, C.P., Wang, M.H., Cheng, C.H., Lin, W.L.: Fault Diagnosis of Steam Turbinegenerator Using CMAC Neural Network Approach. In: Proceeding of International Joint Conference on Neural Network Publication, vol. 334 (2003)
20. Albus, J.S.: A New Approach to Manipulator Control: the Cerebeller Model Articulation Controller (CMAC)[1]. Trans. ASME J. Dynam., Syst. Meas. and Contr. 97, 220–227 (1975)
21. Hung, C.P., Wang, M.H.: Fault Diagnosis of Air-conditioning System Using CMAC Neural Network Approach. In: Advances in Soft Computing—Engineering, Design and Manufacturing, pp. 1–11. Springer, Heidelberg (2003)
22. Hung, C.P., Wang, M.H.: Diagnosis of Incipient Faults in Power Transfers Using CMAC Neural Network Approach. Electric Power Systems Research 71, 235–244 (2004)
23. Zhu, J.N., Miao, S.E., Wang, C.Q.: Implementation of Long-distance Vibration Monitor and Fault Diagnosis System for High Output Turbo-generator Units. Turbine Technology (China) 45(1), 5–7 (2003)
24. Chen, C.: The Study and Application of Turbo-generator Vibration Fault Diagnosis Based on Multi-symptom Input Neural Network Method. Master dissertation of Zhejiang University (2006)
25. Handeiman, D.A., Lane, S.H., Gelfand, J.J.: Integrating Neural Networks and Knowledge-based Systems for Intelligent Robotic Control. IEEE Control Syst. Mag., 77–86 (1990)
26. Wong, Y.F., Sideris, A.: Learning Convergence in the Cerebellar Model Articulation Controller. IEEE Trans. on Neural Network 3(1), 115–121 (1992)

The Fault Diagnosis of Analog Circuits Based on Extension Theory

Meng-Hui Wang, Yu-Kuo Chung, and Wen-Tsai Sung

Department of Electrical Engineering National Chin-Yi University of Technology
35 Lane 215 Chung-Shan Rd. Sec. 1 Taiping City, Taichung County 411 Taiwan, R.O.C.
`wangmh@ncut.edu.tw`

Abstract. This paper proposed a new fault diagnosis method based on the extension theory for analog circuits. The responses of an analog circuit were difference at some node with the normal and failure conditions. However, the identification of the faulted location was not easily task due to the variability of circuit components. So this paper presented a novel extension method for fault diagnosis of analog circuit, which is based on the matter-element model and extended relation functions. The proposed method has been tested on a practical analog circuit, and compared with the multilayer neural network (MNN) based methods and k-means classification method. The application of this new method to some testing cases has given promising results.

Keywords: Analog circuit; Fault diagnosis; Extension theory (ET).

1   Introduction

Faults in analog circuits are classified into two categories: catastrophic faults (hard faults) and parametric faults (soft faults) [1]. The hard faults are due to break of circuit component, and soft faults are due to a variation of one (or more) circuit component values over the tolerance range and deviation of about 50% of the faulty element from their nominal value [2-3]. The deviation of the component condition is complex, the poor fault models, component tolerances, nonlinear effects [4-5]. Therefore, the soft fault is difficult to detect in electric circuits, because soft faults do not change the circuit topology, but cause the circuit to operate outside its specifications. Hence, there were many fault diagnosis methods using artificial intelligence (AI) techniques to solve the problems in the past [6-7].

In recent years, extension theory (ET) has proposed practical applications on different applications for fault diagnosis [8-9]. Cai originally created the concept of extension theory to solve contradictions and incompatibility problems in 1983 [10]. In this world, there are some problems that cannot be directly solved by given conditions, but the problem may become easier or solvable through some proper transformation. For example, the Laplace transformation is one of the commonly used techniques in engineering fields and the concept of fuzzy sets is a generalization of well-known standard sets to extend application fields. Therefore, the concept of an extension set is to extend the fuzzy logic value from [0,1] to (-∞,∞), which allows

us to define any data in the domain and has given promising results in many fields [10-13]. Therefore, this paper proposed a new method using the extension theory and develops a fault diagnosis scheme for soft fault of analog circuit. The proposed extension diagnosis method is based on a matter-element model and extended set. The first step of the extended fault diagnosis method is to formulate matter-element models of fault types according to the simulated results. At the diagnosis stage, the response of the circuit under test (CUT) is measured and compared to the matter-element models, and then the faults of analog circuit can be directly identified by the degrees of extended relational functions. Results from applications to some testing cases show that the proposed method was faster and as a practical solution.

2   Review of Extension Theory

In the standard set, an element either belongs to or, so the range of the standard set is {0,1}, which can be used to solve a two-valued problem. In contrast to the standard set, the fuzzy set allows for the description of concepts in which the boundary is not explicit. It concerns not only whether an element belongs to the set but also to what degree it belongs to. The range of a fuzzy set is [0,1]. The extension set extends the fuzzy set from [0,1] to $(-\infty,\infty)$. As a result, it allows us to define a set that includes any data in the domain. Extension theory tries to solve the incompatibility or contradiction problems by the transformation of the matter element [14]. The comparisons of the standard sets, fuzzy sets and extension sets are shown in Table 1. Some definitions of extension theory are introduced in the next section.

Table 1. Three different sorts of mathematical sets

Compared item	Standard set	Fuzzy set	Extension set
Research objects	Data variables	Linguistic variables	Contradictory problems
Model	Mathematics model	Fuzzy mathematics model	Matter-element model
Descriptive function	Transfer function	Membership function	Correlation function
Descriptive property	Precision	Ambiguity	Extension
Range of set	$C_A(x) \in (0,1)$	$\mu_A(x) \in [0,1]$	$K_A(x) \in (-\infty,\infty)$

2.1   Matter-Element Theory

(1) Definition of matter-element
Defining the name of matter as N, the characteristics of the matter as c, and the value of c as v, a matter-element in extension theory can be described as follows:

$$R = (N, c, v) \tag{1}$$

Where N, c, and v were called the three fundamental elements of the matter-element. For example, $R = (Wang, Weight, 75\ kg)$ can be used to state that Wang's weight is 75 kg. If the value of the characteristic has a classical domain or a range, then we define the matter-element for the classical domain as follows:

$$R = (N,c,v) = \left(N,c,\left\langle v^{L},v^{U}\right\rangle\right)$$ (2)

where v^{L} and v^{U} are the lower bound and upper bound of a classical domain.

(2) Multi-dimensional matter-element

Assuming $R=(N, c, v)$ to be a multidimensional matter-element, $C=[c_1,c_2,...,c_n]$ to be a characteristic vector and $V=[v_1,v_2,...,v_n]$ to be a value vector of c, then a multidimensional matter-dement is defined as:

$$R = \begin{bmatrix} N & c_1 & v_1 \\ & c_2 & v_2 \\ & \vdots & \vdots \\ & c_n & v_n \end{bmatrix} = \begin{bmatrix} R_1 \\ R_2 \\ \vdots \\ R_{n1} \end{bmatrix}$$ (3)

where $R_i = (N, c_i ,v_i),(i = 1, 2, . . . ,n)$ is defined as the sub-matter-element of R.

2.2 Extension Set Theory

(1) Definition of extension set

Let U be the universe of discourse, then an extension set $\tilde{A}$ on U is defined as a set of ordered pairs as follows:

$$\tilde{A} = \left\{(x, y)\,\middle|\, x \in U, y = K(x)\in (-\infty,\infty)\right\}$$ (4)

Where $y = K(x)$ is called the correlation function for extension set $\tilde{A}$. The $K(x)$ maps each element of U to a membership grade between $-\infty$ and ∞. The higher degree and the more elements belong to the set. In a special condition, when $0 \le K(x) \le 1$, it corresponds to a normal fuzzy set. $K(x) \le 1$ implies that the element x has no chance to belong to the set. When $-1 < K(x) < 0$, it is called an extension domain, which means that the element x still has a chance to become part of the set.

(2) Definition of correlation function

The correlation functions have many forms dependent on application. If we set $X_o=<a, b>$ and $X=<c , d>$ are two intervals in the real number field, and $X_o \in X$, then the correlation function in the extension theory can be defined as follows:

$$K(x) = \frac{\rho(x, X_o)}{D(x, X_o , X)}$$ (5)

where

$$D(x, X_o, X) = \begin{cases} \rho(x, X) - \rho(x, X_o) & x \notin X_o \\ -\rho(x, X) & x \in X_o \end{cases} \quad (6)$$

$$\rho(x, X_o) = \left| x - \frac{a+b}{2} \right| - \frac{b-a}{2} \quad \rho(x, X) = \left| x - \frac{c+d}{2} \right| - \frac{d-c}{2} \quad (7)$$

The correlation function can be used to calculate the membership grade between x and X_o as shown in Fig. 1 when $K(x) \geq 0$. It indicates the degrees to which x belongs to X_o. When $K(x) < 0$ it describes the degree to which x does not belong to X_o. When -$1 < K(x) < 0$, it is called the extension domain, which means that the element x still has a chance to become part of the set if conditions change.
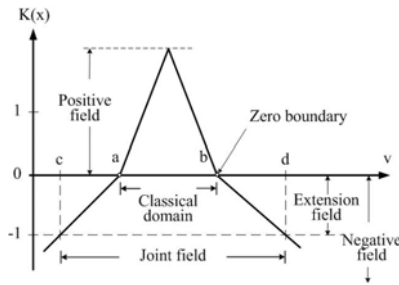
Fig. 1. The proposed extended relation function

3   The Proposed Fault Diagnosis Method

The tested circuit of this research was the second-order low-pass filter as shown in Fig. 2. A low-pass filter is an analog circuit that passes low-frequency signals but attenuates signals with frequencies higher than the cutoff frequency. It is called a high-cut filter when it used in audio applications. In this paper, the DC gain of the
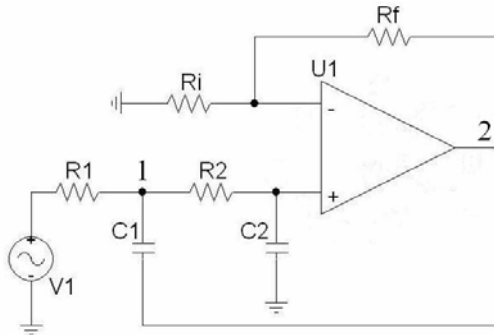
Fig. 2. The second-order low-pass filter

Table 2. The nominal value of low-pass filter

Components	Nominal values
R1	100KΩ
R2	85KΩ
Ri	2.2MΩ
Rf	0.1KΩ
C1	5nF
C2	2nF

filter is set to 1 and the cutoff frequency is set to 1 kHz. Node 1 and 2 are the testing node where voltage can be measured or simulated. A list of the nominal values of components is presented in Table 2.

3.1 The Various Situations of Analog Circuits

The CUT was simulated both at normal and all faulty conditions by using PSPICE software, the tolerance rang of normal was selected deviation of about ±5% of the nominal values, and soft fault was deviation of about ±20%. By analyzing sensitivity of the CUT, R1, R2, C1 and C2 are selected to be the fault components, the pre-selected set of single faults of the test circuit are presented in Table 3.

In this research was discovered that the transient voltage and phase spectrum was different between normal and other faults condition at nodes, Fig. 3, Fig. 4, Fig. 5 and Fig. 6 demonstrated their difference at node1 and node2, the RMS value of transient voltage and phase spectrum can made the pattern of CUT. 80 Monte-Carlo simulations are carried out for each fault scenario, in the training stage; every type of matter-element model was built by 30 pieces of data. And in the testing stage; there are 50 pieces of fault data for each fault recognized.
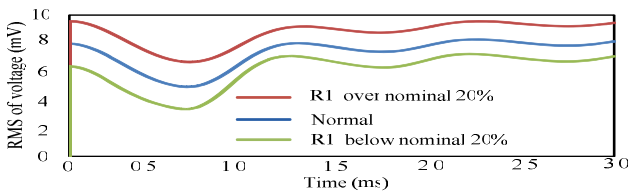
Fig. 3. The RMS value of voltage between normal and soft fault at node 1
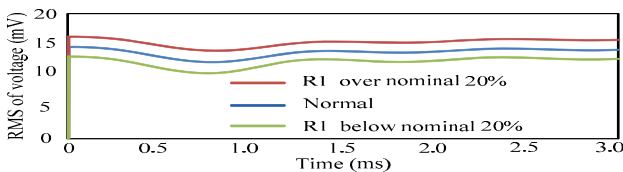
Fig. 4. The RMS value of voltage between normal and soft fault at node 2
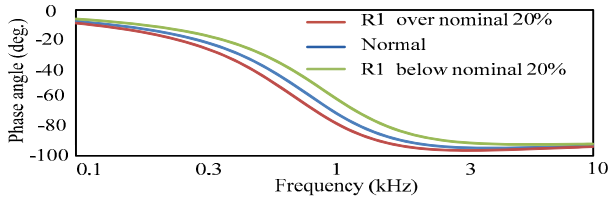
Fig. 5. The phase spectrum between normal and soft fault at node 1
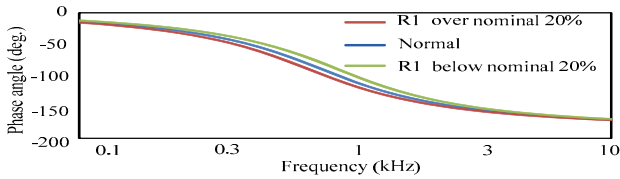
Fig. 6. The phase spectrum between normal and soft fault at node 2

Table 3. A list of the components that tolerance range of normal and single soft faults

Components	Nominal values	Variables ±5% normal	Variables ±20% single soft faults
R1	100 KΩ	105 KΩ 95 KΩ	120 KΩ 80 KΩ
R2	85 KΩ	90 KΩ 80 KΩ	102 KΩ 68 KΩ
C1	5 nF	5.5 nF 4.5 nF	6 nF 4 nF
C2	2 nF	2.2 nF 1.8 nF	2.4 nF 1.6 nF

3.2   The Matter-Element Model of Fault

The fault types divided to 9 kinds(one normal and eight soft fault), and F_i represents i fault type. Moreover, C_i is the features. The neighborhood domain R_P defined as the maximum range of every characteristic is set as follows:

$$R_p = (F, C, V_p) = \begin{cases} F & C_1 & V_1\,min, V_1\,max \\ & C_2 & V_2\,min, V_2\,max \\ & C_3 & P_1\,min, P_1\,max \\ & C_4 & P_2\,min, P_2\,max \end{cases} \tag{8}$$

After establishing the matter-element model based on the extension method, the soft fault can be identification properly.

3.3 The fault Diagnosis Method

The extension fault diagnosis method of the proposed method is shown below:

step 1: Establish the matter-element model of every soft fault category, where Cj is the j-th feature of the Fi, and Vij = <aij, bij> is the classical region of every characteristic set.

step 2: set the matter-element of the circuit under test data as follows:

$$R_t = (F, C, V_t) = \begin{cases} F & C_1 & V_{t1} \\ & C_2 & V_{t2} \\ & C_3 & P_{t1} \\ & C_4 & P_{t2} \end{cases} \tag{9}$$

step 3: Calculate the correlation function of the tested soft fault with the proposed extended correlation function as follows:

$$K_{ij}(v_{tj}) = \begin{cases} \dfrac{-\rho(v_{tj}, V_{ij})}{|V_{ij}|} & if \ v_{tj} \in V_{ij} \\ \dfrac{\rho(v_{tj}, V_{ij})}{\rho(v_{tj}, V_{pj}) - \rho(v_{tj}, V_{ij})} & if \ v_{tj} \notin V_{ij} \end{cases} \tag{10}$$

$$i = 1,2,\ldots,9 \ ; \ j = 1,2,\ldots,4$$

step 4: Assign weights to the soft fault characteristic such as W_{j1}, W_{j2}, W_{j3}, W_{j4}, denoting the significance of every soft fault characteristic.

step 5: Calculate the correlation function of every soft fault category:

$$\lambda_i = \sum_{j=1}^{4} W_{ij} K_{ij} \ , \ i = 1,2,\ldots,4 \tag{11}$$

step 6: Make each correlation functions of every soft fault in <1, -2>, that it's easy to diagnosis.

$$\lambda_i' = \frac{3\lambda - \lambda_{min} - 2\lambda_{max}}{\lambda_{max} - \lambda_{min}} \ , \ i = 1,2,\ldots,9 \tag{12}$$

Which

$$\lambda_{max} = \max_{1 \le i \le 9}\{\lambda_i\} \qquad \lambda_{min} = \min_{1 \le i \le 9}\{\lambda_i\} \tag{13}$$

$\lambda_f' = 1$ were means I_f to be anomalism. The value of correlation degree is closing 1 which would be anomalism.

step 7: Going back to step 2 for the next case when the diagnosis of one has been completed, until all have been done.

4 Experiment Results and Discuss

This paper used 450 tested data to test the practicability of the proposed method. There are 9 kinds of faults, each soft fault of circuits were simulated 50 data, the

partial relation degrees of soft faults are shown in Table 5. It is very easy to diagnose the fault types in circuits from Table 4. For example, in case no.20, the relation degree with the fault type no.1 equals to 1 (or maximum value), which is indicative of fault type no. 1, or normal. In comparison, the relation degrees with other fault types are very small; moreover, the proposed method can not only diagnose the main fault types of analog circuit; it can also provide useful information for future trend analysis by the relation degree.

Table 4. The circuit data with different fault types (Partial samples)

Cases	V_1	V_2	P_1	P_2	Fault no.	Actual fault type
20	7.749	14.434	-93.204	-151.301	1	Normal
70	9.923	16.691	-97.915	-156.404	2	R1 over nominal
120	9.924	16.691	-88.829	-147.318	3	R1 below nominal
187	7.531	16.560	-91.174	-156.134	4	R2 over nominal
236	7.685	11.709	-96.250	-143.379	5	R2 below nominal
285	7.143	13.658	-101.683	-159.625	6	C1 over nominal
335	7.5005	13.853	-84.7055	-142.6476	7	C1 below nominal
383	7.527	14.353	-87.139	-152.612	8	C2 over nominal
429	7.983	13.980	-102.364	-149.876	9	C2 below nominal

The Table 6 shows the correct rate by using the MNN, k-means, and extension based method to diagnosis the soft fault. The maximum correct rate is 92.6% in the MNN, and the correct rate is 84.67% in k-means based method. But the method of the proposed by this paper has 97.6% for correct rate. This verified extension diagnosis method to be batter then other methods.

To test the diagnosis performance of the proposed method, diagnosis accuracy with different noise percentages is shown in Table 7. The sources of error include the tolerances of other normal components which could lead to data uncertainties. To take into account the errors and uncertainties, the testing data was adding ±5% and ±10% random uniform-distributed samples to test the robustness of the proposed method. the proposed method in noise percentage ±5, ±10 was still has a significantly higher diagnosis accuracy of 97.7% and 83.3% than the k-means and MNN, Moreover, the proposed method shows good tolerance to add errors, and has high accuracy of 83.3% in extreme error of ±10%.

Table 5. Relation degrees λf' by the extended diagnosis method (partial results)

NO	1	2	3	4	5	6	7	8	9	Diagnosis result
20	1.00	-1.92	-2.00	-1.77	-1.52	-1.10	-0.34	-0.67	0.10	1
70	-1.57	1.00	-1.29	-1.25	-2.44	-1.05	-2.00	-1.68	-1.49	2
120	-1.69	-0.11	1.00	-1.64	-2.49	-2.00	-1.67	-1.15	-1.68	3
187	-0.99	-0.63	-1.21	1.00	-2.53	-0.96	-1.78	-1.00	-2.00	4
236	-1.16	-1.95	-1.98	-1.72	1.00	-1.72	-0.94	-1.87	-2.00	5
285	-0.72	-1.56	-2.00	-1.11	-1.56	1.00	-1.58	-0.62	-0.61	6
335	0.09	-2.00	-0.89	-0.51	-0.58	0.014	1.00	0.485	-1.37	7
383	0.19	-2.00	-1.20	-0.57	-1.61	-0.64	-0.70	1.00	-0.97	8
429	-0.56	-1.84	-2.00	-1.80	-1.01	-0.18	-0.85	-0.82	1.00	9

Table 6. Diagnosis performances of different methods

Methods	Training time (Epochs)	Accuracy rate (%)
Proposed method	N/A	98.6%
K-means	N/A	84.67%
MNN(4-7-9)	1000	91.3%
MNN(4-8-9)	1000	85.56%
MNN(4-9-9)	1000	92.6%

Table 7. Diagnosis performances of errors added

Noise Percentage (%)	Proposed Method	K-means	MNN
±0	98.6%	88.44%	91.3%
±5	97.7%	87.56%	88.89%
±10	83.3%	74.89%	76.6%

5   Conclusion

This paper presents a novel fault diagnosis method based on extension theory for analog circuits. Compared with other traditional AI methods, the proposed method does not require particular artificial parameters and learning processes. The calculation of the proposed diagnosis algorithm is fast and very simple. It can be easily implemented by PC software. In addition, by the test results, we can see that this method is not only diagnosis of the main soft fault. But it also using correlation function of value recognizes the other soft fault.

References

1. Yongkui, S., Guangju, C., Hui, L.: Analog Circuits Fault Diagnosis Using Support Vector Machine. In: International Conf. on Communications, Circuits and Systems, 2007. ICCCAS 2007, pp. 1003–1006 (2007)
2. Catelani, M., Fort, A., Alippi, C.: A Fuzzy Approach for Soft Fault Detection in Analog Circuits. Meas. 32, 73–83 (2002)
3. Slamani, M., Kaminska, B.: Fault Observability Analysis of Analog Circuits in Frequency Domain. IEEE Trans. on Circuits Systems II, Analog and Digital Signal Proc., 134–139 (1996)
4. Tadeusiewicz, M., Halgas, S., Korzybski, M.: An Algorithm for Soft-Fault Diagnosis of Linear and Nonlinear Circuits. IEEE Trans. on Circuits and Systems, Fundamental Theory and Applications 49(11), 1648–1653 (2002)
5. Catelani, M., Fort, A.: Soft Fault Detection and Isolation in Analog Circuits: some Results and a Comparison between a Fuzzy Approach and Radial Basis Function Networks. IEEE Trans. on Instrumentation and Meas. 51, 196–202 (2002)
6. Yanghong, T., Yigang, H., Chun, C., Guanyuan, Q.: A Novel Method for Analog Fault Diagnosis Based on Neural Networks and Genetic Algorithms. IEEE Trans. on Nstrumentation and Meas. 57(11), 2631–2639 (2008)
7. Spina, R., Upadhyaya, S.: Linear Circuit Fault Diagnosis Using Neuron Morphed Analyzers. IEEE Trans. on Circuits Systems II, Analog Digital Signal Proc. 44(3), 188–196 (1997)
8. Wang, M.H.: A Novel Extension Method for Transformer Fault Diagnosis. IEEE Trans. on Power Delivery 18(1), 164–169 (2002)
9. Wang, M.H.: Application of Extension Theory to Vibration Fault Diagnosis of Generator Sets. IEE Proc. Generation, Transmission and Distribution 151(4), 503–508 (2004)
10. Cai, W.: The Extension Set and Incompatibility Problem. J. of Scientific Exploration 1, 81–93 (1983)
11. Cai, W.: Extension Set, Fuzzy Set and Classical Set. In: First Congress of International Fuzzy System Association, Spain (1985)
12. Li, J., Wang, S.: Primary Research on Extension Control Information System, vol. 1. International Academic Publishers (1991)
13. Huang, Y.P., Chen, H.J.: The Extension-Based Fuzzy Modeling Method and its Applications. In: IEEE Canadian Conf. on Electrical and Computer Engineering, vol. 2, pp. 977–982 (1999)
14. Huang, Y.C., Yang, H.T., Huang, C.L.: Developing a New Transformer Diagnosis System through Evolutionary Fuzzy Logic. IEEE Trans. on Power Delivery 12(2), 761–767 (1997)

Improvement and Light Weight of Twin Seat Underframe in Multiple Unit Train

RenLiang Wang, Hu Huang, XinTian Liu, and LiHui Zhao

College of Automotive Engineering, Shanghai University of Engineering
Science, Shanghai 201620, China
`wangrenliangcelia@163.com`

Abstract. To improve the structure of the twin seat underframe in multiple unit train and lightweight design, FEM is used to analyze this problem. First of all, the 3D geometric model of the existing twin seat underframe is built. And according to the requirements of enterprises and experiment, the finite element model is built and analyzed. Then results of FEA and experiment are compared. According to the comparison, the finite element model that was simplified is effective. On the basis of effective finite element model, five schemes of the twin seat underframe are proposed and structural strength of twin seat underframes that were improved is researched. Finally according to influence on manufacturing procedure of the twin seat underframe and the error of the FEA, the ideal solution is proposed.

Keywords: Multiple Unit Train; U-shaped Load Method; Light Weight.

1 Introduction

With the rapid development of high-speed railway traffic in China, on the technical platform of the speed of 200 km/h, the speed of multiple unit train will reach 300 km/h through independent research and development. At the same time the reliability and safety of the train seat underframe is proposed higher request. There is a very important issue for us to design the train seat underframe that can meet with the requirement of the enterprises on the safety and economy [1].

Relying on the experimental method and researchers' experience, it will be very difficult to analyze the structure performance of the train seat underframe. A lot of financial, material resources and manpower will be required. The effective method to solve the question is using FEM (Finite Element Method) and software technology, and at the same time the cycle of product development will be shorten and much research expenses will be saved. According to the requirements of enterprises, the improved seat underframes will be analyzed and the ideal solution will be proposed through using FEM. [2]

2 Theoretical Backgrounds

According to the flexural rigidity theory in mechanics of materials, the flexural rigidity of seat underframe can be defined as the loads which make the underframe have

specific flexivity. To the underframe longeron, if the middle of span is loaded F, the maximum flexivity value of the longeron is

$$Y_{max} = \frac{Fl^3}{48EJ}$$

Where Y_{max} is the maximum flexivity of the longeron; l is the span of the longeron; J is the bending resistance section factor; E is the elasticity modulus of longeron material; F is the concentrated load added in the middle of longeron. According to requirements of the enterprises, the middle of longeron is loaded 2000N and the maximum flexivity value should be less than 1mm.

3   Establishment of the 3D-Model of the Twin Seat Underframe

According to the existing seat underframe, the 3D-model of the twin seat underframe will be established (Fig.1). Then three schemes of improved longeron sections will be proposed, and according to the different sections, the stress and displacement will be analyzed.
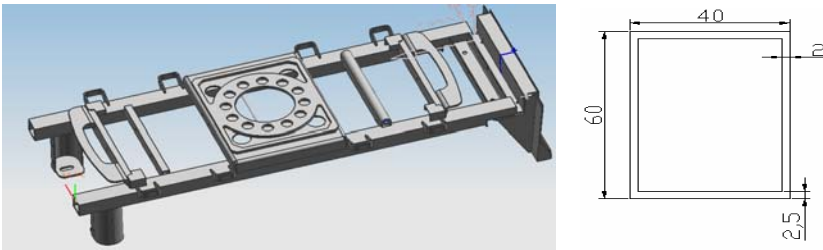
Fig. 1. 3D-model of twin seat underframe (Right picture shows section of the longeron, the size of the section is 2.5×2)

4   Pretreatment of the Twin Seat Underframe Model

According to the actual load condition on the twin seat underframe, the 3D model should be simplified. And according to the actual situation of seat underframe model, most parts are thin-wall structures, so two-dimensional grid can be used to partition model, not only the calculation work can be greatly reduced, but also the precision can be guaranteed. In addition, to some irregular parts, three-dimensional grids can be used and the appropriate size of elements should be chosen.

Through meshed and quality checked, the complete finite element model of twin seat underframe can be acquired (Fig.2). The model was mainly meshed with two-dimensional grid; there are 32 assembly parts, 136624 nodes, 134767 elements, 132970 quad elements and 1797 triangular elements. The proportion of triangular elements is only 1.4%. [3]

Fig. 2. Finite element model of the twin seat underframe

5   Constraint of the Seat Underframe

5.1   Distribution of Material Attributes

According to the requirement of the enterprises, 7003 aluminum is used to manufacture seat underframes. Aluminum's specific gravity is small, approximately one-third of the steel, and the machining performance is good, belonging to the high-strength aluminum alloy. The mechanical performance of the 7003 aluminum is simliar to 7005 aluminum, and both of them are usually applied to frame. The 7003 aluminum's elongation strength is 345Mpa, fatigue strength is 130Mpa, elastic modulus is 74Gpa and Poisson's ratio is 0.3.

5.2   Boundary Conditions

The principle to establish the boundary condition is that the load can be reasonably simplified on the premise that the situation can be truly reflected. In the process of the boundary conditions establishment, the model structure and the purposes of analysis, test data, and other factors should be comprehensively considered. In order to simulate the real situation in this test and make the stress on the seat underframe conform to the reality, the U-shaped load method and the staircase distribution load method will be adopted to analyze.
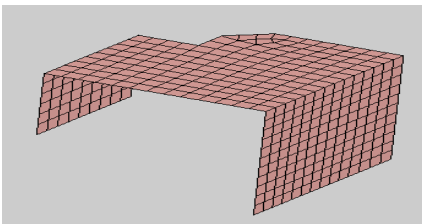
Fig. 3. Around Supporting Part
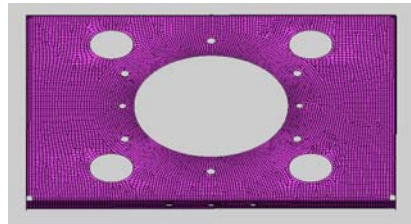
Fig. 4. The Turnable Part
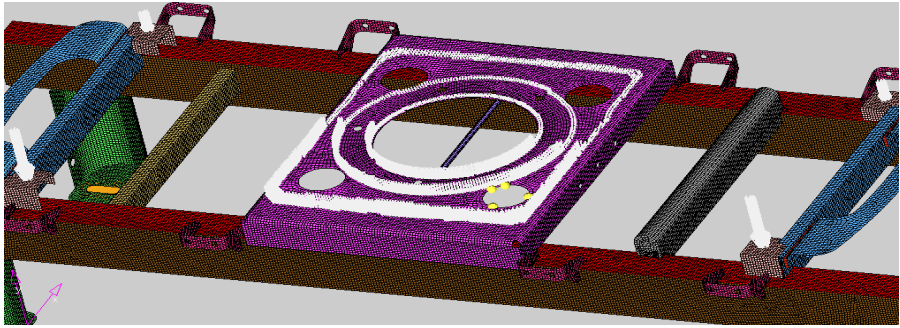
Fig. 5. U-shaped load method on the turnable part

According to the pressure distribution of sitting posture and the experimental data, 2000N is loaded on the twin seat underframe. 200N is distributed on the Around Supporting Part (Fig.3). 1800N is distributed on the Turnable Part (Fig.4).

First load method: U-shaped load method simulate the pressure distribution when passengers completely contact with the seat, according to the pressure distribution of sitting posture and load requirements from the enterprise, 1440N and 360N are added on the Turnable Part. The ratio of 4:1 of the force distributed on the Turnable Part should be guaranteed.
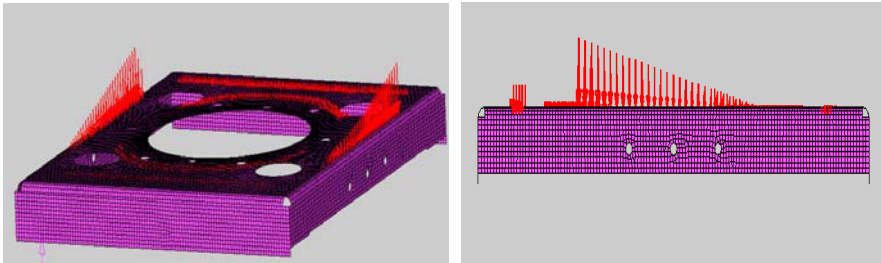
Fig. 6. The load method of staircase distribution on the turnable part

Second load method: the staircase distribution load method simulates the pressure distribution in the instant moment when passengers contact with the seat, the strength of the force become bigger from the middle to both sides and from after to front side. And the ratio of 4:1 of the force distributed on the Turnable Part should also be guaranteed.

The twin seat underframe is connected with floor by screws, so joint between the seat underframe and the floor will be restricted at 6 dof.

6   CAE Simulation of the Twin Seat Underframe

It has the same basic theory when the structural static analysis is calculated through CAE. The object of analysis is node. According to the condition of static equilibrium,

the static equilibrium equations of the structure are established on all nodes, and then put them together to get the balance equation of the structure. The unknown node's displacement can be solved through the equation. [4]

Due to the stress distribution on the seat underframe, the displacement is mainly concentrated in the former longeron. Because the Turnable Part is not fixed in the middle of the longeron, the maximal displacement on the longeron didn't correspond with the installation position of the Turnable Part.
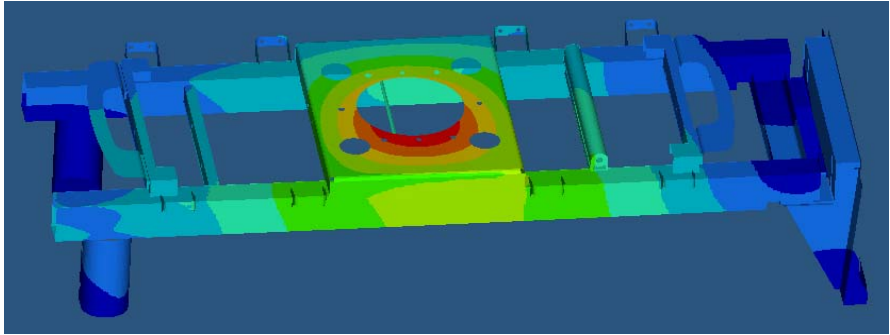
Fig. 7. CAE simulation of twin seat underframe

6.1   Analysis and Verification of the Finite Element Model

In the process of product development and design, how to guarantee the effectiveness of the results of finite element method will directly related to the finite element method whether can correctly guide product development and design.

Selecting the most reasonable model and analysis method is the prerequisite to guaranteeing the accuracy and efficiency of the results. [5]

Table1 summarises the data from CAE simulation and from the experiment for comparison.

Because the error between the data from experiment and the data from simulation calculation is less than 10%, the finite element model of twin seat underframe is reliable. [6]

Table 1. The contrast of enterprises and CAE simulation experiment data

Data sources	Specification (mm)	Displacement (mm)		Stress (Mpa)	
		U-shaped load	Staircase distribution load	U-shaped load	Staircase distribution load
CAE Simulation Experiment	80×40 (5×2.5)	0.43	0.443	54.4	53.2
			0.46		——

6.2   The Analysis of the Improved Model

The first scheme: the original section size of 80×40 is modified to the section size of 60×40, and eight kinds of wall thickness of this section will be analyzed to obtain the data (Table 2).

Table 2. The analysis result of different thickness on the section of 60×40 model

Specification of wall thinkness (mm)	Decreasement of the seat weight (%)	Displacement (mm)		Stress (Mpa)	
		U-shaped load	Staircase distribution load	U-shaped load	Staircase distribution load
5×2.5	6.9%	0.54	0.56	99.7	101.2
2.5×2.5	18.8%	0.75	0.78	106	107
2.5×2	22.5%	0.799	0.82	107	108
2.5×1.5	26.3%	0.86	0.89	107	108
2.5×1	20.7%	0.95	0.97	107	119
2×2	25%	0.89	0.92	109	110
2×1.5	28.8%	0.97	0.99	110	111
2×1	32.5%	1.076	1.12	122	119

In the above table, there are seven schemes that the displacement is less than 1mm. Because there is little error between the data from the experiment and the data from simulation calculation, the selected scheme should have enough redundancy. The schemes that the diaplacement is less than 0.9mm can be selected.

The second scheme: the section size of 80×40 is modified to the section size of 40×40. After analyzing this model, the displacement is still more than 1mm when the wall thickness is 5×2.5, and its economy is not ideal when the wall thickness is 5×5 (Table.3).

Table 3. The analysis result of different thickness on the section of 40×40 model

Specification of wall thinkness (mm)	Decreasement of the seat weight (%)	Displacement (mm)		Stress (Mpa)	
		U-shaped load	Staircase distribution load	U-shaped load	Staircase distribution load
5×2.5	31.1%	1.09	1.12	144	147
5×5	3.4%	0.32	0.35	53	58

The second scheme: the section size of 80×60 is analyzed, in the condition of different thickness, the following analysis results is obtained. In consideration of the safety and economy, two kinds of schemes can be selected (Table 4).

Table 4. The analysis result of different thickness on the section of 80×60 model

Specification of wall thinkness (mm)	Decreasement of the seat weight (%)	Displacement (mm)		Stress (Mpa)	
		U-shaped load	Staircase distribution load	U-shaped load	Staircase distribution load
2.5×1.5	15.1%	0.61	0.65	97.8	101.2
2.5×1	20.2%	0.72	0.74	104	106.7
2×2	13.7%	0.68	0.71	112	115.4
2×1.5	18.9%	0.73	0.78	135	136.9
2×1	24.3%	0.86	0.89	150	152.7
1.5×1.5	22.7%	0.97	0.99	187	189.2

6.3 The Analysis of Stress Concentration

Through the analysis of finite element model, the stress distribution of the model can be obtained. After the comparison of various schemes, the maximum stress appears in the joints between the longeron and the supporting board (Fig.8). The maximum is between 107Mpa and 150Mpa, all the data don't exceed allowable stress of the material, and where the maximun stress appears is in the same place with the experimental result.
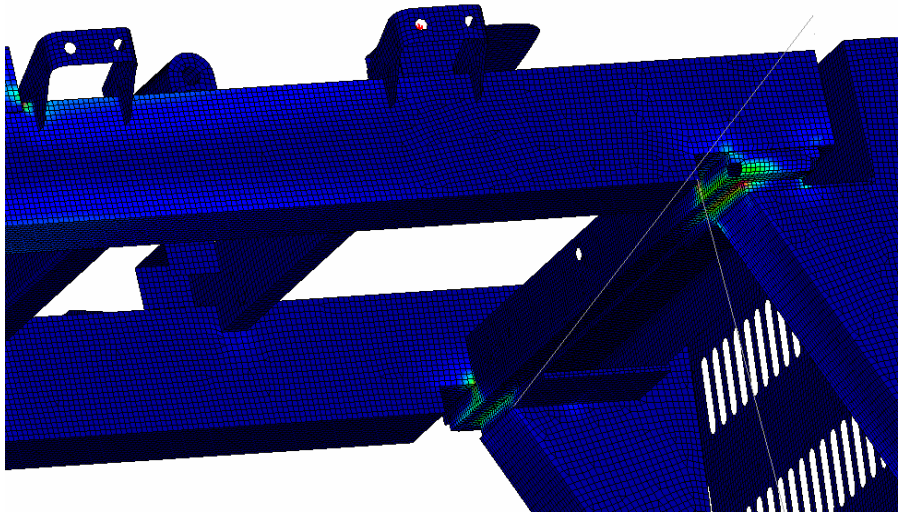
Fig. 8. The stress concentration on the model

7 Conclusions

In this paper, according to the requirements of enterprises, the various sections and wall thickness of the multiple unit train were analyzed through the load method of U-shaped

and staircase distribution and then several feasibility schemes were determined. The most reasonable scheme was selected under the comprehensive consideration of the safety and economy (Table 5). Through the finite element method the development cycle of the enterprises is greatly shorten, the research funding of the enterprises is saved and a better development platform is provided to the enterprises.

Table 5. The best scheme

Scheme	Decreasement of the seat weight(%)	Displacement (mm)	Stress (Mpa)
60×40(2.5×2)	22.5%	0.82	108

Acknowledgements

References

1. Arno, L.: Casting Simulation as a Tool in Concurrent Engineering. In: International ADI and Simulation Conference, May 28-30 (1997)
2. Ferrera, W., Martins, F., Kameoka, S., Salloum, A.S., Raeya, J.T.: Structural Optimization of Automotive Components Applied to Durability Problems. SAE paper 2003-01-3547 (2003)
3. Schiling, A., Bidmon, K., Sommer, O., et al.: Filling Arbitrary Holes in Finite Element Models. In: 17th International Meshing Roundtable Pittburgh, pp. 12–15 (2008)
4. Lotz, K.D.: Finite Element Analysis of the Torsiona l Stiffness of a Convertible Car Body. Cranfield University (1991)
5. Ma, X., Zhao, Y.P.: Finite Element Analysis on the Stiffness of the Body Structure of a Light-duty Bus, pp. 86–88 (2001)
6. Fu, M.W., Yong, M.S., Tong, K.K., Muramatsu, T.: A Methodology for Evaluation of Metal Forming System Design and Performance via CAE Simulation. Int. J. Prod. Res. 44, 1075–1092 (2006)

Implementation of LED Array Color Temperature Controlled Lighting System Using RISC IP Core

Cheol-Hong Moon and Woo-Chun Jang

Gwangju University, Gwangju, Korea
`chmoon@gwangju.ac.kr, zone45@nate.com`

Abstract. In this article, an LED Array Color Temperature Controlled Lighting System has been implemented using an 8 bit RISC IP Core for the lighting control system, as well as a Color Temperature Controlling IP and Delta-Sigma DAC IP designed to control the system. The light sources are made of an LED Array, and the LEDs are configured to have 10 stick bars, such as 3 chips of white (30EA), daylight (30EA), red (30EA), green (30EA), and blue (30EA) 0.1W SMD. The time information is acquired through a Real Time Clock, and bio rhythm compatible presentation is made through the LED Array Color Temperature control. The temperature control IP and Delta-Sigma DAC IP are interfaced by accessing the SFR address of the 8 bit RISC IP Core. The system is configured so that the Delta-Sigma DAC IP would produce 0V~3.3V analogue signals through a low bandwidth passing filter and control the lighting system through the serial communications with a PC using the serial port.

1   Introduction

As the depletion of the energy sources becomes a critical problem, many studies on saving energy have been actively pursued regardless of the field of research. As interest in energy saving products grows, LEDs which consume less power and converts the energy into light more efficiently, has emerged as the light source for the next generation of lighting systems to replace the conventional light bulbs and fluorescent lights.[1]. Light is very important not only as its simple function to light a space and so provide a certain level of light regardless of the use or function of the space, but also in psychological aspects of human beings. Therefore a lighting system should be implemented that can provide high level functionality for the living space for healthy life, as well as to provide pleasant lighting environments.[2].

In this article, the lighting system is designed to use a reusable 8 bit RISC structure Core and Control IP (Intellectual Property) in order to save time and money.[5]. For the lighting in the system, LEDs, the next generation light source (white, daylight, red, green and blue) have been used to save energy and to provide pleasant lighting environments that do not disturb natural human bio-rhythms, by controlling the Color Temperature.

2   The Lighting System

The lights are controlled using an 8 bit RISC IP Core with a Main CPU. To control the LED Array, a Color Temperature IP and Delta Sigma DAC IP, to control the

lightings via analogue means have been designed. Additionally, to control the Color Temperature according to the time information, a Real Time Clock and Dallas Company's DS1302 IC, which is more cost efficient than any other product because of its built-in register, are used here. In this paper, all of the 8 bit RISC IP Cores are designed to process the instruction set of ASM51 and 80C31. The MCU Core is able to interface with various kinds of peripheral circuits without making any changes in the Core circuit.

Fig. 1. Lighting System Block Diagram

3   Lighting Control IP Design

3.1   Color Temperature Control IP

1) Color Temperature
The color of the light that an ideal black body emits is designated by its temperature, according to Planck's Law of Radiation. When an object shines, emitting visible lights, and the color is identical to that which a black body radiates at a certain temperature, it is considered that the temperature of the black body and the object is the same. This temperature is called the Color Temperature of the object.

2) Color Temperature Control IP Design
When LEDs are used as the light source, they have a great advantage over other light sources since they can make digital pulse type operations. To change the color temperature of LEDs, the voltage passed to LED by a PWM (Pulse Width Modulation) signal needs to be changed. Using the built-in SFR address of the 8 bit RISC IP Core, it accesses the Color Temperature Control IP. Each register setting allots and uses PWM0~PWM4 settings through the SFR address Map to obtain the different colors. Fig.2 shows the interface of the RISC IP Core and the Color Temperature Control IP.

Fig. 2. Color Temperature Control IP Interface

A data register is used to store the counter value, and the data value decides the required pulse width so that an Up/Down counter can load a new value from the data register when the counter value reaches TC (Terminal Count). A Toggle Flip Flop issues the PWM signal.
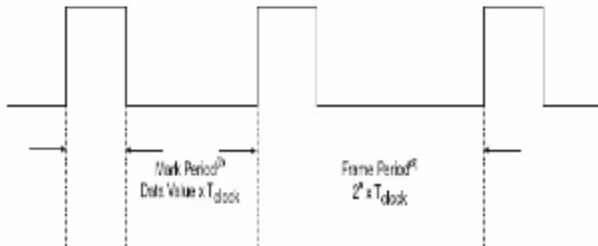
Fig. 3. Sample PWM Output Wave

Fig.3 shows the output signal of the PWM generator and equation. According to the entered data 0x19, 0x40, 0x80, 0xC0, 0xE6, the duty cycle is output as 10%, 25%, 50%, 75% and 90%, respectively.

3.2 Delta-Sigma DAC IP

The DAC (Digital-to-Analog Converter) converts binary signals to the analogue voltages proportional to the digital value. The external circuits needed to make the DAC are; 1 resistance for the low bandwidth passing filter, and 1 capacitor, only. The consumption of the FPGA resources is also low. As the Delta-Sigma DAC uses a digital method, it is independent to the temperature changes and can be implemented with programmable logic. Delta-Sigma is actually a high speed single bit DAC, and it
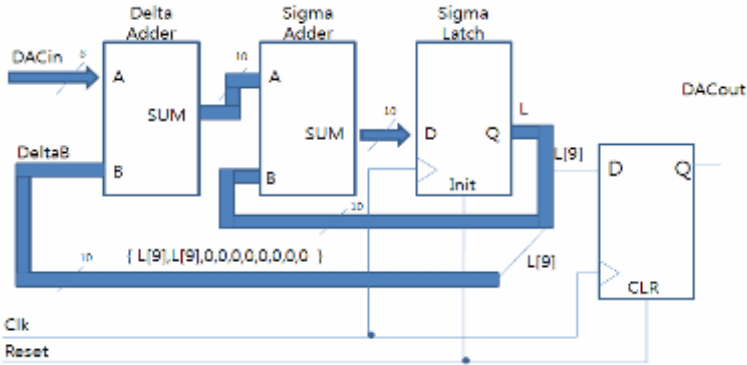
Fig. 4. Block Diagram of Delta-Sigma DAC

creates pulse arrays using digital feedback. The average duty cycle of the pulse arrays is proportional to the binary input values. When pulse arrays pass the low bandwidth passing filter, analogue signals are created.[10].

Here we use the notation Delta to mean the arithmetic difference, and Sigma to mean the arithmetic sum, where the difference and the sum are calculated using a binary adder. The data should be entered in 8 bits. DeltaB is the value stored in the Sigma Latch and is composed of 10 bits, so it is expressed as binary number by copying the uppermost bit L [9] to make two, converting the rest of the 8 bits to 0 and entering them into DeltaB. Each bit shows the value calculated and stored in the Adder and the Latch. Although the Delta Adder entry values do not have plus or minus signs, the output of the two adders is considered as having a sign. The Delta Adder calculates the difference between DACin and the current DAC output. This value, and the previous output stored in the Sigma Latch, is added in the Sigma Adder.

4   Experiments and Results

4.1   Color Temperature Control IP Simulation

The Color Temperature Control IP logic consists of a data register, an UP/DOWN counter, Cnt_out_int_mux, Rco_int and Pwm_int. When Sfraddr signals and the Color Temperature Control IP addresses are identical, the algorithm receives 8 bit data through the data register where the data is stored to perform the Up/Down counter by receiving 8 bit data at Sfrdatai (7: 0) signals. This data decides the duty cycle. In the SFR Map, PWM0_ID is set as 0x7B(0xFB), PWM1_ID as 0x7C(0xFC), PWM2_ID as 0x7D(0xFD), PWM3_ID as 0x7E(0xFE) and PWM4_ID as 0x7F(0xFF). The program shows that the 0x19 value is loaded in PWM_data0, 0x40 in PWM_data1, 0x80 in PWM_data2, 0xC0 in PWM_data3 and 0xE6 in PWM_data4. The output signal time of Pwm0_o is 2160ns, Pwm1_o 5280ns, Pwm2_o 10,400ns, Pwm3_o 15,520ns and Pwm4_o 18,560ns respectively. Accordingly, each duty cycle is confirmed as 10%, 25%, 50%, 75% and 90% respectively, throughout the simulations.
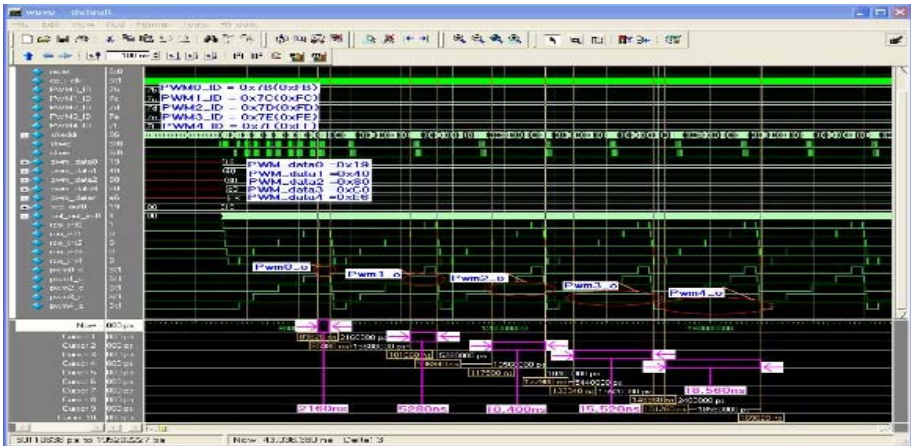
Fig. 5. 5 Channel Color Temperature Control IP Simulations

The PWM waves, emitted when 0x80, 0xC0 and 0xE6 values, are entered in the PWM0 register can be checked through the simulation results.
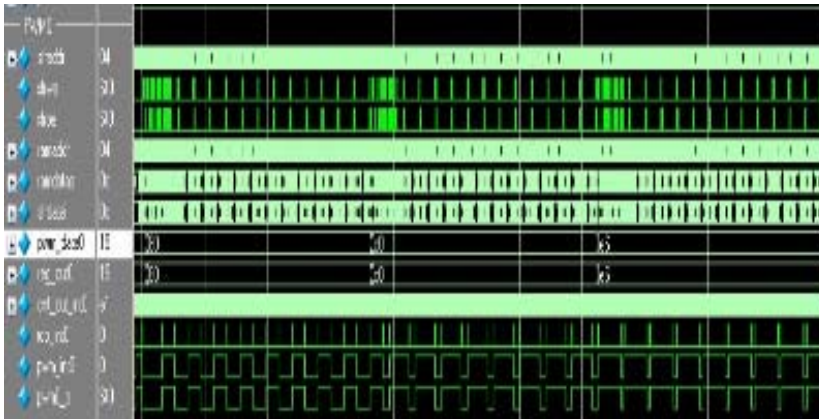
Fig. 6. Wave 2: 50%(0x80); 75%(0xC0); and 90%(0xE6)

As shown in the simulation results, when 8 bit data values are entered in to the register, the desired duty cycle of PWM signal can be acquired. PWM0 controls white light, PWM1 gives daylight, PWM2 gives red, PWM3 gives green and PWM4 outputs blue light.

4.2   Delta-Sigma DAC IP Simulations

The DAC IP logic consists of a Delta Adder block, a Sigma Adder block and a Sigma Latch block. When the Sfraddr signal and the Sfrwe signal are consistent with the

DAC IP addresses, 8 bit data is entered in the DeltaA of the Delta Adder. Then 10 bit output signals from the Sigma Latch block are entered to the DeltaB which is expressed as a binary number by copying the upper most bit L[9]. The rest of the 8 bits are converted to 0 and the 10 bits are entered to DeltaB. The Delta Adder calculates the difference between DACin, the 8 bit input signal and the current Sigma Latch output value. The 10 bit output value, calculated in the Delta Adder, and the previous output value, stored in the Sigma Latch, is added in the Sigma Adder block. In the Sigma Latch block, when the Clk signal is Active, the upper most bit L[9] is output to DACout, and the output value from Sigma Latch gives the feedback to the Sigma Adder.

Fig. 7. DACin Entry Value 0x80

The DACout signals are confirmed according to the input data values using Modelsim. Fig.7 show that when the Sfrwe signal is Active, and the Sfraddr value is 0x78, it is consistent with DAC_ID=0x78. Therefore, 0x80 is entered to the DACin. As the initial value of the output value of 00 from SigmaAdder is 0, the initial load value in the DeltaAdder is 0x080.

4.3   Experimental Results

IPs, as explained in this paper, are modeled using Verilog HDL. As all the units are designed in RTL (Register Transfer Level), they have advantages in composition and optimization. Experimental results show that the desired movement and output signals can be acquired in the simulation of each IP.

Fig.8 shows the actual lighting system implementing the features explained in this paper. It is composed of an Xilinx FPGA device, a peripheral EEPROM, the LED operation driver AMC7140, a power supply part, an LED Array, and Real Time Clock. The lighting control system is programmed to control the color temperature of the LED Array (W, W_W, R G, B) through the serial communication, and to change

Fig. 8. Lighting System

Table 1. Presentation Table for Bio-rhythms

	Description
1 stage	6 AM color temperature 3500k, intensity of illumination 15lx (3000k 85% + 6500k 15% about 3525k)
2 stage	8 AM color temperature 4000k, intensity of illumination 60lx (3000k 70% + 6500k 30% about 4050k)
3 stage	12 PM color temperature 5500k, intensity of illumination 101lx
4 stage	2 PM color temperature 6500k, intensity of illumination 150lx
5 stage	6 PM color temperature 4000k, intensity of illumination 100lx
6 stage	12 AM color temperature 3000k, intensity of illumination 10lx

Table 2. Measurement Data by Stage

	Intensity of illumination(lx)	Color temperature(K)	Electric current(A)	Electric power(W)
1 stage	15	3520	0.09	2.01
2 stage	60	4000	0.13	3.62
3 stage	101	5523	0.14	5.98
4 stage	160	6400	0.16	9.43
5 stage	103	4000	0.15	7.04
6 stage	10	3000	0.08	1.03

the color temperature from 2,000°K to 8,000°K according to the color temperature of natural light. The system is configured to control the Delta-Sigma DAC IP output signal level.

Table1 shows the measurement value of the intensity of illumination and color temperature in each given situation, and amperes and watts accordingly. Fig.9 shows the result of the experiments to change the intensity of illumination and color temperature according to the presentation table of Table2, by creating a particular space (60*60*90) and using Color Temperature Control IP.

Fig. 9. Presentation through the Color Temperature Control

5   Conclusions

The purpose of lighting control is not only for energy saving in social environments and it should also aim to create pleasant environments while not reducing the quality of the lighting. The LEDs required, are those that are more efficient and which consume less energy, to develop inexpensive lighting control chips that can perform the essential control functions. In this paper, to implement the ASIC chip (lighting control IC) for lighting control, an 8 bit RISC IP Core has been used. By designing a Color Temperature Control IP, and implementing it on a FPGA, the Color Temperature Control lighting system has been successfully implemented. For the FGPA chip, a 54Mhz chip is used, and for the main CPU, 27Mhz is used by dividing the frequency into two.

In the 8 bit RISC IP Core, all the buses are allotted to the SFR so that the peripheral circuits can be expanded easily without modifying the Core circuit. To load the Boot Code to the External EEPROM, an I2C IP has been designed and the Color Temperature Control IP in this paper is controlled by accessing the built-in SFR address of the RISC IP Core.

For the color temperature presentation, it should have a spectrum distribution that changes continuously within the wavelength of visible light. Therefore, the LED array has been composed by applying white light, daylight, red light, green light and blue light LEDs, which have different wavelengths, respectively. The Color temperature of the LED Array can be controlled in a similar way to that of natural light, by modulating the duty cycle (%) into 255 steps according to the 8 bit data value that is calculated. The time information is acquired through a Real Time Clock, the time and control values are displayed using LCDs and the lighting system is controlled by serial communication with a PC connected through the serial port. The Delta-Sigma DAC IP is interfaced by accessing the SFR address of the RISC IP Core. With 8 bit data values (0x00~0xFF), the system outputs 0V~3.3V analogue signals through the low bandwidth passing filter using the DAC, and it is able to control the lighting system with the analogue signals. More studies and verification are needed to produce an

effective lighting control ASIC chip in the future. Additionally, an actual lighting system should be installed to verify and analyze the issues of residential environments, and modified based on the customer satisfaction survey results.

Acknowledgements

References

1. Jang, J.H.: A study on the LED Implement which is Proper for Space and User Performs Color Temperature and Color (2005)
2. Choi, A.S.: Development and Application of Health Lighting Plan in Residential Areas. Journal of Architectural Institute of Korea (2004)
3. Sung, H.K.: Disign of 8bit RISC processor. Journal of Research Institute of Energy and Resources (2003)
4. Song, S.W.: 16bit RISC processor Design (2008)
5. Lee, K.Y.: Disign of Microcontroller Core Chip and Fabrication for ASIC. Journal of The Industrial Technology Institute (1996)
6. Lee, Y.S.: Disign of High Performance 8bit Microcontroller Using Pipeline (2005)
7. Kim, H.S.: Disign of Low Power 8bit RISC Processor Using Clock gating (2006)
8. Philips: The I2C Bus Specification Ver 2.1 (2000)
9. Kong, W.T.: Control of LED Sensor By Surrounding Temperature (2007)
10. Kang, K.S.: Class-D Digital Audio Amplifier Using 1-bit 4th-order Delta-Sigma Modulation. Journal of The Institute of Electronics Engineers of Korea (2008)
11. Park, S.U.: The Modeling of Pipeline A/D Converter Using Verilog-A. Journal of Korea Information and Communications Society (2007)
12. Mentor, G.: ModelSim SE 6.0 Specification Manual (2004)
13. Tektronix: TLA5203 logic Analyzer Manual (2004)

A Universal Data Access Server for Distributed Data Acquisition and Monitoring Systems

Dae-Seung Yoo, Vu Van Tan, and Myeong-Jae Yi⋆

School of Computer Engineering and Information Technology
University of Ulsan, San-29, Moogu-2 dong, Namgu, Ulsan 680-749, Korea
ooseboss@ulsan.ac.kr, {vvtan,ymj}@mail.ulsan.ac.kr

Abstract. This paper introduces a universal data access (DA) server for modern distributed data acquisition and monitoring systems used in process and factory automation. This system is proposed with utilization of the OPC (Openness, Productivity, and Collaboration) technology and XML to achieving interoperability and platform independence. It allows to easily aggregate a large number of existing OPC DA servers and new OPC XML-DA servers into a unified and flexible system that supports exchange of data among these servers. By using binary data encoding to the SOAP messages, the proposed system has a sufficient good performance. The security consideration is discussed to provide more information to technical-level readers. The comparison of the proposed system with the existing approaches is also presented.

Keywords: Automation, data acquisition, OPC technology, Universal DA Server, XML.

1 Introduction

Web technology is playing an important role in process and factory automation today, especially for control systems of relatively slow processes [9]. Besides, XML is the preferred format to encoding and moving the structured data in the independent system to achieve the interoperability and platform independence. As a result the OPC XML-DA standard was defined to move the same type of plant floor data as existing OPC DA [23]. Such OPC standards are enhanced the ability of exchanging complex data by using the OPC complex data [22].

Due to the developments of process and factory automation, i.e., process monitoring and control systems, it requires to aggregate hundreds or even thousands of existing OPC DA servers and new OPC XML-DA servers into a unified and flexible system that allows exchange of complex data not only between the client and server, but also between these OPC servers. Recently, a commercially available system, `Universal XML-DA Server`, developed by the Advosol [18] has the same functionality mentioned above, but the performance of such system is not good because of using textual XML data representation.

To transfer data in the heterogeneous environments, the bandwidth problem needs to be considered and solved adequately. The fundamental approach to

---

⋆ Corresponding author.

tackle such problem today is using a binary data representation with the agreement about a binary exchange format. The consideration for the memory buffer, CPU resources, and conversion between the XML representation and the binary representation to hold the whole data for this approach is therefore required. Moreover, only the fast possibility to transmit a large amount of data is to use the binary representation rather than the XML representation. But different platforms or even different compilers of the same platform use different binary representations such as different floating point formats, different string formats, and so forth [24,7]. Some universal standards should be chosen to transport data between different platforms. The OPC XML-DA standard has a big disadvantage that used the XML data representation, i.e., the size of a SOAP (Simple Object Access Protocol) message exchanged between the client and server is very large. This causes much more network traffic to transfer data [7].

The study of this article aims at proposing and developing a unified and flexible system, called `Universal DA Server`, for distributed data acquisition and monitoring systems. The proposed system allows a large number of existing OPC DA servers and new OPC XML-DA servers to be aggregated into the Universal DA Server that also supports the horizontal data exchange between these OPC servers. This system is designed and developed including interfaces, architecture, and modules. They make us easy to identify implemented components to reuse for the developments of modern distributed data acquisition system softwares.

2   Background and Problem Statements

2.1   Relevant OPC Specifications

The OPC Foundation is an independent, non-profit organization that comprises leading manufacturers and solution providers in factory and process automation as well as providers of enterprise solutions. This section reviews several relevant OPC specifications. Firstly, the OPC DA specification defines a set of standard COM (Component Object Model) objects, methods, and properties that specifically address interoperability requirements for the factory automation, process control, and condition monitoring applications, and so forth [23]. The OPC DA technology leverages DCOM allowing the client-server applications to access the plant floor via an Ethernet network distributed across the manufacturing enterprise. The OPC DA applications are only compatible with Microsoft Platforms.

Secondly, the OPC XML-DA defines a new specification to move the same type of plant floor data as the existing OPC DA products [24]. This provides vertical integration between the plant floor and condition, monitoring, maintenance, etc., using XML, HTTP, SOAP, and industrial standards. The OPC XML-DA provides better connectivity and interoperability for production management and enterprise applications such as Manufacturing Execution System (MES), Enterprise Resource Planning (ERP), Enterprise Asset Management (EAM), and plant optimization that need to access the plant floor data. It was designed to allow the existing OPC DA products to be wrapped by the OPC XML-DA Interfaces and to support both interfaces from the same OPC server.

Thirdly, the OPC Complex Data will provide a full way for the OPC clients to read and decode any type of the data from field devices on the plant floor [22]. Complex data mean that an OPC Item is defined as a structure. The item includes read-only information, run-time status and writeable control points. The complex data consist of complex data items that can include non-structured items, structured items, XML data, OPC binary, and so forth [22].

Finally, the OPC DX (Data eXchange) has well-defined objects that are based on the OPC COM-DA and XML-DA objects [21]. The OPC DX addresses the simple mechanism for moving data between source and destination. It sets out the rules associated with the when, what, and how of moving the data between endpoints. Moreover for dealing with exception conditions, it defines the value to be written to the target or to be maintained at the target when good data are not available from the source. The OPC DX adds some key extensions by leveraging the OPC DA and OPC XML-DA standards to exchange the data horizontally between peer level OPC applications. The OPC DX solution also extents data access to enable server-to-server data exchange during run-time.

2.2   Problem Statements

More recently, hundreds or even thousands of existing OPC DA servers and new OPC XML-DA servers need to be aggregated into a system, which allows complex data to be exchanged not only between the OPC servers, but also between the OPC servers and OPC clients. A commercially available system, i.e., `Universal XML-DA Server`, was developed by the Advosol [18]. This system allows that the OPC DA servers and OPC XML-DA servers can be configured as data source for client accessing. Nevertheless, the performance of this system is weak because of using the XML textual data representation.

Several systems based on the OPC XML-DA specification have been proposed and developed successfully [3,7,11,15,10]. However, the OPC XML-DA solution has a big disadvantage because of using the XML data representation [24,7]. The XML data representation causes much more network traffic to transfer data. The OPC XML-DA requires XML messages to be very descriptive about the data being transferred. For example, instead of "0.555" the record <value xsi:type = "xsd:float">0.555< /value> is sent, meaning that the bandwidth is increased about six times or even more [7]. Furthermore, data alignments are often required to transport data in native representation. The OPC Complex Data based on the use of the XML data presentation requires big amount of memory and high intensity of memory management operations. In addition, more CPU resources and memory buffer for transformation between the native data representation and XML are required [4,3]. In terms of software engineering, the ability to reuse and upgrade the components of a developed system is an important factor. It will make the cost of an application reducible as well as possible. Moreover, the compatibility and interoperability of a new system with existing systems and middleware are strictly required in designing and implementing a system.

To solve bandwidth problem, the fundamental approach to transfer data in the heterogeneous environments today is using a binary data representation,

which is integrated into XML to reduce the size of a message. This approach makes a challenge to consider the memory buffer, CPU resources, and conversion between the XML representation and the binary representation to hold the whole data [3]. Different platforms or even different compilers of the same platform use different binary representations such as different floating formats, different string formats, etc. The interoperability of a system makes us difficult to guarantee.

To tackle the problems as aforementioned, a distributed data acquisition and monitoring system is proposed including the following features:

1. Allowing to easily aggregate existing OPC DA servers and new OPC XML-DA servers into a *unified and flexible* system. This system allows data to be exchanged not only between these OPC servers, but also between the OPC servers and clients.
2. The proposed system can be configured online from local or remote clients for the configuration changes. The OPC clients can also access both the OPC DA servers and OPC XML-DA servers in a full way.
3. By using the binary data representation, the performance of the proposed system is improved much and reaches an acceptable performance.
4. The proposed system is designed and developed including interfaces, architectures, functionalities, and modules. They make us easy to reuse, upgrade, and maintain in order to expand new features.
5. The interoperability of the proposed system is guaranteed based on the investigation of `XML libraries` and by using the OPC specifications.

3   Designing and Developing Universal DA Server

3.1   Architectural Overview

A gateway application for OPC DA servers, OPC XML-DA servers, and other gateways is required that can provide a sound base for new applications and an easier migration path for the hundreds (or thousands) of the OPC DA products today. Therefore, Universal DA Server based on the OPC DX specification is proposed and developed. This system allows clients to access both types of the OPC DA server and OPC XML-DA server. It can be configured the access rights such as *write-only*, *read-only*, and *read-write* from online or local configuration clients. The architecture of the Universal DA server is shown in Fig. 1.

A number of the OPC DA servers and OPC XML-DA servers can be configured as data sources for clients to access. The address space of the Universal DA Server maps the address spaces of both types of such servers into a *unified and flexible* address space. The Universal DA Server allows the OPC clients to read and decode any type of data from field devices (see [15,17] for the detailed technique and implementation).

The Universal DA Server is responsible for managing the data-flow on a connection from source to target. In general, it subscribes to data from the sources and copies these data to the target when it is received. The flow-diagram of updating the data from source servers to target items is shown in Fig. 2. On the
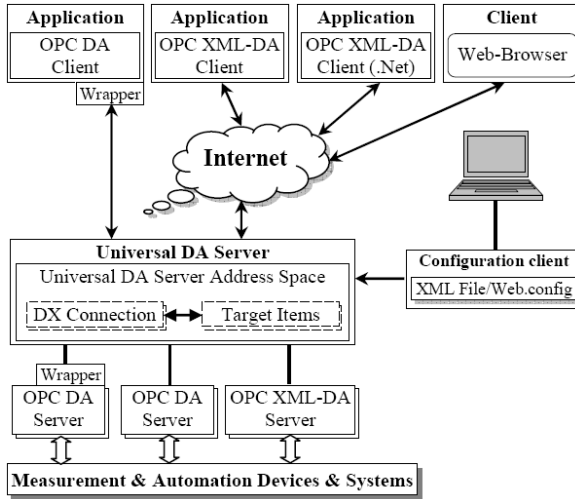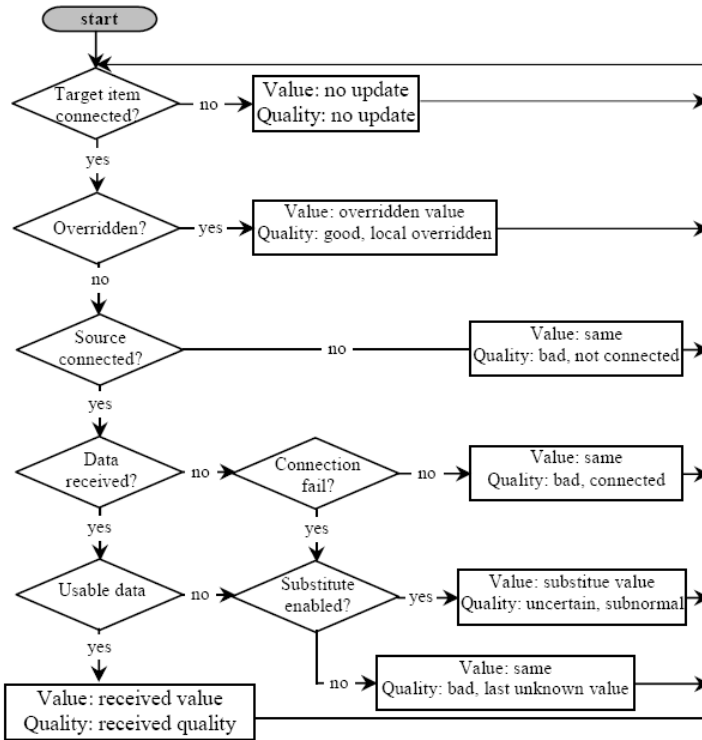
Fig. 1. The architecture of Universal DA Server

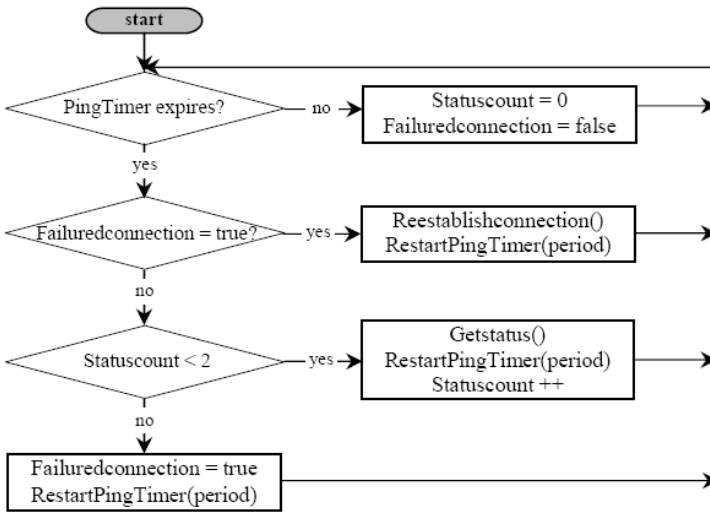**Fig. 2.** The flow diagram of updating data to target items

Fig. 3. The flow diagram of recovering a connection failure

other hand the Universal DA Server is responsible for maintaining the connection to source servers. When the connection to a source server is lost, either during startup or runtime, it tries to reestablish the connection in the minimum time between two attempts to reestablish a connection by using a period value, i.e., the period value can be set by the administrators or operators, e.g., `10 seconds`. The flow diagram of recovering a connection failure between the Universal DA Server and source servers is shown in Fig. 3.

3.2 Modular Design of the Universal DA Server Implementation

The Universal DA Server that can access to a number of the OPC DA servers and new OPC XML-DA servers might be configured by configuration clients. The OPC items of all configured OPC servers are mapped into single address that can be structured in any suitable way. Both operations `Read` and `Write` are supported. The components of the Universal DA Server for aggregating multiple OPC servers as shown in Fig. 4. By this way the OPC DA servers and OPC XML-DA servers are considered as data sources. The data achieved from the source servers depend on the types of the source servers, e.g., OPC DA or OPC XML-DA. In general, the mechanism is referred to a `Subscription`. Subscribing to the data is supported by `Callback` service in COM-DA and by `Subscribe` service in XML-DA. In the case of COM-DA, the Universal DA Server creates one or more groups for each source server. With defined groups, the Universal DA Server adds data items to them to indicate the source server which items to access. In the case of XML-DA, the Universal DA Server uses `Subscribe` and `SubscriptionPolledRefresh` services to acquire source data. The `Subscribe` service allows the clients to define a single operation for a set of items of a source server to access.
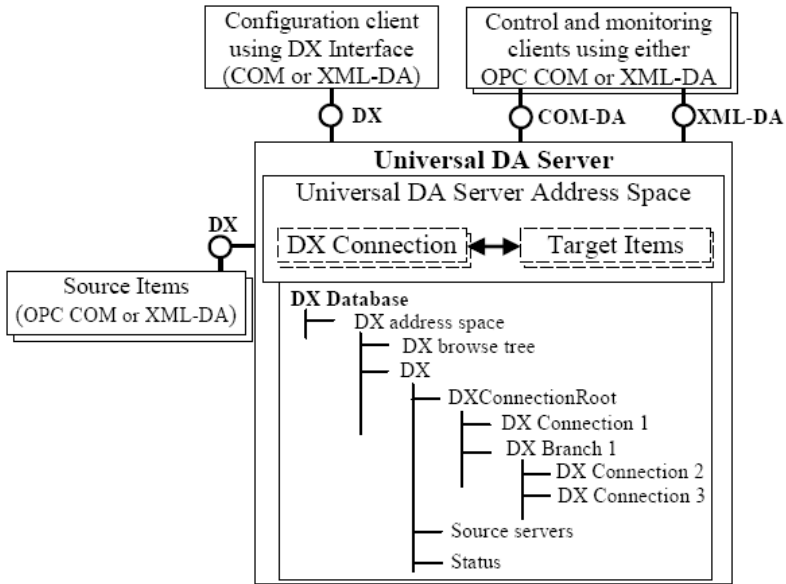
Fig. 4. Components of the Universal DA Server

The Universal DA Server is composed of three components: (i) The COM-DA component exposes OPC DA-Interface and DX-Configuration Interface implemented as COM-Interface. (ii) The XML-DA component handles SOAP requests and exposes both OPC XML-DA Interface and DX-Configuration Interface implemented as Web Service Interface using Web Service Description Language (WSDL). (iii) The DX component contains parts of functionality of the Universal DA Server and marshals data to/from the OPC servers. It can create new connections, remove connections, and modify the status of each connection. The inconsistency and synchronization issues are required solving because DX-Connections can be modified and controlled by the SOAP and COM clients simultaneously.

3.3 Connecting and Representing Data

This section will concern about issues for connecting and representing data exchanged between the Universal DA Server and clients.

Data Representation: The fundamental approach to solve the bandwidth problem in the heterogeneous environments is using the binary data representation, which is integrated into XML such as BXML [2], BXSA [5], etc. Recently, several proposals are available to satisfy these conditions like SOAP message with attachment [1] or HTTP message using XLink (XML Linking Language) [6]. The WS-attachment technology [1,8] could be used to incorporate binary data into the SOAP messages. The SOAP message with attachment of binary data encoding will reduce the required size of a message about six times or even

Fig. 5. The comparison between query languages. SQL is the fastest access method while XSU is feasible for querying in platform independent databases.

more: (i) the SOAP header is still XML and (ii) the body of the SOAP message should be encoded as binary data.

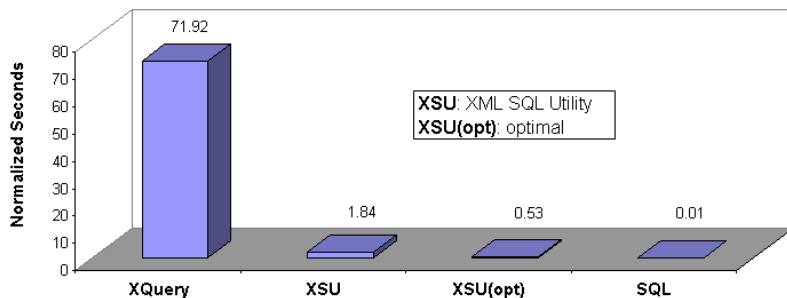**XML Libraries Comparison:** There are many available XML library benchmarking projects such as XMark [25], XML Benchmark [26], and so forth (see [13] for more recent XML benchmarking projects). Unfortunately, they tested only one or several aspects of XML processing with some predefined sequences of XML data. The performance tests can be divided into the following phases of XML processing stages: `Shema Validation`, `XML Transformation`, `XML Security`, `XML Parsing`, etc. `LiXML` is effectively high performance [4] and it is therefore selected to apply for the development of the Universal DA Server.

Connection between the Universal DA Server and Database: To provide an open and flexible system for implementing and aggregating the OPC HDA [20] into the proposed system for the future development, the connection between the Universal DA Server and database is investigated and provided. Standard SQL (Structure Query Language) as a basic query language for database searches in many cases is not efficient. Consequently, each database has its own extension to SQL such as PL/SQL for Oracle, PL/pgSQL for postage-SQL. However, SQL solutions are usually not platform independent. The platform independence can be achieved by using the XQuery language [27]. But in large applications the XQuery language is still too slow and too memory consuming [7] (see Fig. 5 –time was measured when searching 1000 records and fours records are transferred for one timestamp).

When the Universal DA Server is defined as a web server and supports storage and exchange of the complex data, it should be designed to expand the connection with database. The ability to extend and integrate the proposed system when adding new OPC products or combining the existing OPC-based products should be guaranteed satisfactorily.

The situation of XQuery language will change rapidly in near future due to efforts put into the further development of XML. The solution to tackle this development of XML should be started with the product depending query language XSU and to migrate to XQuery [27,7].

3.4   Security Consideration

Security to remote invocations in the heterogenous environments is an important issue including the authentication of clients, encryption of messages, and the access control. The security objectives to guarantee these issues will include *integrity*, *confidentiality*, *availability*, *authentication*, *non-repudiation*, *accountability*, and *reliability* [14]. The XML signature surely meets the requirements of the security objectives like *integrity*, *non-repudiation*, and *authentication*. It also implicitly meets the security objectives *accountability* and *confidentiality* due to the *authentication*. Therefore, digital signatures are suitable for securing a modern distributed data acquisition and monitoring system that used in process and factory automation in the heterogeneous environments.

Several security solutions used for the application domains of OPC-based monitoring and control systems have been proposed and presented by Tan et al. [16]. The optimal approach complying with the XML distributed signature approach is applied to the Universal DA Server. This approach was modified from the XML distributed signature approach proposed by Miyauchi [12] to reduce the data transfer between the clients and the Universal DA Server (see [16] for more details about descriptions of the solutions). Complying with the consideration of security requirements to the Universal DA Server, the modified security solution can meet the security objectives as aforementioned.

4   Comparison with Existing Approaches

The comparison of the proposed system with other approaches is difficult due to the conceptual nature, different architectures, and wide range of production environments. A qualitative comparison of the proposed system and existing approaches could be made. The structural characteristics are used to compare with existing systems: (i) `Universal XML-DA Server` – a commercially available system – [18], (ii) `KATRIN Slow Control` [7], and (iii) another system developed by Chilingargyan and Eppler [3]. The proposed system has a number of advantages as follows:

1. The proposed system allows to easily aggregate existing OPC DA servers and new OPC XML-DA servers into a *unified and flexible* system which permits exchange of the complex data not only between the OPC servers and clients, but also between these OPC servers. In practice, a commercially available system [18] has the same functionality, but the performance of this system is weak due to using the XML textual data representation.
2. With the utilization of both the binary data encoding in SOAP message and optimal XML library, the required size of a SOAP message is reduced about six times or even more compared to the size of a pure SOAP message. Thus network bandwidth and system resources are improved much.
3. The proposed system can be configured online from a local or remote client that makes it flexible, optimal, and robust. A large number of the OPC DA servers and XML-DA servers can be configured as data sources for client to

access with the configured access rights such as *write-only*, *read-only*, and *read-write* while this feature cannot be executed by `KATRIN Slow Control` [7] and Chilingargyan and Eppler's system [3].

4. The proposed system is designed and developed in a systematic way including architecture, interfaces, functionalities, and modules. By this way, in terms of *framework component* and *software engineering*, these components can be reused to apply to the development of specific applications where the control is distributed and large amounts of distributed data are collected. These components are easy to identify and they will be reused for the future development of modern automation softwares.

5. The proposed system is open for extension by adding new OPC-based products. Therefore, the further development of OPC systems according to the OPC HDA specification [20] and the OPC AE specification [19] can be incorporated into the proposed system. The connection between the Universal DA Server and database was investigated for the development of the OPC HDA standard [20].

5 Concluding Remarks

The work presented in this paper has three main contributions: Firstly, the `Universal UA Server` was proposed that allows multiple OPC DA servers and new OPC XML-DA servers to be integrated into a *unified and flexible* system, which supports the complex data exchange not only between the OPC servers, but also between the OPC servers and clients. The proposed system allows integration of a large number of OPC servers. It can be configured online using a remote or local client that makes it flexible and optimal. Secondly, the `XML Libraries` were investigated and compared in order to choose XML technology as a candidate for the proposed system. The `LibXML` was used to solve requirements of monitoring and control systems today, e.g., platform independence, acceptable performance, etc. Thus the proposed system has a sufficient good performance and it will be accepted to apply to several real industrial applications in the future. The third contribution is the security solution that was adequately discussed to use a good security solution. This solution provides a seamless security approach to the proposed system that is communicated using the Internet environments. The security objectives are also provided to technical-level readers. Furthermore, some open issues for the Universal DA Server are provided for developing modern distributed data acquisition and monitoring systems with cooperating other OPC specifications-based products.

In terms of software engineering and framework component, the proposed system was designed and developed to reach an ultimate solution in designing and developing process and factory automation softwares. Thus, this solution will be reused for the further development of process and factory automation applications where control is distributed and large amounts of distributed data are collected.

Acknowledgements

References

1. Barton, J.J., Thattle, S., Nielsen, H.F.: SOAP Message with Attachments. W3C Recommendation (2006), http://www.w3.org/TR/SOAP-attachments
2. Bruce, C.S.: Cubewerx Position Paper for Binary XML Encoding (2002), http://www.cubewerx.com/main/HTML/Binary_XML_Encoding.html
3. Chilingaryan, S., Eppler, W.: High Speed Data Exchange Protocol for Modern Distributed Data Acquisition Systems based on OPC XML-DA. In: Proceedings of the 14th IEEE-NPSS Real-time Conference, pp. 352–356 (2005)
4. Chilingaria, S.: Fast, Multiplatform XML Toolkits Comparison: Functionality, Speed and Memory Usage (2003), http://prdownloads.sourceforge.net/xmlbench/features.pdf, http://prdownloads.sourceforge.net/xmlbench/benchmark.pdf
5. Chiu, K., Devadithya, T., Lu, W., Slominski, A.: A Binary XML for Scientific. In: Proceedings of the 1st International Conference on e-Science and Grid Computing, pp. 336–343 (2005)
6. DeRose, S., Maler, E., Orchard, D.: XLink (XML Linking Language), Version 1.0. W3C Recommendation (2001), http://www.w3.org/TR/xlink/
7. Eppler, W., Beglaria, A., Chilingarian, S., Kelly, S., Hartmann, V., Gemmeke, H.: New Control System Aspects for Physical Experiments. IEEE Transactions on Nuclear Science 51(3), 482–488 (2004)
8. Freisleben, B., Friese, T., Heinzl, S., Mathes, M., Smith, M.: Flex-SwA: Flexible Exchange of Binary Data based on SOAP Messages with Attachments. In: Proceedings of the IEEE International Conference on Web Services, pp. 3–10 (2006)
9. Holley, D.W.: Understanding and Using OPC for Maintenance and Reliability Applications. IEE Computing and Control Engineering, 28–31 (2004)
10. Jia, Z., Li, X.: OPC-based architecture of embedded web server. In: Wu, Z., Chen, C., Guo, M., Bu, J. (eds.) ICESS 2004. LNCS, vol. 3605, pp. 362–367. Springer, Heidelberg (2005)
11. Katsuji, U., Shin-ichi, S., Hidehiko, W.: A Prototype Embedded XML-DA Server and its Evaluations. In: Proceedings of the SICE-ICASE International Joint Conference, pp. 4331–4336 (2006)
12. Miauchi, X.M.L.: Signature/Encryption – The Basic of Web Service Security. NEC Journal of Advanced Technology 2(1), 35–39 (2005)
13. Mlýnková, I.: XML Benchmarking: Limitations and Opportunities. Technical Report, Department of Software Engineering, Charles University, Czech Republic (2008), http://www.ksi.mff.cuni.cz/~mlynkova/doc/tr2008-1.pdf
14. Singh, M.P., Huhns, M.N.: Service-Oriented Computing - Semantics, Processes, Agents. John Wiley & Sons, Ltd., Chichester (2005)
15. Tan, V.V., Yoo, D.S., Yi, M.J.: Design and Implementation of Web Service by Using OPC XML-DA and OPC Complex Data for Automation and Control Systems. In: Proceedings of the 6th IEEE International Conference on Computer and Information Technology, p. 63 (2006)

16. Tan, V.V., Yoo, D.S., Yi, M.J.: Security in Automation and Control Systems based on OPC Techniques. In: Proceedings of the 2nd International Forum on Stategic Technology, pp. 136–140 (2007)
17. Tan, V.V., Yoo, D.S., Yi, M.J.: Modern Distributed Data Acquisition and Control Systems based on OPC Techniques. In: Proceedings of the 14th Annual IEEE International Conference and Workshops on the Engineering of Computer-Based Systems, pp. 115–122 (2007)
18. The Advosol, Inc.: Universal XML-DA Server, Version 1.01 (2008), http://www.advosol.com/c-5-xml-da-gateways.aspx
19. The OPC Foundation: The OPC Alarms and Events Access Specification, Version 1.0 (2002), http://www.opcfoundation.org/Downloads.aspx
20. The OPC Foundation: The OPC Historical Data Access Specification, Version 1.0 (2003), http://www.opcfoundation.org/Downloads.aspx
21. The OPC Foundation: The OPC Data eXchange Specification, Version 1.0 (2003), http://www.opcfoundation.org/Downloads.aspx
22. The OPC Foundation: The OPC Complex Data Specification, Version 1.0 (2003), http://www.opcfoundation.org/Downloads.aspx
23. The OPC Foundation: The OPC Data Access Specification, Version 3.0 (2003), http://www.opcfoundation.org/Downloads.aspx
24. The OPC Foundation: The OPC XML Data Access Specification, Version 1.01 (2004), http://www.opcfoundation.org/Downloads.aspx
25. XMark: an XML Benchmark Project (2002), http://monetdb.cwi.nl/xml/index.html
26. XMLBenchmark: XML Benchmark (2002), http://www.sosnoski.com/opensrc/xmlbench/-download.html
27. XQuery: XQuery 1.0 – An XML Query Language (2007), http://www.w3.org/TR/xquery/

A Petri Net-Based Ladder Logic Diagram Design Method for the Logic and Sequence Control of Photo Mask Transport

Yun Liu

College of Engineering Technology, Jimei University, Xiamen City, Fujian Province, 361021, P.R. China

Abstract. In this paper, a Petri net-based logic and sequence control model of photo mask transport was constructed. A characteristic analysis was made on the established Petri net model in behavior and structure. This checked the safeness and reliability of the constructed control model off line. In order to implement the Petri net-based control model of photo mask transport via a Programmable Logic controller (PLC), an approach was proposed to generate Ladder Logic Diagram (LLD) programs from the Petri net-based control model. Execution of the generated LLD programs validated the effectiveness of the proposed method.

Keywords: Petri net, Ladder logic diagram (LLD), Photo mask transport, Programmable Logic Controller (PLC).

1 Introduction

For an industrial course control with logic and sequence control taking a large portion, the commonly used control model design methods are Ladder Logic Diagram (LLD) and Petri net. LLD is a conventional and graphical control program design approach to planning the logic and sequences of controlled objects' action. However, LLD will be very complex when there is a large number of control objects and great complexity in the control courses. Moreover, due to LLD's poor readability, and inconvenience in making an off-line evaluation of LLD in behavior and structure [1], locating the failure positions in LLD is difficult. Flexibility, reusability and good maintainability are the fundamental requirements for software design [2]. From this point of view, LLD is not an attractive control program design method. However, Petri net, as an effective logic and sequence control design tool, has been extensively used to model, analyze, simulate and control industrial systems [3] [4] [8]. In addition, the programs generated by Petri net-based control models can be more conveniently extended to incorporate new functions, compared to the programs based on LLD [10]. As the hardware of the control system, for our photo mask transport, is based on a PLC, the program can only be written either in LLD, or in function block diagram (FBD), or in standard template library (STL)[5]. In order to evade the disadvantages of LLD programming, and exert the advantages of Petri net in modeling, simulation and

reparation, we build a logic and sequence control model based on Petri net, and make a characteristic analysis on the control model in behavior and structure. This is an effective way to make an off-line evaluation of safeness and reliability of the control program, which is produced subsequently by the Petri net-based control model [6] [7]. With the assignment of special control properties to the transitions and places of the established Petri net-based control model, we convert the Petri net-based control model into LLD programs, which control the course of photo mask transport directly.

The remainder of the paper is organized as follows. Section 2 gives an introduction about Petri net and the art of photo mask transport. In Section 3, we build Petri net-based models for the overall control of photo mask transport and the control of photo mask loading, respectively, and make an analysis about the established Petri net models in structure and behavior. Section 4 proposes an approach to converting Petri net-based control models into LLD programs. The paper is concluded with Section 5.

2 Preliminaries

2.1 Petri Net

A marked Petri net $Z = (P, T, I, O, m_0)$ [13], where P is a set of places, graphically represented by circles; T is a set of transitions, graphically represented by bars, with $P \bigcup T \neq \emptyset$ and $P \bigcap T = \emptyset$; $I : P \times T \to N$ is the input function and $I(p_i, t_j)$ specifies the connecting weight on the arc directed from p_i to t_j; $O : T \times P$ is the output function and $O(t_i, p_j)$ specifies the connecting weight on the arc directed from t_i to p_j; $m : P \to N$ is a marking whose ith component is the number of tokens graphically represented by dots in the ith place. m_0 is an initial marking where $N = \{0, 1, 2, \ldots\}$.

The behavior of a Petri net is determined by exercising enabling and firing rules: 1) A transition $t \in T$ is enabled iff $m(p) \geq I(p, t), \forall p \in P$; 2) An enabled transition t may fire at marking m', yielding the new marking

$$m(p_i) = m'(p_i) + O(p_i, t) - I(p_i, t), \forall i = 1, 2, \ldots, |P|. \tag{1}$$

The marking m is said to be reachable from m'.

By (1), the $n \times l \, (n = |P|, l = |T|)$ Incidence Matrix matrix composed of integers for a Petri net model is defined as

$$C := O - I, \tag{2}$$

where $c_{ij} = O(p_i, t_j) - I(p_i, t_j), i = 1, 2, \ldots, |P|, j = 1, 2, \ldots, |T|$. The state of a Petri net model starts from its initial marking m_0, experiences a transition serial S, and reaches its new state with a marking $m \in R(m_0)$, a reachable state set from m_0. Thus we can obtain an equation of state transition

$$m = m_0 + C \cdot S, \tag{3}$$

where $S : T \to N$ is an enabling sequence, $S = [s_1, s_2, \ldots, s_n]^T$ with s_i the times that t_i has been fired in the Petri net model's transition from m_0 to m.

A Petri net model has two kinds of properties (behavior properties and structure properties) [11]. Behavior properties are closely related to its initial state. In application, behavior properties involve reachability, boundedness, conservativeness and activeness. However, structure properties are irrelevant to its initial state, and are only determined by its topology.

In order to check whether a Petri net-based logic and sequence control model can work normally, it's very necessary to analyze the model's properties of conservatism, consistence, repeatability and boundedness. A Petri net model Z is conservative iff $\exists$ a weighting vector $w > 0$ such that

$$w(P)^T m(P) = \sum_{i=1}^{n} w(p_i) m(p_i) = K, \forall m \in R(Z, m_0), \qquad (4)$$

where m_0 is the system's initial marking, K a given constant, and $R(Z, m_0)$ a reachable marking set from marking m_0. The system is consistent if each transition in the finite transition set is fired as least once in a recycling state evolution from m_0 to m_0. If each transition (or a portion of transitions) is (are) fired infinite times, the system will be called a repetitive (or partially repetitive) system. If a system is K-bounded, the number of the tokens in every place is less than K when a Petri net model is in any state $m \in R(m_0)$. A Petri net model is structurally bounded if the system is bounded for any initial sate m_0. As the places in a Petri net model usually denote the resources and buffers in the Petri net modeled system, the number of tokens in each place must be bounded for a normal system. A Petri net model is active iff $\forall t_j \in T, j = 1, \ldots, |T|$, and $m \in R(Z, m_0)$, there exists only one transition serial $[t_1, t_2, \ldots, t_l]^T$, $l \leq |T|$, which can be enabled. A Petri net model is reversible iff $\forall m \in R(Z, m_0)$, $m_0 \in R(Z, m)$.

2.2   The Art of Photo Mask Transport

The art of photo mask transport consists of two main parts: photo mask loading and unloading. In the course of photo mask loading, the manipulator accesses a photo mask from a mask library, transports the photo mask to a coarse pre-alignment position, performs a coarse pre-alignment, transports the photo mask to a fine pre-alignment position on the completion of the coarse pre-alignment, implements a fine pre-alignment, and hands the pre-aligned photo mask over to an explosion table. In the course of photo mask unloading, the manipulator receives a photo mask from the explosion table, and transports it back to a specified grid of an ordered photo mask library.

3   Petri Net-Based Logic and Sequence Control Models

3.1   Methods for Design and Analysis

Two commonly used methods to construct a Petri net model are Top-Down and Bottom-up methods. From the photo mask transport art described in Section 2.2, we know that photo mask transport consists of two main parts of mask loading

and unloading. Furthermore, the main course of mask loading consists of a lot of operations such as accessing mask from mask library, bar code reading, a coarse pre-alignment, mask transporting between working places, a fine pre-alignment, fault diagnosis and processing, and so on. Considering the tree like structure of the art, we choose the Top-Down method to construct a Petri net-based logic and sequence model for photo mask transport. First, we divide the entire course of photo mask transport into several main blocks according to their functions, and establish an overall Petri net model for the main course. Then, we build a Petri net model for each main block to fulfill its function. With these two steps, we obtain a Petri net-based logic and sequence control model for photo mask transport.

According to the art of photo mask transport, we construct a Petri net-based logic and sequence control model. However we cannot guarantee the established Petri net will work as expected. Therefore, a characteristic analysis on the Petri net model is necessary. There are a lot of methods for the property analysis on the Petri net model. In [9], a net extension method is used to check whether a Petri net model is reachable. Reachable graph [12] is employed to analyze a Petri net model of a Feasible Manufacture System (FMS). In [13], Matrix theory and Linear Matrix Inequality are used to analyze the Incidence Matrix of a Petri net model and the criterions are proposed for a Petri net model to be conservative, consistent, bounded and repeatable. In [14], several methods are given to reduce a Petri net model's structure while maintaining its original reachable property, and thus simplify the course of characteristic analysis. As there are only a small number of transitions and places in the established Petri net model, we prefer a reachable tree to analyze it.

3.2   An Overall Control Model Based on Petri Net

The Overall Control Model Based on Petri Net. The Top-Down method is used to construct a Petri net-based logic and sequence model for mask transport. First, the whole course is divided into three parts: mask loading, mask unloading, and mask delivering to an explosion table. Second, fulfill the function for every part. Finally, we build an overall Petri net-based logic and sequence control model, shown in Fig.1(a) with an illustration of the corresponding places and transitions listed Table 1, for photo mask transport by combining the three parts together according to the art flow.

Structure Analysis. According to the Petri net model shown in Fig.1(a), we obtain an input matrix I, an output matrix O, and consequently an incidence matrix C, which are expressed by

$$I = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 \end{bmatrix}, \quad O = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 1 \end{bmatrix}, \quad C = O - I = \begin{bmatrix} 1 & -1 & 0 & 0 \\ 0 & 1 & -1 & 0 \\ 0 & 0 & 1 & -1 \\ -1 & 1 & -1 & 1 \\ -1 & 0 & 0 & 1 \end{bmatrix}, \quad (5)$$
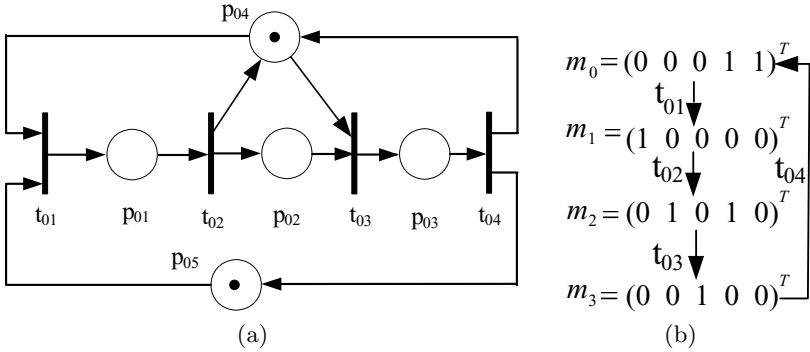
respectively.

$$m_0 = (0\ 0\ 0\ 1\ 1)^T$$
$$m_1 = (1\ 0\ 0\ 0\ 0)^T$$
$$m_2 = (0\ 1\ 0\ 1\ 0)^T$$
$$m_3 = (0\ 0\ 1\ 0\ 0)^T$$

(a)   (b)

Fig. 1. (a) The overall Petri net-based logic and sequence control model of photo mask transport, and (b) The reachable graph for the Petri net-based logic and sequence control model of photo mask transport.

Table 1. An illustration of places and transitions in Fig. 1(a)

Place	Description	Transition	Description
p_{01}	photo mask loading	t_{01}	start to load mask
p_{02}	explosion table receiving mask	t_{02}	completion of mask loading
p_{03}	photo mask unloading	t_{03}	start to unload mask
p_{04}	libraries and manipulator in free	t_{04}	completion of mask unloading
p_{05}	explosion table in free		

In order to solve for a $P-$invariable, assume $X = [x_1, x_2, x_3, x_4, x_5]^T$, and let

$$C^T X = 0, \tag{6}$$

for which, we find a P-variable $X = (2\,1\,2\,1\,1)^T$. In order to solve for a T-invariable, suppose $Y = (y_1, y_2, y_3, y_4)^T$, and let

$$CY = 0, \tag{7}$$

for which we find a $T-$invariable $Y = (1\,1\,1\,1)^T$. As there exist both $P-$ and $T-$invariables with positive and integral components for the incidence matrix, the Petri net is bounded and consistent in its structure. From the discussion above, we observe that the overall Petri net-based control model for mask loading and unloading is bounded and consistent in both behavior and topology. Hence, the overall Petri net-based control model can be said to be safe and reliable in its structure.

The Peri Net-Based Control Model. The art flow of photo mask loading is shown in Fig. 2(a). Firstly, the manipulator accesses a ordered photo mask from a mask library. Secondly, the bar code reader scans the code on the photo mask accessed by the manipulator. Thirdly, the system judges whether the accessed photo mask is the ordered one by comparing the scanned bar code with

Table 2. An illustration of places and transitions in Fig. 2(c)

| Place	Description	Transition	Description
p_{11}	manipulator moving to access mask	t_{11}	start mask loading
p_{12}	opening the gate of mask library	t_{12}	start opening the gate
p_{13}	library moves to specified grid	t_{13}	preparation is completed
p_{14}	access mask from library	t_{14}	start to bar code checking
p_{15}	bar code checking	t_{15}	bar code is correct
p_{16}	failure processing	t_{16}	bar code is incorrect
p_{17}	transport and align	t_{17}	failure process is completed
p_{18}	manipulator and library in free	t_{18}	mask loading is completed
p_{19}	explosion table in free		
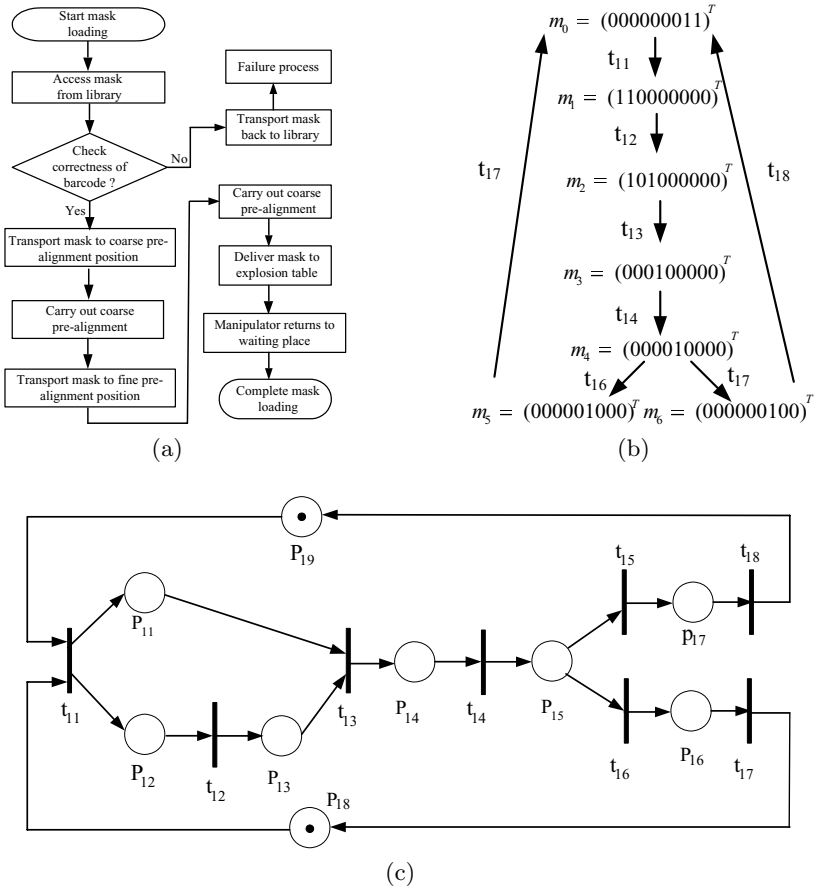
Fig. 2. (a) The art flow of photo mask loading, (b) The reachable graph of the Petri net-based logic and sequence control model of photo mask loading, and (c) The Petri net-based logic and sequence control model of photo mask loading

the desired bar code on the ordered photo mask. If the two bar codes do not match, the manipulator will send the photo mask back to its original place in the mask library, and the system will begin to process this failure; otherwise, the manipulator will transport the photo mask to a coarse pre-alignment position. Fourthly, the manipulator will carry out a coarse pre-alignment to deal with the orientation and position error in a large range for the photo mask. Fifthly, on the completion of the coarse pre-alignment, the manipulator will transport the photo mask to a fine pre-alignment working place and execute a fine pre-alignment to reduce the photo mask's position and orientation error in a small range. Finally, the manipulator will deliver the photo mask to a photo mask explosion table when the fine pre-alignment is completed, and move to awaiting position. This is the entire course for photo mask loading. By the art, a Petri net-based sequence and logic control model for photo mask loading is constructed as shown in Fig. 2(c) with an illustration of the transitions and places listed in Table 2.

Behavior and Structure Analysis. Using the depth first method, we obtain a reachable graph shown in Fig. 2(b), from which, we observe that $\forall p \in P$, $m(p) \leq 1$, and then conclude that the Peri net-based logic and sequence control model is safe and bounded in its behavior. From the Petri net model, we get its incidence matrix

$$
C = \begin{bmatrix}
1 & 0 & -1 & 0 & 0 & 1 & 0 & 0 \\
1 & -1 & 0 & 0 & 0 & 1 & 0 & 0 \\
0 & 1 & -1 & 0 & 0 & 1 & 0 & 0 \\
0 & 0 & 1 & -1 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 1 & -1 & -1 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 1 & -1 & 0 \\
0 & 0 & 0 & 0 & 1 & 0 & 0 & -1 \\
-1 & 0 & 0 & 0 & 0 & 0 & 1 & 1 \\
-1 & 0 & 0 & 0 & 0 & 0 & 1 & 1
\end{bmatrix} .
$$

Assume $Y = [y_1, y_2, y_3, y_4, y_5, y_6, y_7, y_8]^T$, and let

$$CY = 0, \tag{8}$$

for which, there exist a positive and integral solution $Y = (2\,2\,2\,2\,2\,1\,1\,1\,1)^T$, a T-invariable for this Petri net model. Similarly, suppose $X = [x_1, x_2, x_3, x_4, x_5, x_6, x_7, x_8, x_9]^T$, and let $C^T X = 0$, we can also obtain a positive and integral solution $X = (1\,1\,1\,2\,2\,2\,2\,1\,1)^T$, a P-invariable for this Petri net model. Since there exist both T- and P-invariables (positive and integral) for the incidence matrix, the Petri net-based logic and sequence control model for the photo mask loading is bounded, conservative and repeatable. In summary, the established Petri net-based logic and sequence control model for Photo mask loading is active, safe and reliable in both topology and behavior.

As the photo mask loading and unloading have the similar operation courses, the Petri net modeling and analyzing methods stated above also apply to the photo mask unloading course. For the limit of space, the details of design and analysis are omitted here. From the art of mask transport, we know that there is

no increase in resource or resource shared with other parts when the Petri net-based control models of the mask loading and unloading are inserted into the overall Petri net-based logic and sequence control model. Hence, by the theorems stated in [13], the final Petri net-based control model containing detailed mask loading and unloading operations will remain active, repeatable and safe.

4   The Petri Net-Based LLD Programming Method

A lot of specialized hardware supports the Petri net controller, and can directly use the Petri net model to control the process. However, the equipments in our experiment do not support the Petri net controller, we have to discuss the conversion from a Petri net model to LLDs. As discussed in Section 2, the state evolution of a Petri net model is governed by the rules of transition enabling and token changing. Now we will find the logic expressions composed of input, output, operation and constraints imposed by the rules, and convert the logic expressions into LLDs.

The enabling conditions for LLDs are different from the enabling rules for a Petri net. In LLD design, the input events and control conditions must be taken into account. Let E_i and G_i be the input events and control conditions related to transition t_i, respectively. Then the enabling condition is

$$t_i(\tau) = \prod_{p_j \in {}^* T_i} (p_j(\tau) \cdot G_i(\tau) \cdot E_i(\tau)), \tag{9}$$

where if $p_j(\tau)$ stands for an action or a process is in progress, then

$$p_j(\tau) = \begin{cases} 1 & \text{if } p_j(\tau) \text{ action or process is in progress at the instant } \tau, \\ 0 & \text{otherwise;} \end{cases} \tag{10}$$

if $p_j(\tau)$ stands for resource and $m(p_i(\tau))$ denotes the number of resources, then

$$p_j(\tau) = \begin{cases} 1 & \text{if } m(p_j(\tau)) \geq \text{ the weight on the arc dricted from } p_j \text{ to } t_i, \\ 0 & \text{otherwise;} \end{cases} \tag{11}$$

$G_i(\tau)$ represents the sate of the control conditions at $t = \tau$, thus

$$G_i(\tau) = \begin{cases} 1 & \text{if all the input conditions at } \tau \text{ are true,} \\ 0 & \text{otherwise.} \end{cases} \tag{12}$$

If $p_j \in^* T_i$ and p_j is an operating place, then

$$E_i = \prod_{p_j \in {}^* T_i} Z(p_j), Z(p_j) = \begin{cases} 1 & \text{if } p_j \text{ is completed and } p_i = 1, \\ 0 & \text{otherwise.} \end{cases} \tag{13}$$

If $t_j(\tau) = 1$, t_j will be enabled immediately. At $t = \tau + \Delta\tau$, the sate of p_i can be expressed as

$$p_i(\tau + \Delta\tau) = (p_i(\tau) + \sum_{t_j \in {}^* p_i} \alpha_{ji} t_j(\tau)) \cdot \prod_{t_j \in {}^* p_i} \overline{t_j(\tau)}, \tag{14}$$

where α_{ji} is the connecting weight between $t_j \in {}^* p_i$ and p_i.

According to the above rules converting a Petri net model to LLDs, we obtain the following logic relations for the Petri net model of photo mask loading: $t_{01} = p_{04} \cdot p_{05}, t_{12} = p_{12} \cdot Z(p_{12}), t_{13} = p_{13} \cdot p_{11} \cdot Z(p_{13}) \cdot Z(p_{11}), t_{30} = p_{30} \cdot Z(p_{30}), t_{31} = p_{31} \cdot Z(p_{31}), t_{32} = p_{32} \cdot Z(p_{32}), t_{14} = p_{33} \cdot Z(p_{33}), t_{14} = p_{33} \cdot Z(p_{33}), t_{15} = p_{15} \cdot Z^+(p_{15}), t_{16} = p_{15} \cdot Z^-(p_{15}), t_{50} = p_{50} \cdot Z(p_{50}), t_{51} = p_{51} \cdot Z(p_{51}), t_{52} = p_{52} \cdot Z(p_{52}), t_{53} = p_{53} \cdot Z(p_{53}), t_{54} = p_{54} \cdot Z(p_{54}), t_{55} = p_{56} \cdot Z(p_{56}), t_{56} = p_{57} \cdot Z(p_{57}) \cdot p_{55} \cdot Z(p_{55}), t_{40} = p_{40} \cdot Z(p_{40}), t_{41} = p_{41} \cdot Z(p_{41}), t_{42} = p_{42} \cdot Z(p_{42}), t_{43} = p_{43} \cdot Z(p_{43}), t_{44} = p_{44} \cdot Z(p_{44}), t_{45} = p_{45} \cdot Z(p_{45}), t_{46} = p_{44} \cdot Z(p_{44}), p_{46} \cdot Z(p_{46})$, where $Z^+(p_{15})$ stands for correct bar code checking, and $Z^-(p_{15})$ for incorrect bar code checking; $p_{11} = t_{01} \cdot \overline{t}_{13}, p_{12} = t_{01} \cdot \overline{t}_{12}, p_{13} = t_{12} \cdot \overline{t}_{13}, p_{30} = t_{13} \cdot \overline{t}_{30}, p_{31} = t_{30} \cdot \overline{t}_{31}, p_{32} = t_{31} \cdot \overline{t}_{32}, p_{33} = t_{33} \cdot \overline{t}_{14}, p_{15} = t_{14} \cdot \overline{t}_{15} \cdot \overline{t}_{16}, p_{40} = t_{16} \cdot \overline{t}_{40}, p_{41} = t_{40} \cdot \overline{t}_{41}, p_{42} = t_{41} \cdot \overline{t}_{42}, p_{43} = t_{42} \cdot \overline{t}_{43}, p_{44} = t_{43} \cdot \overline{t}_{46}, p_{45} = t_{43} \cdot \overline{t}_{45}, p_{46} = t_{45} \cdot \overline{t}_{46}, p_{04} = p_{05} = t_{46} \cdot \overline{t}_{01}; p_{50} = t_{15} \cdot \overline{t}_{50}, p_{51} = t_{50} \cdot \overline{t}_{51}, p_{52} = t_{51} \cdot \overline{t}_{52}, p_{53} = t_{52} \cdot \overline{t}_{53}, p_{54} = t_{53} \cdot \overline{t}_{54}, p_{55} = t_{54} \cdot \overline{t}_{56}, p_{57} = t_{55} \cdot \overline{t}_{56}, p_{04} = t_{56}, p_{02} = t_{54}$.

As the PLC scanning sequence is from top to bottom, and from left to right, LLD programs are ought to be written in a inverse sequence of the actions to avoid the avalanche in LLD [15]. The LLD for photo mask loading can be easily generated according to above logic expressions. The details are omitted here. Similarly, the LLD programs for photo mask can be obtained from its Petri-net model. The safe and reliable running of the LLD programs on our PLC controlling system verified the correctness of the Petri net-based logic and sequence control model. In addition, the convenience in debugging and function extension manifested that the Petri net-based logic and sequence control model had good maintainability.

5   Conclusion

In this paper, we established a Petri net-based logic and sequence control model for photo mask transport, made a characteristic analysis of the established control model in structure and behavior, and proposed an approach to converting the established control model into LLD programs. As the Petri net-based logic and sequence control model and the LLD programs, thereafter generated by the Petri net-based control, are equivalent in structure and behavior, the characteristic analysis on the Petri net-based control model is an effective way to predict off line the characteristics of the LLD programs. This may provide a new idea for generating safe and reliable LLD programs for sequence and logic control. Experiment verified the effectiveness of the proposed method.

References

1. Venkatesh, K., Zhou, M., Reggie, J.: Comparing Ladder Logic Diagrams and Petri Nets for Sequence Controller Design through a Discrete Manufacturing System. IEEE Transactions on Indutrial Electronics 41, 611–619 (1994)
2. Richardsson, J., Fabin, M.: Automatic Generation of PLC Programs for Control of Flexible Manufacturing Cells. In: IEEE Conference on emerging technologies and automation, ETFA 2003, vol. 2, pp. 337–344 (2003)
3. Han, Z., Xiong, D.: Hybrid Petri nets for modeling of manufacturing systems. Tsinghua Science and Technology 42, 1429–1432 (2002)
4. Tang, Y., Zhou, M., Qiu, R.G.: Virtual Production Lines Design for Back-end Semiconductor Manufacturing Systems. IEEE Transactions on Semiconductor manufacturing 16, 543–550 (2000)
5. OMRON SYSMAC CS1-serial Programming Manual, OMRON(China) (1999)
6. Simon, C.: Developing Software Controllers with Petri Nets and Logic of Actions. In: Proceeedings of the 2001 IEEE International Conference on Robotics and Automation, Seoul Kerea, May 2001, pp. 58–63 (2002)
7. Liao, D., Jeng, M., Zhou, M.: Petri Net Modeling and Lagranian Relxation Approach to Vehicle Scheduling in 300 mm Semiconductor Manufacturing. In: Proceeedings of the 2004 IEEE International Conference on Robotics and Automation, Neworleans, LA, April 2004, pp. 5301–5306 (2004)
8. Charr, J.K., Charr, A.V., Edward, S.D.: An Integrated Approach to Developing Manufacturing Control Software. In: Proceeedings of the 1991 IEEE International Conference on Robotics and Automation, Sacramento Califoria, April 1991, pp. 1979–1984 (1991)
9. Ramachandran, P.: An Acyclic Transformation Technique for the Reachability Analysis of Petri nets: [Disseration of PHD], Graduate college of Oklahoma State University (2002)
10. Iordache, M.V.: Method for the Supervisory Control of Concurred Systems Based on Petri Net Abstractions: [Dissertation of PHD], Graduate school of University of Notre Dame (2003)
11. Park, J.: Structural Analysis Control of Resource Allocation Systems Using Petri Nets: [Dissertation of PHD]. Grduate school of Geogia Institute of Technology (2000)
12. Bouyekhf, R., Moudni, A.E.: On the Analysis of Some Strutcutre Properties of Petri Nets. IEEE Transactions on Systems, Man and Cebernetics-Part A: Systems and Humans 35, 784–794 (2005)
13. Zhou, M., DiCesare, F.: Petri Net Synthesis for Discrete Event Control of Manufacture Systems. Kluwer Academic Publishers, Norwell (1993)
14. Han, Z., Lee, G.B.: Reduction Method for Reachability Analysis of Petri Nets. Tsinghua Science and Technology 43, 231–235 (2003)
15. Fabian, M., Hellgren: PLC-based implementation of supervisory control of discrete event systems. In: Proc. of the 37th IEEE Conference on Decision and Control, Tampa, Florida, USA, December 3, pp. 3305–3310 (1998)

Modeling of Micro-Piezoelectric Motion Platform for Compensation and Neuro-PID Controller Design

Van-tsai Liu, Ming-jen Chen, and Wei-chih Yang

Department of Electrical Engineering, National Formosa University,
Huwei, Yunlin 632, Taiwan
`vtliu@nfu.edu.tw, n665118@moon.nfu.edu.tw`

Abstract. The purpose of this study is to design a tracking controller for micro-piezoelectric motion platform applications. The hysteresis effect is originated from the piezoelectric actuated platform that provides nonlinear behaviors. A Prandtl-Ishlinskii model is constructed to describe the hysteresis behavior of piezoelectric actuators. The weights of hysteresis model are identified by using the LMS(Least-Mean-Square) algorithm. Based on the Prandtl-Ishlinskii model, a feed-forward controller is developed for compensating the hysteresis nonlinearity. A self-tuning neuro-PID controller is introduced to suppress the tracking errors due to the modeling inaccuracy and hence to get precision tracking errors. These approaches are numerically and experimentally verified which demonstrate the performance and applicability of the proposed designs under a variety of operating conditions.

Keywords: Hysteresis, Prandtl-Ishlinskii model, Piezoelectric actuator, Self-tuning neuro-PID controller.

1 Introduction

In recent years, nanoscale science and engineering is an emerging technological field and have immense economic and societal impacts, with the development in precision engineering and micro-positioning applications, there is an increasing demand in high frequency high precision actuators. Nanotechnology will fundamentally change the manufacturing processes of many materials and devices.

Piezoelectric ceramics is one of the preferred solutions because they not only have rapid response, and high repeatability, they also do not produce magnetic fields and thus do not interfere with a number of equipment, but also in industrial production such as scanning tunneling microscopy and diamond turning machines. However the piezoelectric actuators have the disadvantage of hysteresis behavior, which severely limits system performance such as giving rise to undesirable inaccuracy or oscillations, even leading to instability. Without the aid of further control technique to overcome this problem, it is only possible to achieve a limited positioning accuracy.

The hysteresis phenomenon occurs in all the smart material-based actuators, such as piezoceramics and shape memory alloys. When a nonlinear plant is preceded by the hysteresis nonlinearity, the system usually exhibits undesirable inaccuracies or

oscillations and even instability due to the nondifferentiable and nonmemoryless character of the hysteresis. The development of control techniques to mitigate the effects of hystereses has been studied for decades and has recently reattracted significant attention.

The compensation methods for piezoelectric hysteresis have been widely investigated**,** such as Approximated Polynomial Model[1], Bouc-Wen Model[2], Duhem Model [3], Generic Differential Model[4], Kim Model[5], Maxwell Slip Model[6], Preisach Model[7], Prandtl-Ishlinskii Model. Recently Prandtl-Ishlinskii operator is used to control piezoelectric actuator. The present paper describes an alternative solution for the compensation of the hysteresis of piezoelectric actuator. A Prandtl-Ishlinskii model is introduced. The hysteresis referenced curves are derived by identification of LMS(Least-Mean-Square) adaptive algorithm[8].

Current research in hysteresis modeling and compensation can be broadly classified into closed-loop displacement control and linear control with the feedforward inverse hysteresis model. Other proposed closed-loop schemes to treat hysteresis include linearizing the hysteretic nonlinearity[9], using adaptive control with an approximate model of the hysteresis[10], training a neural network to learn the nonlinearity[11], or a combination of neural network with adaptive control[12].

In this paper we focus on the Prandtl-Ishlinskii model[13,14,15] to describe the hysteresis behavior of piezoelectric actuators. The weights of hysteresis model are identified by using the LMS(Least-Mean-Square) algorithm. Based on the Prandtl-Ishlinskii model, a feed-forward controller is developed for compensating the hysteresis nonlinearity. A self-tuning neuro-PID controller is introduced to suppress the tracking error due to the modeling inaccuracy and hence to extract precision tracking error[16, 17].

2 Hysteresis Model

A piezoelectric actuator is an electrically controllable positioning element, which functions on the basis of the piezoelectric effect. A major limitation of piezoelectric actuators, however, is rate-independent hysteresis exhibited between input and output. The output signal of a system with hysteresis depends not only on the present value of the input signal but also on the order of their amplitudes, especially their extrema, but not on their rate in the past. Because of its phenomenological character, the concept of hysteresis operator developed by Kranosel'skii and Pokrovskii allows a very precise modeling of hysteresis system behavior. The basic idea consists in the modeling of the real hysteretic transfer characteristic by a weighted superposition of many elementary hysteresis operators which are the so-called Prandtl-lshlinskii operators.

This model is based on a same principle as a Preisach model, where the idea was to sum the partial nonlinearities. This conception is also called main block. The nonlinearity type backlash is used as an basic block in Prandtl-Ishlinskii model. In this case is implemented the weight function as well. The output of this model is given by summing of partial nonlinearities multiplied be defined weight function. This function has been mentioned above. They changed the partial nonlinearity type backlash with saturation instead backlash type nonlinearity.

2.1 Prandtl-Ishlinskii Operator

The elementary operator in the PI hysteresis model is a backlash operator. A backlash operator is defined by[13]

$$z(t) = B_r[u, y_0](t)$$
$$= \max\{u(t) - r, \min\{u(t) + r, y(t - T)\}\} \tag{1}$$

where u is the control input, z is the actuator response, r is the control input threshold value of the backlash operator, and T is the sampling period. The initial condition of (1) is denoted as[13]

$$z(0) = \max\{u(0) - r, \min\{u(0) + r, y_0\}\} \tag{2}$$

where y_0 is a real number which is usually initialized to 0. Multiplying the backlash operator B by a weight function value w_h can defined as[13]

$$z(t) = w_h B_r[u, y_0](t) \tag{3}$$

where weight function w_h defines the gain of the backlash operator.

A PI hysteresis operator is modeled by a weighted superposition of several backlash operators with different threshold and weighting values as[13]

$$z(t) = \vec{w}_h^T \vec{B}_r[u, \vec{y}_0](t) \tag{4}$$

2.2 Model Parameter Identification

To develop a suitable control law and adaptation scheme for updating the parameters in the hysteresis model, it is important to apply a simple and quick algorithm. The steepest descent algorithm is a convenient way to solve for the optimal tap weight vector, but it needs to know the exact gradient at each time step. A more practical algorithm is the least mean square (LMS) algorithm, which uses the instantaneous estimate of the gradient at each time step, given as

$$Error[x, z](\vec{w}_h, t) = y(t) - z(t)$$
$$= y(t) - \vec{w}_h^T \vec{H}_r[x, \vec{y}_0](t) \tag{5}$$

where $y(t)$ is the plant output and $z(t)$ is the Prandtl-Ishlinskii hysteresis model output. Equation (5) is the weight update equation for the LMS algorithm. The simplicity of this algorithm is apparent, since at each time step the weight update is dependent only on the input vector x multiplied by the current error function, $Error(\vec{w}_h, t)$. If the difference $Error(\vec{w}_h, t)$ is sufficiently small, the weights will be acquired and the Prandtl-Ishlinskii hysteresis model can then be applied.

The Prandtl-Ishlinskii hysteresis model uses seven backlash operators. These parameters are selected by an iterative process, whereby the order of the operators is systematically increased until the modeling performance improvement becomes insignificant.

The identification of the Prandtl-Ishlinskii model weights is based on the measured response of the piezoelectric actuators to a $3 \, \mu m$ p-p sinusoidal control input signal. The Prandtl-Ishlinskii model weights will be obtained by identification of LMS(Least-Mean-Square) adaptive algorithm. Table 1 show, the backlash operator weights of control systems.

Table 1. The optimal values of the PI operator weights

n	1	2	3	4	5	6	7
r_n	0	0.15	0.3	0.45	0.6	0.75	0.9
w_{hn}	1.3	1.3	1.4	1.6	1.7	1.8	1.9

2.3 Feed-Forward Compensator

The main idea of an inverse feedforward controller is to cascade the inverse PI operator Γ^{-1} with the actual hysteresis, which is represented by the hysteresis operator Γ, the operation of the inverse feed-forward is depicted in Fig. 1.
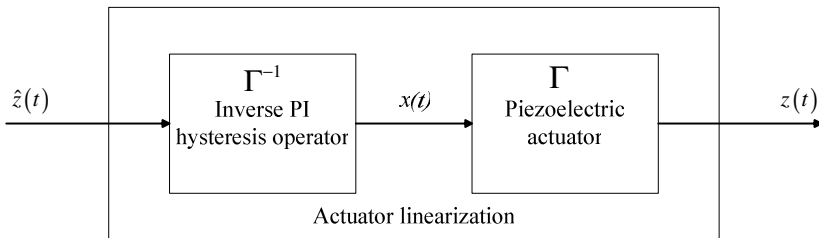

Fig. 1. Inverse feed-forward controller

The inverse PI operator is also of the PI operator. The inverse PI operator is given by

$$\Gamma^{-1}\left[\hat{z}\right](t) = \vec{w}_h''^T \vec{B}_{r'}\left[\hat{z}, \vec{y}_0'\right](t). \tag{6}$$

The inverse PI operator parameters can be calculated by

$$w_{h0}' = \frac{1}{w_{h0}}, \; w_{hk}' = \frac{-w_{hk}}{\left(\sum_{j=0}^{k} w_{hj}\right)\left(\sum_{j=0}^{k-1} w_{hj}\right)}, k = 1 \cdots n, \tag{7}$$

$$r_k' = \sum_{j=0}^{k} w_{hj}\left(r_k - r_j\right), \; k = 1 \cdots n, \tag{8}$$

$$y'_{0k} = \sum_{j=0}^{k} w_{hj} y_{0k} + \sum_{j=k+1}^{n} w_{hj} y_{0j} , \quad k = 1 \cdots n . \tag{9}$$

3   Self-tuning Neuro-PID Controller

The control structure of the self-tuning neuro-PID controller can be shown in Fig. 2 where the outputs of the neural network are proportional, integral, and derivative gains and the inputs are selected in a suitable way according to the specific problem.
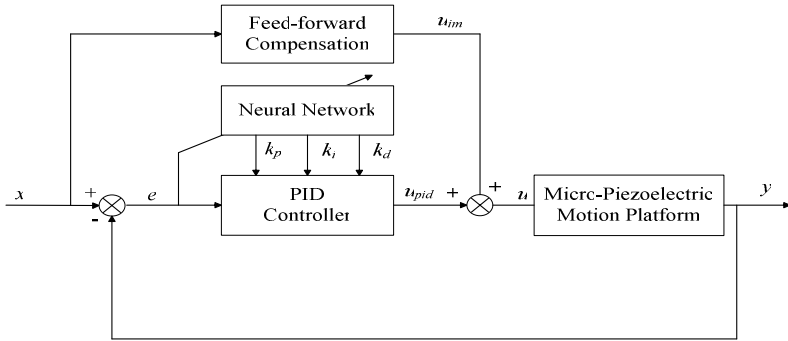

Fig. 2. Self-tunning neuro-PID control scheme

In order to derive the self-tuning neuro-PID control algorithm, can define mathematical notations of neuron models. Total input to neuron j is denoted by net_j and given by

$$net_j = \sum_{i=1}^{n} w_{ji} O_i + b_j , \tag{10}$$

where b_j denotes a threshold. Then the output of neuron j becomes

$$O_j = f\left(net_j\right), \quad f\left(x\right) = \frac{1}{1+e^{-x}} . \tag{11}$$

Using the BP algorithm, can derive the self-tuning neuro-PID control algorithm, note that the PID controller in discrete-time is described as

$$u(k) = u(k-1) + K_P\left(e(k) - e(k-1)\right) + K_I e(k) + K_D\left(e(k) - 2e(k-1) + e(k-2)\right), \tag{12}$$

where $u(k)$ denotes the plant input at kT, and $e(k)$ is the error between a desired value $r(k)$ and the actual output $y(k)$, given as

$$e(k) = r(k) - y(k) . \tag{13}$$

Here, T denotes a sampling time.

In order to adjust K_P, K_I, and K_D adaptively, we use a three layered neural network based on the BP method. Each layer consists of N_1, N_2, and N_3 neurons where N_1 and N_2 can be selected by trial and error according to complexity of the plant and N_3 is three which are equal to the number of PID gains. For the BP method, need a cost function E which should be minimized. Here, the following E is denoted as[17]

$$E = \frac{1}{2} e(k)^2 ,$$ (14)

The connection weights w_{lj} at the output layer are updated by the following relation[17].

$$\Delta w_{lj}(k) = \eta \delta_l O_j + \alpha \Delta w_{lj}(k-1),$$ (15)

where

$$\delta_l = -\frac{\partial E}{\partial net_j},$$ (16)

Using the chain rule for derivative, δ is given as[17]

$$\delta_l = e(k) \cdot \frac{\partial y(k)}{\partial \Delta u(k)} \cdot \frac{\partial \Delta u(k)}{\partial O_l} \cdot \frac{\partial O_l}{\partial net_l} .$$ (17)

Furthermore, from Eq. (12) can obtain as[17]

$$\frac{\partial \Delta u(k)}{\partial O_l} = \begin{cases} e(k) - e(k-1) & l = 1 \\ e(k) & , \ l = 2 \\ e(k) - 2e(k-1) + e(k-2) & l = 3 \end{cases}$$ (18)

where $O_1 = K_P$, $O_2 = K_I$, and $O_3 = K_D$.

In Neural Network setup, forward-Backpropagation Neural Network has three layers : input layer、 hidden layer、 output layer, the nodes number are 4,5,3; the learning rate is 0.5; random number is 0.5; the initial value of network weight value is the random number between +0.5 and -0.5.

4　Experimental Results

The experimental micro-piezoelectric motion platform is shown in Figure 3. Experimental and simulation results for the hysteresis characteristics of the piezoelectric actuator are displayed in Figures 4. A sinusoidal signal is used as a reference command with amplitude 30 µm and frequency 1 Hz. Figure 5 show,

respectively, the position responses of control system with and without feedback control. When applying a feed-forward compensator, the root-mean-square(rms) tracking error was 0.968 µm, the rms tracking error was 3.23 % and the maximum error was 1.5587 µm. Incorporating the feed-forward compensator with the self-tuning neuro-PID controller reduced the rms tracking error was 0.1765 µm and the maximum
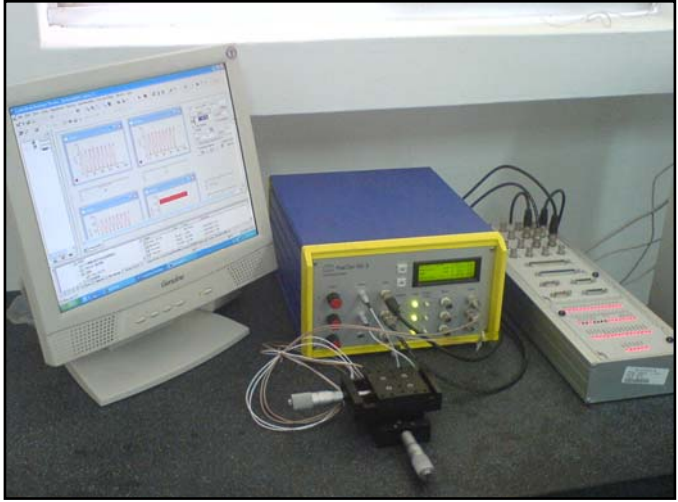

Fig. 3. The experimental micro-piezoelectric motion platform
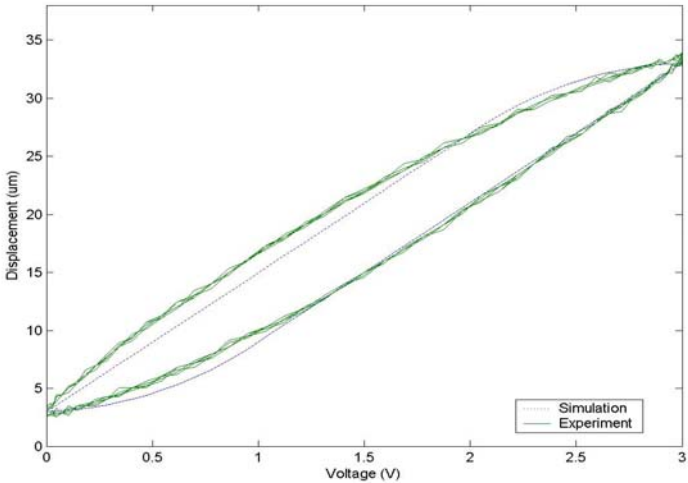
Fig. 4. Experiment and simulation of the hysteresis loop
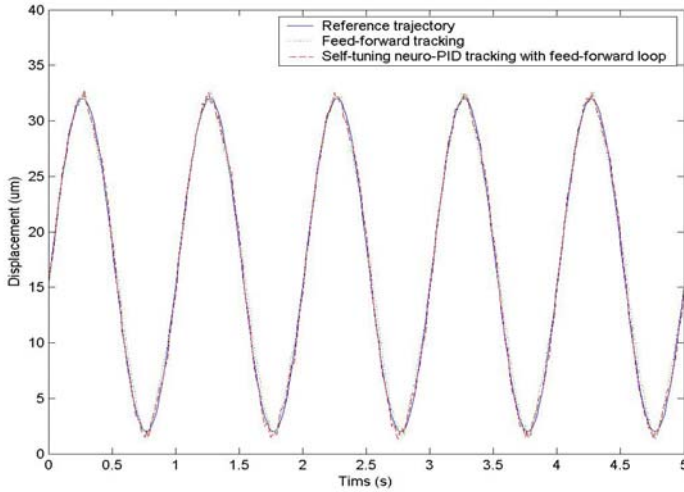
Fig. 5. Reference command and actual position response with feed-forward and the self-tuning neuro-PID controller

Fig. 6. Reference command and actual tracking error with feed-forward and the self-tuning neuro-PID controller

error was 0.613 µm. Figure 6 show, respectively, the tracking error. Figure 7 show the proportional, integral, and derivative gains. Obviously, the self-tuning neuro-PID controller improves the rms error and tracking error and generates satisfactory positioning accuracy.

Fig. 7. Outputs of the neural network are proportional, integral, and derivative gains

5   Conclusions

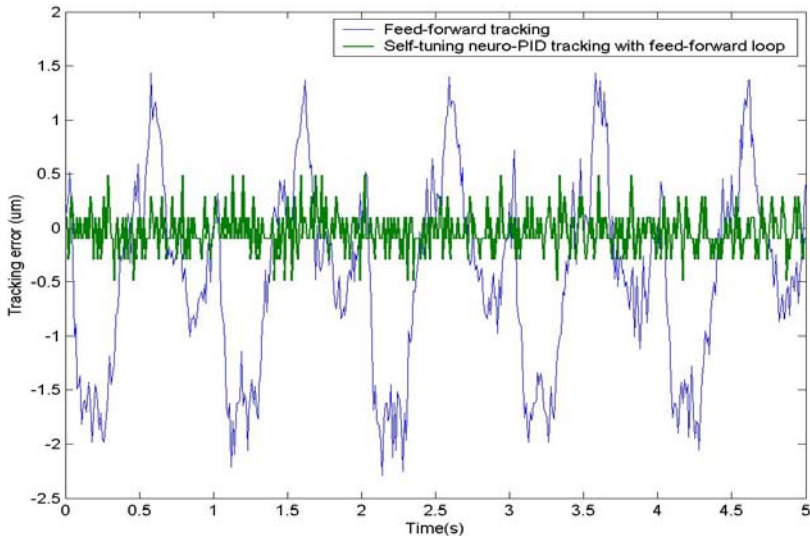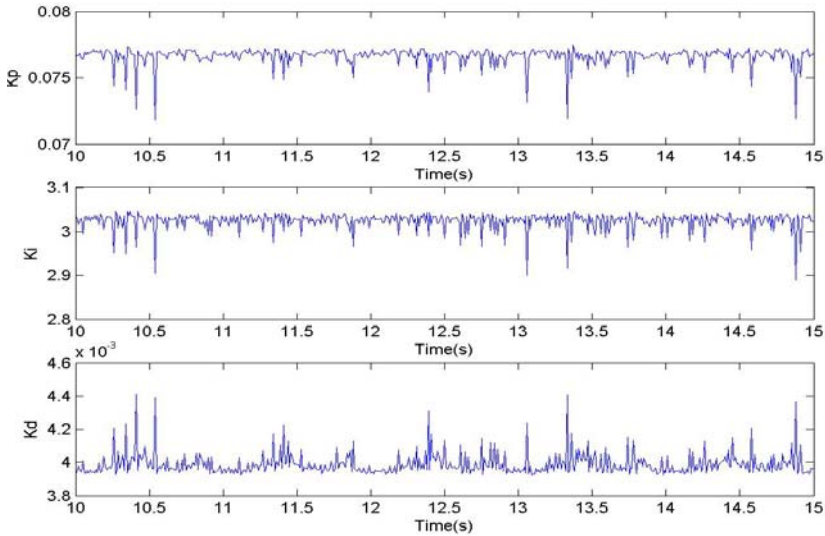In order to improve the control precision of piezoelectric actuator, the main contribution of this paper is to apply the Prandtl-Ishlinskii hysteresis model to describe hysteresis, apply LMS algorithm to identified the weights of main hysteresis loop and an inverse feedforward controller is to cascade the inverse operator with the actual hysteresis. Then, utilizing a self-tuning neuro-PID controller is introduced to suppress the tracking error due to the modeling inaccuracy and hence to get precision tracking error. Experiments were performed on a piezoelectric motion platform driven by piezoelectric actuators and the motion tracking experimental results were presented that demonstrated substantial improvements in positioning precision.

Acknowledgments

References

1. Kim, J.D., Nam, S.R.: Apiezoelectrically driven micro-positioning system for the ductile-mode griding of brittle materials. Journal of Materials Processing Technology 61, 309–319 (1999)
2. Chen, B.M., Lee, T.H., Hang, C.C., Guo, Y., Weerasorriya, S.: An H ∞ almost disturbance decoupling robust controller design for a piezoceramic bimorph actuator with Hysteresis. IEEE Transaction on Control System Technology 7(2), 160–173 (1999)

3. Adriaens, H.J.M.T.A., Koning, W.L.D., Banning, R.: Modeling piezoelectric actuators. IEEE/ASME Transactions on mechatronics 5(4), 331–341 (2000)
4. Reinder, B., de Willem, L., Han, K., Adriaens, J.M.T.A., Richard, K.: Koops: "State-space analysis and identification for a class of hysteretic systems". Automatica 37, 1883–1892 (2001)
5. Kim, J.D., Nam, S.R.: A piezoelectrically driven micro-positioning system for the ductile-mode griding of brittle materials. Journal of Materials Processing Technology 61, 309–319 (1999)
6. Michael, G., Nikola, C.: Modeling piezoelectric stack actuators for control of micromanipulation. IEEE Control Systems Magazine 17, 69–79 (1997)
7. Yu, Y., Naganathan, N., Dukkipati, R.: Preisach modeling of Hysteresis for piezoceramic actuators system. Mechanism and Machine Theory 37, 49–59 (2002)
8. Ru, C.H., Sun, L., Kong, M.X.: Adaptive inverse control for piezoelectric actuator based on hysteresis model. In: Proceedings of 2005 International Conference on Machine Learning and Cybernetics, vol. 5, pp. 3189–3193 (2005)
9. Jan, C., Hwang, C.L.: Robust control design for a piezoelectric actuator system with dominant hysteresis. In: IEEE Int. Conf. on Industrial Electronics, Control and Instrumentation, Nagoya, Japan, pp. 1515–1520 (2000)
10. Tao, G., Kokotovic, P.V.: Adaptive control of plants with unknown hysteresis. IEEE Trans. Automatic Control 40(2), 200–212 (1995)
11. Ku, S.S., Pinsopon, U., Cetinkunt, S., Nakajima, S.I.: Design, Fabrication, and Real-Time Neural Network Control of a Three-Degrees-of-Freedom Nanopositioner. IEEE/ASME Transactions on Mechatronics 5(3), 273–280 (2000)
12. Hwang, C.L., Jan, C.: A reinforcement discrete neuro-adaptive control for unknown piezoelectric actuator systems with dominant hysteresis. IEEE Trans. Neural Networks 14(1), 66–78 (2003)
13. Ang, W.T., Riviere, C.N., Khosla, P.K.: Feedforward Controller with Inverse Rate-Dependent Model for Piezoelectric Actuators in Trajectory Tracking Applications. ASME/IEEE Transactions on Mechatronics 12(2), 1–8 (2007)
14. Tan, U.X., Win, T.L., Ang, W.T.: Modeling Piezoelectric Actuator Hysteresis with Singularity Free Prandtl-Ishlinskii Model. In: IEEE International Conference on Robotics and Biomimetics, Kunming, China, pp. 251–256 (2006)
15. Tan, U.X., Win, T.L., Tanjaya, M., Wirawan, H.T., Shee, C.Y., Ang, W.T.: Real-Time Disturbance Compensation with Accelerometers & Piezoelectric-Driven Mechanism. In: IEEE International Symposium on Computational Intelligence in Robotics and Automation, Florida, USA (2007)
16. Yamamoto, T., Kaneda, M., Oki, T., Watanabe, E., Tanaka, K.: Intelligent tuning PID controllers. In: IEEE International on Systems, Man and Cybernetics, Intelligent Systems for the 21st Century, vol. 3, pp. 2610–2615 (1995)
17. Omatu, S., Yoshioka, M.: Self-Tuning Neuro-PID Control and Applications. In: IEEE International Conference on System, Man, and Cybernetics, vol. 3, pp. 1985–1989 (1997)

The Fuzzy PI Control
for the DSTATCOM Based on the Balance of
Instantaneous Power

Qun-Feng Zhu, Lei Huang, Zhan-Bin Hu, and Jie Tang

Department of Electronic Engineering, Shaoyang University, Shaoyang 422000,
Hunan Province, China
smallbeeseagull@163.com

Abstract. The DSTATCOM regulates the voltage at the point of common coupling (PCC) by injecting reactive power to the PCC, making it meet the requirement of the voltage quality. First, this paper deduces from the balance theory of instantaneous power of the DSTATCOM system the direct output voltage control strategy, in which the current detection circuit is not demanded. Compared with the cascade control strategy, it has the merits of a simple structure and fast response, but it doesn't perform well when the parameters of the system are changed. So the paper proposes the fuzzy PI control to solve the problem. The validity and effectiveness of the control strategy has been verified by theoretical analysis and digital simulation.

Keywords: DSTATCOM, Voltage regulation, Fuzzy control, Instantaneous power.

1 Introduction

This paper is focused on the control strategy for the DSTATCOM to stabilize the voltage at the point of common coupling (PCC) at the given value. The DSTATCOM regulates the voltage at the PCC by injecting reactive power to the PCC, making it meet the requirement of the voltage quality. Many control strategies have been proposed, e.g., the cascade control strategy using voltage outer loop and current inner loop, in which it is difficult to design the controller as too many PI regulators should be embodied in the control system [6-8]. Some researchers use multi-variable control theory to design the controller through establishing a state.

We would like to draw your attention to the fact that it is not possible to modify a paper in any way, once it has been published. This applies to both the printed book and the online version of the publication. Every detail, including the order of the names of the authors, should be checked before the paper is sent to the Volume Editors.

2 The Balance of Instantaneous Power of the DSTATCOM and Direct Voltage Control

Fig. 1 shows the single phase equivalent circuit of the DSTATCOM, in which three phase voltage inverter topology is applied and the direct current capacitance offers the

direct current voltage support. The distribution system of the DSTATCOM is represented by Thevenin's equivalent circuit, in which Rf and Lf stand for the equivalent resistance and impedance of the transformer and the wave filter, e for the output voltage of the DSTATCOM transformer, Upcc for the voltage at the coupling function for the DSTATCOM and using the linear approximation at the working point [9,10]. But the performance of the DSTATCOM is influenced by the changes of the system's functioning condition. So it is especially valuable to find an easy-to-be-designed high-quality control strategy.

The direct voltage control strategy for the DSTATCOM has been proposed according to the balance theory of instantaneous power [11]. It doesn't need a mutual inductor and current detection circuit and makes control easier and response faster. But its performance will worsen when the parameters are changed as it relies too much on the equivalent electronic parameters of the coupling transformer, the filter as well as the circuit. To solve the problem, this paper proposes a fuzzy-PI-based direct voltage control strategy, maintaining its merits such as structural simplicity and fast response, giving it a good self-adjusting robust features. Its validity and effectiveness is verified by the digital simulation results offered in the last part of the paper.
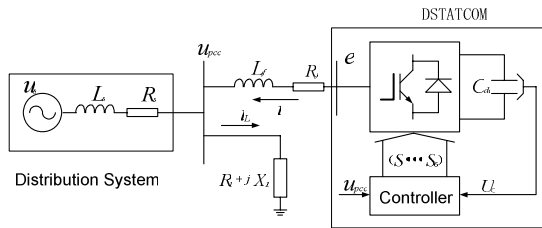
Fig. 1. Single phase equivalent circuit of the DSTATCOM

Now we'll apply the balance theory of instantaneous power to deduce the direct voltage control strategy for the DSTATCOM.

According to the balance theory, the output power of the inverter(P_e+jQ_e) should be equal to the sum of the power of the injection system($Po+jQo$) and that consumed by the equivalent resistance and impedance of the coupling transformer and the filter($Pf+jQf$), i.e.,

$$P_e = P_o + P_f .\tag{1}$$

$$Q_e = Q_o + Q_f .\tag{2}$$

Suppose the voltage of the three phase grid is symmetrical. In the two-phase synchronous rotation axis, we according to the theory of balance of instantaneous power, can get:

$$P_e = \frac{3}{2}(e_d i_d + e_q i_q) .\tag{3}$$

$$Q_e = \frac{3}{2}(e_q i_d - e_d i_q).$$ (4)

To carry out a more convenient analysis, we choose to make the synchronous rotating coordinate axis d in the coincidence with the voltage vector →Upcc of the PCC given u=the voltage vector, we can get:

$$u_d = u.$$ (5)

$$u_q = 0.$$ (6)

So:

$$P_o = \frac{3}{2} u_d i_d = \frac{3}{2} u i_d.$$ (7)

$$Q_o = -\frac{3}{2} u_d i_q = -\frac{3}{2} u i_q.$$ (8)

and the power consumed by the coupling transformer and the filter is:

$$P_f = \frac{3}{2} i^2 R_f = \frac{3}{2}(i_d^2 + i_q^2) R_f$$ (9)

$$Q_f = \frac{3}{2} i^2 \omega L_f = \frac{3}{2}(i_d^2 + i_q^2)\omega L_f.$$ (10)

Put the above formulae in the balance formulae(1), (2), we get:

$$e_d = R_f i_d - \omega L_f i_q + u.$$ (11)

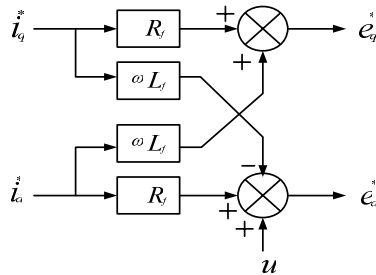$$e_q = R_f i_q + \omega L_f i_d.$$ (12)

Fig. 2. Schematics of current to voltage converting

Obviously, formulae (1), (2), as shown in the schematic diagram (Fig. 2), realize the transformation from the current id, iq to ed, eq in the dq0 coordinate system. If id, iq are the reference current, then ed, eq are the reference voltage output by the

DSTATCOM. As there is no need of a current regulator to convert the current instruction to the voltage instruction, we can clearly see the superiority of the direct voltage control strategy over the traditional voltage-outer-ring and current-inner-ring control strategy.

Fig.3 is a diagram showing the direct output voltage PI control of the DSTATCOM deduced from the aim of the DSTATCOM compensation, namely, the voltage at the coupling point and the direct current capacity.
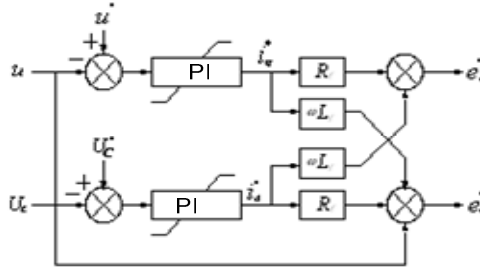

Fig. 3. Diagram of the direct output voltage control of the DSTATCOM

As shown in the diagram, after the PI regulation, the error between the reference voltage at the coupling point and the actually measured value becomes the reference reactive current signal I*q, and the error between the reference direct current capacity and its actually measured value becomes the reference active current signal I*d. After the current to voltage transformation, I*q and I*d become respectively the reference voltage signal e*q and e*d, which, as regulating signals, produce PWM drive signals thereafter.

Compared with the cascade double-loop voltage control strategy for the DSTATCOM, the control method shown in Fig. 3 doesn't need the current inner loop, thus making it more conveniently to design the voltage controller and improving the speed of the device's response. But from formulae (11), (12) and Fig. 3, we can see that the reference output voltages e*q and e*d of the DSTATCOM inverter are too closely related to the coupling transformer's and the wave filter's equivalent electrical parameters, whose changes will influence the controller's performance. Therefore, it becomes necessary to introduce a self-adjusting mechanism into the traditional PI controller to eliminate the influences. Fuzzy PI control is an effective solution to this problem. It regulates the PI parameters according to the input errors and their changing rate in the controller, giving it self-adjusting robust characteristic so as to meet the requirement of the DSTATCOM voltage controller.

3   The Fuzzy PI Control of the Direct Voltage for the DSTATCOM

Now the paper will offer the approach and process of the fuzzy PI controller's design, taking the control of the voltage at the PCC as an example, which is similar to the control of the direct capacity's voltage.
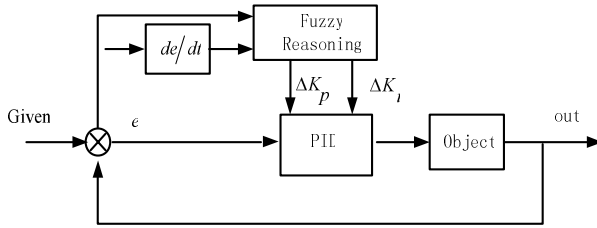
Fig. 4. Block diagram of the fuzzy-PI controller

3.1 The Principles and Structure of the Fuzzy PI Controller

Fig. 4 is the block diagram of the DSTATCOM's fuzzy PI controller. According to fuzzy mathematic theories and approaches we summarize the regulating experiences of the operators and the technologies as IF(condition)-THEN(result)-form fuzzy rules, which, together with other related data(the original PI parameters), are stored in the computer. The DSTATCOM's fuzzy PI controller inputs the error(e) between the reference voltage at the coupling point and the actually measured value, as well as the deviation(Δe) of them. Given e=Uref$-$U, Δe=Δ(Uref$-$U), then we can get ΔKp and Δki at that moment through fuzzy computing so as to realize the best regulation of the PI parameters. Then, input e, Δe and output Kp, ΔKi in the correspondent maximum domain[-6,-5,-4,-3,-2,-1,0,1,2,3,4,5,6], in which the input and output reference fuzzy sets are respectively defined as E, ΔE and ΔKp, ΔKi.

3.2 The Designing of the Fuzzy-PI Controller

The function of the regular PI controller may be described by this position arithmetic:

$$i(n) = K_p [e(n) + \frac{T}{K_i} \sum_{i=0}^{n-1} e(i)], \qquad (13)$$

In it, e(n) is the input(deviation) of the controller at the nth sampling moment; i (n) is the output of the controller(controlling value) at the nth sampling moment, and i (n-1) the output at the (n-1)th moment; T is the sampling period; Ki is the integral constant; and Kp is the proportional increment.

From formula (13), we get the incremental formula of controllers I(n) and I(n-1):

$$\Delta i(n) = i(n) - i(n-1)$$

$$= K_p [e(n) - e(n-1) + \frac{T}{K_i} e(n-1)]' \qquad (14)$$

$$\Delta e = e(n) - e(n-1). \qquad (15)$$

We apply Zigler-Nichols approach (at present a commonly-used direct solution to the PID controller's parameters) to solve K'p and K'I so as to realize the pre-adjusting of the parameters. The steps are: first increase K'p until the system vibrates given K'p=K'I=0, record the value of K'p at this critical point, i.e. Ker, and the vibrating period Ter; then can determine the controller's parameters—K'p=0.45Ker, K'I=0.83 Ter.

Divide the values of the input and output variables into 7 equal language values, i.e., {NB, NM, NS, 0, PS, PM, PB}; use the highly sensitive triangular function as the needed recognition function. In order to improve the system's robuster feature and increase the resolving power of the recognition function(shown in Fig. 5), we can make a steeper function curve around value zero.
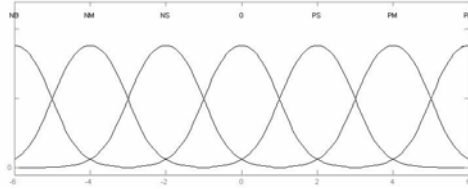
Fig. 5. Recognition function

The core of fuzzy deduction is the rules of parameter tuning for the controller that are obtained from on-the-spot regulation and experts' experiences. When the deviation appears, the controller will function to prevent it from increasing.

The actual operation is as follows:

① If |e| is big, i.e., e∈{NB, PB}, then maintain Kp at a high value to give the system a fairly good fast-tracking performance, at the same time set a limit to the value of Ki(normally Ki=0) to avoid a big over-shoot of the system's response.

② If |e| is medium, i.e., e∈ {NM, PM}, then choose to make Kp smaller and Ki larger accordingly to avoid a big over-shot of the system's response.

③ If |e| is small, i.e., e∈{NS, PS}, then Kp is to further decrease as designed, and Ki should be of a suitable value and will increase as |e| is getting smaller, so that the system will soon be stabilized, its stationary deviation will be removed, and the accuracy of control will be greatly improved.

Δe is taken into account simultaneously: when Δe is in the same direction as e and the input is deviating from the stable values, we should increase Kp to a suitable value; or in contrary conditions decrease its value.

According to the object's actual characteristics and our regulating experiences, we summarize as the following rules as shown in Table 1 and Table 2.

Table 1. Control rules of ΔKp

E / △E	NB	NM	NS	0	PS	PM	PB
NB	PB	PB	NB	PM	PS	PS	0
NM	PB	PB	NM	PM	PS	0	0
NS	PM	PM	NS	PS	0	NS	NM
0	PM	PS	0	0	NS	NM	NM
PS	PS	PS	0	NS	NS	NM	NM
PM	0	0	NS	NM	NM	NM	NB
PB	0	NS	NS	NM	NM	NB	NB

As the input e and Δe reach the equivalent language values, we can get the fuzzy values of the two corrected parameters through fuzzy formula calculation applying the tuning rules.

Table 2. Control rules of ΔKi

E \ ΔE	NB	NM	NS	0	PS	PM	PB
NB	0	0	NB	NM	NM	0	0
NM	0	0	NM	NM	NS	0	0
NS	0	0	NS	NS	0	0	0
0	0	0	NS	NM	PS	0	0
PS	0	0	0	PS	PS	0	0
PM	0	0	PS	PM	PM	0	0
PB	0	0	NS	PM	PB	0	0

After the above fuzzy deduction, we de-fuzzilize (including the maximum recognition approach, center of gravity method, etc) the two corrected parameters regulated by the fuzzy-PI controller and choose its accurate value to calculate the amount of output control. This paper applied center-of-gravity to get the accurate value of output.

$$
\begin{cases}
\Delta K_P = f_P(e,\Delta e) = \dfrac{\sum_{j=1}^{n} \mu_j(e,\Delta e)\Delta K_{Pj}}{\sum_{j=1}^{n} \mu_j(e,\Delta e)} \\[2em]
\Delta K_i = f_i(e,\Delta e) = \dfrac{\sum_{j=1}^{n} \mu_j(e,\Delta e)\Delta K_{ij}}{\sum_{j=1}^{n} \mu_j(e,\Delta e)}
\end{cases}, \tag{16}
$$

As Kp=K'p+ΔKp, Ki=K'I+ΔKi, so Kp, Ki are:

$$
\begin{cases}
K_P = K_P^* + \dfrac{\sum_{j=1}^{n} \mu_j(e,\Delta e)\Delta K_{Pj}}{\sum_{j=1}^{n} \mu_j(e,\Delta e)} \\[2em]
K_i = K_i^* + \dfrac{\sum_{j=1}^{n} \mu_j(e,\Delta e)\Delta K_{ij}}{\sum_{j=1}^{n} \mu_j(e,\Delta e)}
\end{cases}. \tag{17}
$$

4   Simulation Results

To verify the validity and effectiveness of the proposed strategy, we conducted a simulation research and then offered the simulation waveform. In the process, we produced a surge of reactive power when t=0.2s to simulate the fall of voltage at the PCC; the load went back normal when t=0.4s.

The simulation parameters of the system are:

The system's voltage: 380V; its frequency: 50Hz;
The direct current capacity: 2200 μ F, its voltage: 500V;
The SPWM Carrier frequency: 10kHz;
The equivalent inductance of the coupling transformer and the wave filter, etc: 5.5mH;
The equivalent resistance: 0.2 Ω
When applying the regular PI control, the controller's parameters are:
Parameters of the PCC: Kp=6, Ki=200;
Parameters of the direct current capacity: Kp=0.6, Ki=100.

Fig.6 shows the response curve of the Vpcc without the DSTATCOM compensation, from which we can see a fall of the Vpcc caused by the surge of reactive load. So it is necessary to compensate the voltage.
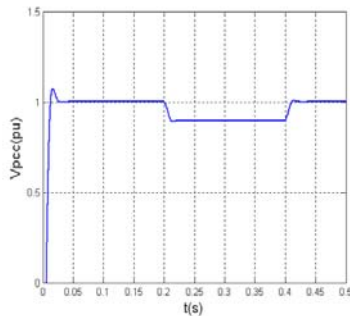

Fig. 6. Response curve of the Vpcc without the DSTATCOM compensation

Fig. 7 shows the response curves of the Vpcc under either the regular or the fuzzy PI control with the above-mentioned parameters.
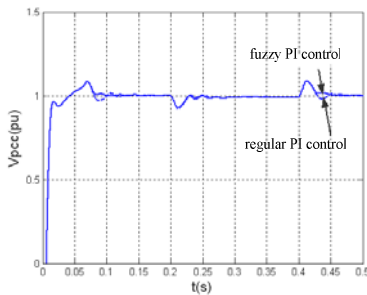

Fig. 7. Response curve of the Vpcc under fixed parameters

We can see from Fig.7 that the voltage is well compensated when the DSTATCOM is functioning, which proves the effectiveness of the direct voltage control strategy based on the balance of instantaneous power. Fig. 7 also implies that the DSTATCOM will have a high quality of control if we accordingly work out a set of PI parameters, that is to say, the regular PI control and the fuzzy PI control can be almost equally effective under the condition of fixed parameters.

To prove the fuzzy PI self-adjusting robust features, we made random changes to the equivalent electrical parameters of the coupling transformer and the wave filter with their equivalent inductance changed to 3 mH, and the equivalent resistance changed to 0.5Ω. The response curves of the Vpcc under either the regular or the fuzzy PI control are shown in Fig. 8.
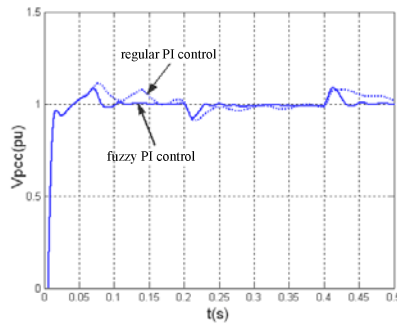
Fig. 8. Response curves of the Vpcc under changed parameters

We can see from Fig. 8 that the performance of the regular PI controller worsens under changed equivalent electrical parameters of the coupling transformer and the wave filter; but the performance of the fuzzy PI controller is not influenced at all by the changed parameters, showing a good self-adjusting robust characteristic. All these prove that the proposed fuzzy PI control strategy is valid and effective.

5   Conclusions

The DSTATCOM regulates the voltage at the PCC by injecting reactive power to it, thus making it meet the requirement of the voltage quality. According to the balance theory of instantaneous power in the DSTATCOM system, we've deduced the current-to-voltage converting relationship under the direct voltage control by the DSTATCOM, realized the no-current sensor control and so improved the DSTATCOM's response rate. To remove the drawback that the direct voltage control for the DSTATCOM based on the balance of instantaneous power is greatly influence by the changes of the equivalent parameters of the coupling transformer and the wave filter, the paper proposes the fuzzy PI control strategy for controlling the Vpcc and the voltage of the direct current capacity. Simulation results prove that the fuzzy-PI direct voltage control strategy has a good self-adjusting robust characteristic suitable for the practical engineering application of the and DSTATCOM. The next step of our research will be how to apply the proposed strategy to the actual DSTATCOM equipment.

Fund Projects

References

1. Wang, P., Jenkins, N., Bollen, M.H.J.: Experimental Investigation of Voltage Sag Mitigation by an Advanced Static var Compensator. IEEE Trans. on Power Delivery 13(4), 1461–1467 (1998)
2. Sensarma, P.S., Padiyar, K.R., Ramanarayanan, V.: Analysis and Performance Evaluation of a Distribution STATCOM for Compensating Voltage Fluctuations. IEEE Trans. on Power delivery 16(2), 259–264 (2001)
3. Yuan, J.X., Chen, B.C., Wan, L., et al.: The Method to Improve Power Quality of Distribution System by Distribution Static Compensentor. Power system technology 28(19), 81–84 (2004)
4. Li, K., Liu, J.J., Wei, B., et al.: Control and Optimization of Static var Generator for Grid Voltage Unbalance Mitigation. Proceedings of the CSEE 26(3), 58–63 (2006)
5. Lee, W.C., Lee, T.K., Hyun, D.S.: A Three-Phase Parallel Active Power Filter Operating with PCC Voltage Compensation with Consideration for an Unbalanced Load. IEEE Tran. on Power electronics 17(5), 807–814 (2002)
6. Freitas, W., Morelato, A., Xu, W., et al.: Impacts of AC Generators and DSTATCOM Devices on the Dynamic Performance of Distribution System. IEEE Tran. on Power delivery 20(2), 1493–1501 (2005)
7. Sun, J., Czarkowski, D., Zabar, Z.: Voltage Flicker Mitigation Using PWM Based Distribution STATCOM. IEEE Power Engineering Society Summer Meeting 1, 616–621 (2002)
8. Lehn, P.W.: A Benchmark System for Simulation of the D-STATCOM. IEEE Power Engineering Society Winter Meeting 1, 496–498 (2002)
9. Lehn, P.W., Iravani, M.R.: Experimental Evaluation of STATCOM Closed Loop Dynamics. IEEE Tran. on Power delivery 13(4), 1378–1384 (1998)
10. Rao, P., Crow, M.L., Yang, Z.P.: STATCOM Control for Power System Voltage Control Applications. IEEE Trans. on Power delivery 15(4), 1311–1317 (2000)
11. Chen, W.L., Hsu, Y.Y.: Direct Output Voltage Control of a Static Synchronous Compensator Using Current Sensorless d-q Vector-Based Power Balancing Scheme. Transmission and Distribution Conference and Exposition, IEEE PES 2, 545–549 (2003)
12. Wang, Z.A., Yang, J., Liu, J.J.: Harmonic Suppression and Reactive Power Compensation. Machine publishing, Beijing (1998)

A Complex Fuzzy Controller for Reducing Torque Ripple of Brushless DC Motor

Zhanyou Wang[1,*], Shunyi Xie[1], and Zhirong Guo[2]

[1] Department of Weaponry Engineering, Naval University of Engineering,
Wuhan 430033, China
[2] Department of weaponry Engineering, Naval Bengbu Petty Officer Academy,
Bengbu 233000, China
wangzhanyou75101@163.com

Abstract. The Brushless DC Motor (BLDC) has been applied widely for its high torque density, high efficiency and small size, but its torque ripple is relatively high. In recent years, some scholars apply direct torque control to BLDC, for reducing torque ripple by means of the fast response of torque. Based on this theory, this paper presents a complex fuzzy controller which comprises two fuzzy controllers. The first controller is designed to select correct voltage vector according to the torque error, stator flux-linkage error and electric angle of stator flux-linkage. The second fuzzy controller with adjustable factor is designed to regulate the action time of voltage vector according to the torque error and torque error differential. For minimizing the torque ripple, the genetic algorithm (GA) is utilized to optimize adjustable factor. The whole system is simple, the control method is convenient to be realized and the effect is significant. Simulation and experiment results verify the effectiveness of the complex fuzzy controller.

Keywords: direct torque control, brushless DC motor, voltage space vector, genetic algorithm, optimization, fuzzy control.

1 Introduction

The Brushless DC Motor (BLDC) has been applied widely for its good performance. But the torque ripple is large relatively. Conventional control strategy is current control, which essentially assumes that the torque is proportional to the phase current. However, since, in practice, the relationship is nonlinear. Various current control strategies have been adopted [1-4], but the effect is not good.

In recent years, some scholars apply direct torque control to BLDC [5-6], for reducing torque ripple by means of the direct control and fast response of torque. Based on this theory, this paper presents a complex fuzzy controller.

Firstly, the DTC of BLDC [5] is introduced in this paper. Then a complex fuzzy controller is designed based on this theory. This controller comprises two fuzzy controllers. The first controller is used to select correct voltage vector in different

---

* Corresponding author.

section. The second controller with fuzzy factor is designed to regulate the action time of voltage vector. In order to minimize the torque ripple, a real-coded genetic Algorithm [7-8] is adopted to optimize the fuzzy factor. Finally, the simulation and experiment are done, the results show that the intelligent controller proposed in this paper can reduce the torque ripple effectively and thus the torque performance is improved.

2 Direct Torque Control of Brushless DC Motor

In general, neglecting the influence of mutual coupling between the direct and quadrature axes, the electromagnetic torque of a permanent magnet brushless machine in the synchronously rotating d-q reference frame can be expressed as [5]:

$$T_e = \frac{3}{2}\frac{p}{2}(\frac{dL_d}{d\theta_e}i_{sd} + \frac{d\psi_{rd}}{d\theta_e} - \psi_{sq}) \; i_{sd} + \frac{3}{2}\frac{p}{2}(\frac{dL_q}{d\theta_e}i_{sq} + \frac{d\psi_{rq}}{d\theta_e} + \psi_{sd}) \; i_{sq} \quad (1)$$

where $\psi_{sd} = L_d i_{sd} + \psi_{rd}, \psi_{sq} = L_q i_{sq} + \psi_{rq}$, and θ_e is the rotor electrical angle, p is the number of poles, i_{sd} and i_{sq} are the d - and q -axis currents, L_d and L_q are the d - and q -axis inductances, respectively, and $\psi_{rd}, \psi_{rq}, \psi_{sd}$ and ψ_{sq} are the d - and q -axis rotor and stator flux-linkages, respectively. For non-salient machines, it can be assumed that L_d and L_q are constant and $L_d = L_q = L_s$, electromagnetic torque can be expressed as:

$$T_e = \frac{3}{2}\frac{p}{2}\left[(\frac{d\psi_{rd}}{d\theta_e} - \psi_{rq}) \; i_{sd} + (\frac{d\psi_{rq}}{d\theta_e} + \psi_{rd}) \; i_{sq}\right]. \quad (2)$$

In the stationary α-β reference frame, electromagnetic torque can be expressed as:

$$T_e = \frac{3}{2}\frac{p}{2}\left[\frac{d\psi_{r\alpha}}{d\theta_e}i_{s\alpha} + \frac{d\psi_{r\beta}}{d\theta_e}i_{s\beta}\right] \quad (3)$$

Where

$$\psi_{r\alpha} = \psi_{rd}\cos\theta_e - \psi_{rq}\sin\theta_e$$
$$\psi_{r\beta} = \psi_{rd}\cos\theta_e + \psi_{rq}\sin\theta_e \quad (4)$$

In direct torque control of Brushless DC Motor, the stator flux-linkage is controlled through terminal voltage, and terminal voltage is relative to the switches. For BLDC, according to switching states, we can get six Non-zero voltage space vectors at intervals of 60 electronic angle: U_1(100001), U_2(001001), U_3(011000), U_4(010010), U_5(000110)and U_6(100100), as shown in Fig.1 (a). As shown in Fig.1 (b), from left to right, the logical values express states of the upper and lower switch signals of phases A, B and C, respectively. The zero voltage space vector is defined as U_0 (000000).
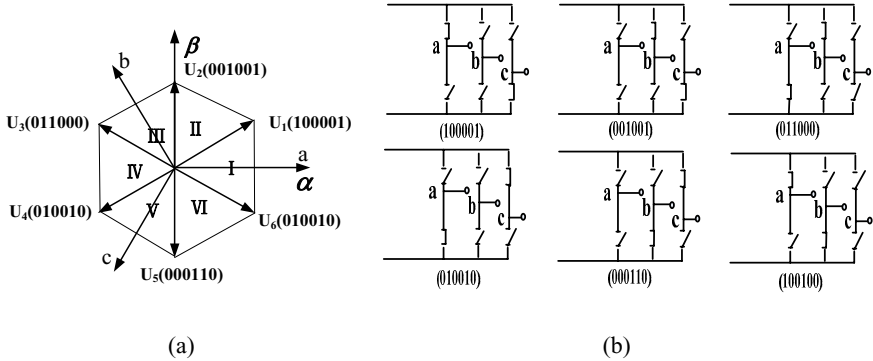
Fig. 1. Non-zero voltage space vectors of BLDC

The stator flux-linkage vector can be obtained from the measured stator voltages, $u_{s\alpha}$, $u_{s\beta}$, and currents, $i_{s\alpha}$, $i_{s\beta}$, as:

$$\begin{cases} \psi_{s\alpha} = \int \left(u_{s\alpha} - Ri_{s\alpha} \right) dt \\ \psi_{s\beta} = \int \left(u_{s\beta} - Ri_{s\beta} \right) dt \end{cases}$$
(5)

Where R is the stator winding resistance. The magnitude and angular position of the stator flux-linkage vector is:

$$\begin{cases} \psi = \sqrt{\psi_{s\alpha}^2 + \psi_{s\beta}^2} \\ \theta = arctg \left(\psi_{s\beta} / \psi_{s\alpha} \right) \end{cases}$$
(6)

The relationship between the stator flux-linkage vector and action time of voltage vector is:

$$\psi_s = \psi_{s0} + U_i \Delta t$$
(7)

where ψ_{s0} is initial flux-linkage vector, U_i is voltage vector and Δt is its action time.

From (5), (6) and (7), it can be seen that, if the correct voltage vector is selected and its action time is controlled, the amplitude and angular position of the stator flux-linkage vector can be controlled, then the torque is controlled. Now, two fuzzy controllers will be designed, one is used to select voltage vector and the other is used to control action time of voltage vector.

3 Design of Fuzzy Controller for Selecting the Correct Voltage Space Vector

From Fig.1 and (7), the fuzzy control rules are designed as:

1) If the command torque is smaller than actual torque, the non-zero voltage space vector which decreases the torque should be selected;
2) If the command torque is equal to actual torque, the voltage space vector which keeps the torque should be selected;
3) If the command torque is greater than actual torque, the non-zero voltage space vector which increases the torque should be selected.

The stator flux-linkage error E_ψ, torque error E_T and electric angle θ are selected as input of fuzzy controller, the output is voltage vector. The universe of discourse of E_ψ and E_T is divided into three fuzzy sets: P, Z, and N, and the membership function is shown in Fig.2. θ is divided into 6 parts, as shown in Fig.1 (a). Thus fuzzy control rules are available, as shown in Tab.1.
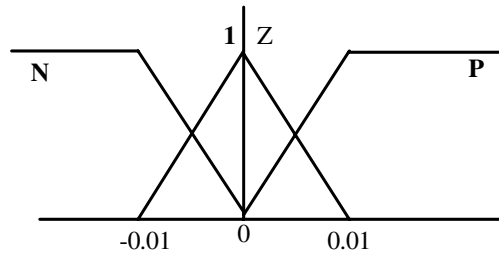
Fig. 2. The membership function of flux-linkage error and torque error

Table 1. The fuzzy control rule-table of voltage space vector

E_T	$E\psi$	I	II	III	IV	V	VI
N	**N**	$U_1(100001)$	$U_2(001001)$	$U_3(011000)$	$U_4(010010)$	$U_5(000110)$	$U_6(100100)$
	Z	$U_6(100100)$	$U_1(100001)$	$U_2(001001)$	$U_3(011000)$	$U_4(010010)$	$U_5(000110)$
	P	$U_5(000110)$	$U_6(100100)$	$U_1(100001)$	$U_2(001001)$	$U_3(011000)$	$U_4(010010)$
Z	**N**	$U_2(001001)$	$U_3(011000)$	$U_4(010010)$	$U_5(000110)$	$U_6(100100)$	$U_1(100001)$
	Z	$U_0(000000)$	$U_0(000000)$	$U_0(000000)$	$U_0(000000)$	$U_0(000000)$	$U_0(000000)$
	P	$U_4(010010)$	$U_5(000110)$	$U_6(100100)$	$U_1(100001)$	$U_2(001001)$	$U_3(011000)$
P	**N**	$U_2(001001)$	$U_3(011000)$	$U_4(010010)$	$U_5(000110)$	$U_6(100100)$	$U_1(100001)$
	Z	$U_3(011000)$	$U_4(010010)$	$U_5(000110)$	$U_6(100100)$	$U_1(100001)$	$U_2(001001)$
	P	$U_4(010010)$	$U_5(000110)$	$U_6(100100)$	$U_1(100001)$	$U_2(001001)$	$U_3(011000)$

4 Design of Fuzzy Controller for Tuning the Action Time of Voltage Space Vector

The torque ripple of Brushless DC Motor is related to the action time of voltage vector to a great extent. If the action time of voltage vector can be tuned according to the torque error E_T and torque error differential $E_T C$ on condition that the correct voltage vector is selected, the torque ripple will decrease greatly. A fuzzy controller based on adjustable factor is designed in this section, and GA is utilized to optimize adjustable factor.

The torque error E_T and torque error differential $E_T C$ are selected as input, control variable U is output. The universe of discourse of E_T, $E_T C$ and U is divided into seven fuzzy sets, it is

$$\{NB, NM, NS, 0, PS, PM, PB\}$$

or

$$\{-3, \ -2, \ -1, \ 0, \ +1, \ +2, \ +3\} \ .$$

The membership function of E_T, $E_T C$ and U is expressed as (8) and shown in Fig.3.

$$\mu(x) = \exp(-(x-c)^2 / 2b^2) \ . \tag{8}$$

Fig. 3. The membership function of torque error, error differential and control variable

The control variable U can be obtained from (9).
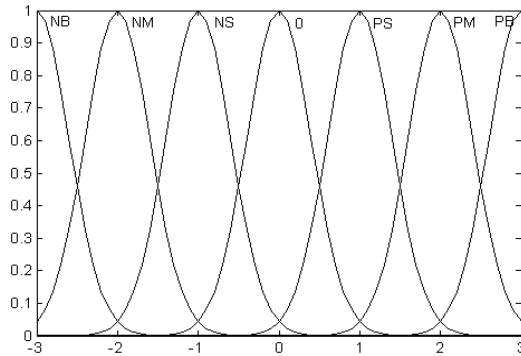
$$U = \begin{cases} \alpha_0 E_T + (1-\alpha_0) E_T C, E_T = 0 \\ \alpha_1 E_T + (1-\alpha_1) E_T C, E_T = \pm 1 \\ \alpha_2 E_T + (1-\alpha_2) E_T C, E_T = \pm 2 \\ \alpha_3 E_T + (1-\alpha_3) E_T C, E_T = \pm 3 \end{cases} \tag{9}$$

where α_0, α_1, α_2 and α_3 are adjustable factors according to the different value of E_T, they meet $\alpha_0 \leq \alpha_1 \leq \alpha_2 \leq \alpha_3$ and are restricted in $[0, 0.25]$, $[0.25, 0.5]$, $[0.5, 0.75]$ and $[0.75, 1]$, respectively.

The rules are made based on the following considerations. When the torque error E_T is very big, it indicates that the actual torque is far from target torque, so the controller should enlarge its control variable in order to make the torque attain balance state as soon as possible. In this case, the error E_T plays a more important role in control variable. However, when the error E_T is very small, it indicates that the actual torque is close to target torque. In order to make the torque steady and restrain the torque ripple as soon as possible, then the role of E_T in the control variable is small while the role of $E_T C$ becomes large.

After the fuzzy number calculation, we will get the fuzzy subset of U. According to the degree of membership, the control variable U is defuzzied and multiplied by factor k_u and we will get the accurate control variable u, viz:

$$u = K_u \cdot U \tag{10}$$

where u is the duty cycle of PWM periodicity.

In order to minimize the torque ripple, the α_0, α_1, α_2 and α_3 should be optimized. Define a dimension array $\alpha = (\alpha_0, \alpha_1, \alpha_2, \alpha_3)$ and objective function F as:

$$F = \min \frac{1}{n} \sum_{i=1}^{n} |T_m - T_i| \tag{11}$$

where T_m is objective torque, T_i is the actual torque at ith sample time.

We adopt GA to optimize objective function F. GA is a stochastic global search method that mimics the metaphor of natural biological evolution. The algorithm differs substantially from more traditional search and optimization methods as follows:

1) GA searches a population of points in parallel instead of searching for a single point;
2) GA does not require derivative information or other auxiliary knowledge, only the objective function and corresponding fitness levels influence the directions of search;
3) GA uses probabilistic transition rules instead of deterministic rules;
4) The nature of the function being optimized is immaterial: both unimodal and multimodal functions can be dealt with successfully. The parallel search capability of a GA avoids the iterations being trapped in local optimum points.

The objective function F will be optimized in a four-dimension space. The adjustable factors, α_0, α_1, α_2 and α_3, are restricted in $[0, 0.25]$, $[0.25, 0.5]$, $[0.5, 0.75]$ and $[0.75, 1]$ respectively. The optimization step of GA is follows:

1. Code the parameter;
2. The initialization of the population;
3. Evaluate the fitness of each member;
4. Selection;
5. Crossover;
6. Mutation;
7. Go to step1 until find the optimum solution.

We take the long real-valued strings as the code of chromosome. A typical coding example of a chromosome is

$$chromosome = \left(\alpha_0, \alpha_1, \alpha_2, \alpha_3 \right) \tag{12}$$

For optimizing the F , the parameters of the GA are set as follows:

1) Initial population size—80;
2) Maximum number of generations—15;
3) Probability of crossover—0.85;
4) Mutation probability—0.02;
5) Performance measure—the mean of absolute value of error between the target torque and the actual torque.

Combining two controllers, we will get a complex controller, as shown in Fig.4.
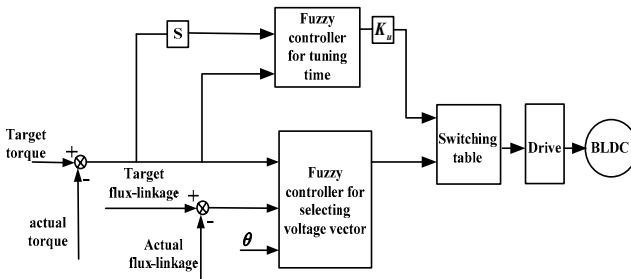
Fig. 4. The complex fuzzy controller of BLDC

5 Experiments and Simulation

Based on the above mentioned, simulations and experiments are carried out.

The experimental system includes a drive board which is designed adopting IR2135 drive chip, ADC board, a position board, a communication board, a TMS320LF2407A DSP, a PC, a BLDC. The phase currents, voltages and the inverter DC link voltage are measured and sampled by ADC board, position board is an interface between the DSP and the rotor position sensor, and the DSP is used to implement the complex fuzzy control algorithm proposed above. The BLDC is driven by drive board.

We adopt the PC to implement GA, DSP2407 is used to control BLDC and calculate function F, the array α is transferred to DSP, and DSP will implement the complex fuzzy controller and calculate function F and transfer it to PC, and they communicate via RS232. Thus, the adjustable factors are optimized off-line. At last, the optimized adjustable factors are stored in ROM of DSP, they can be used for real-time control. The optimized result is $\alpha = (0.19, 0.3, 0.7, 0.82)$.

The simulation is done with Matlab/Simulink. The target torque is 0.6Nm. The simulation and experiment results are shown in Figs.5-7. It can be seen that good agreement is achieved between simulated and experimental results. Comparing fig.6 and fig.7, we can see that the effect of complex fuzzy controller is obvious, and the torque ripple is reduced by 35 percent approximately.
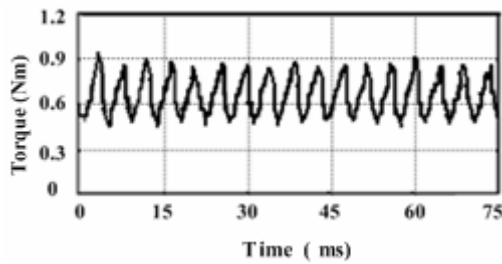
Fig. 5. The simulation result of torque with complex fuzzy controller
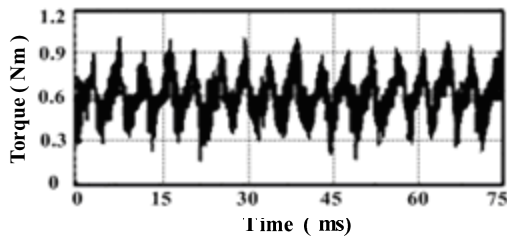
Fig. 6. Experimental result of torque with complex fuzzy controller
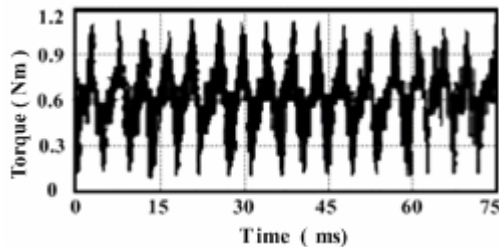
Fig. 7. Experimental result of torque with conventional control strategy

6   Conclusions

A complex fuzzy controller is proposed to restrain the torque ripple of BLDC. the design of two fuzzy controllers is described in detail. The structure of the fuzzy controller with adjustable factor is simple, the calculating methods are convenient and the effect is significant. It can restrain the torque ripple effectively. The off-line optimization and on-line tuning of the adjustable factor realize the minimization of torque ripple.

References

1. Sung, P.J., Han, W.P., Man, L.H., Harashima, F.: A New Approach for Minimum-torque-ripple Maximum-efficiency Control of BLDC Motor. IEEE Transactions on Industrial Electronics 47, 109–114 (2000)
2. French, C., Acarnley, P.: Direct Torque Control of Permanent Magnet Drives. IEEE Transactions on Industry Applications 32, 1080–1088 (1996)
3. Low, T.S., Tseng, K.J., Lee, T.H., Lim, K.W., Lock, K.S.: Strategy for the Instantaneous Torque Control of Permanent-magnet Brushless DC Drives. IEE Proceedings on Electric Power Applications 137, 355–363 (1990)
4. Low, T.S., Tseng, K.J., Lock, K.S., Lim, K.W.: Instantaneous Torque Control. In: Fourth International Conference on Electrical Machines and Drives, vol. 13-15, pp. 100–105 (1989)
5. Zhu, Z.Q.: Direct Torque Control of Brushless DC Drives with Reduced Torque Ripple. IEEE Transactions on Industry Applications 41(2), 599–605 (2005)
6. Low, T.S., Lee, T.H., Tseng, K.J., Lock, K.S.: Servo Performance of a BLDC Drive with Instantaneous Torque Control. IEEE Transactions on Industry Applications 28, 455–462 (1992)
7. Chan, P.T., Rad, A.B., Tsang, K.M.: Optimization of Fused Fuzzy Systems via Genenic Algorithm. Trans. Ind. Electron of IEEE. 49(3), 685–692 (2002)
8. Silva, W.D., Acamley, P.P., Finch, J.W.: Application of Genetic Algorithm to the Online Tuning of Electronic Drive Speed Controllers. Trans. Ind. Electron of IEEEE 47(1), 217–219 (2000)

Multiobjective Permutation Flow Shop Scheduling Using a Memetic Algorithm with an NEH-Based Local Search

Tsung-Che Chiang[1], Hsueh-Chien Cheng[2], and Li-Chen Fu[2]

[1] Department of Computer Science and Information Engineering,
National Taiwan Normal University, Taipei, Taiwan, R.O.C.
[2] Department of Computer Science and Information Engineering,
National Taiwan University, Taipei, Taiwan, R.O.C.
tcchiang@ieee.org, r96922066@ntu.edu.tw, lichen@ntu.edu.tw

Abstract. In this paper we address scheduling of the permutation flow shop with minimization of makespan and total flow time as the objectives. We propose a memetic algorithm (MA) to search for the set of non-dominated solutions (the Pareto optimal solutions). The proposed MA adopts the permutation-based encoding and the fitness assignment mechanism of NSGA-II. The main feature is the introduction of an NEH-based neighborhood function into the local search procedure. We also adjust the size of the neighborhood dynamically during the execution of the MA to strike a balance between exploration and exploitation. Forty public benchmark problem instances are used to compare the performance of our MA with that of twenty-seven existing algorithms. Our MA provides close performance for small-scale instances and much better performance for large-scale instances. It also updates more than 90% of the net set of non-dominated solutions for the large-scale instances.

Keywords: Flow shop, multiobjective, makespan, total flow time, memetic algorithm, evolutionary algorithm.

1 Introduction

Production scheduling problems have been widely studied during the last several decades due to its high problem complexity and practical value. Most production scheduling problems have NP-hard time complexity. This great challenge attracts researchers in the academia and makes production scheduling problems a good platform to test various kinds of optimization algorithms. In the industry, on the other hand, scheduling is a key to the success of a production system since it determines the machine utilization, the work-in-process (WIP) level, and so on. The high practical value also draws the attention of managers in the industry to production scheduling.

In this paper, we address scheduling of the permutation flow shop. In a flow shop, n jobs are to be processed by m machines following an identical route. In other words, each job is processed on machine 1 at first, then on machine 2, …, and finally on machine m. The permutation flow shop is a special case of the flow shop. It requires that all jobs are processed in the same order on all machines. To construct a feasible

solution (schedule), two constraints must be followed. The first constraint is that each machine can process only one job at a time; the second one is that processing of a job on machine k ($k \geq 2$) can start only after processing of the same job on machine $k-1$ is finished.

Among the feasible solutions, those which minimize (or maximize) a certain objective function are called optimal solutions. Two common objective functions are makespan (C_{max}) and total flow time (*TFT*). Both objective functions are to be minimized. Denoting the completion time of job i on machine m by C_i, the makespan is defined by $C_{max} = \max_{i=1...n} C_i$ and the total flow time is defined by $TFT = \sum_{i=1...n} C_i$.

Single-objective scheduling of the permutation flow shop in terms of either C_{max} or *TFT* is already NP-hard and has been studied extensively [1][2]. In the real-world applications, however, the decision makers are usually concerned about multiple objectives simultaneously. Due to the growth of multiobjective metaheuristics in recent years [3][4], more and more researchers joined in the field of multiobjective scheduling [5][6]. In most cases the concerned objectives are conflicting with each other. It means that the improvement on one objective will cause degradation on another one. Therefore, solving a multiobjective scheduling problem does not mean to search for a single optimal solution but for a set of solutions that are not dominated by any other one. (We say one solution x dominates another solution y if the value of each objective of x is no worse than that of y and at least one objective value of x is better than that of y.) In this study, we propose a memetic algorithm (MA), which is usually known as a hybrid of genetic algorithm (GA) and local search, to minimize the makespan and the total flow time in the permutation flow shop. In the literature this problem is usually denoted by $F \mid prmu \mid C_{max}$, *TFT*.

The rest of this paper is organized as follows. Section 2 gives a review of existing work on multiobjective permutation flow shop scheduling. The proposed MA is detailed in Section 3. The benchmark problem instances and benchmark algorithms are described in Section 4. In that section we also present the experimental results of performance comparison. Finally, we draw the conclusions and list the future research directions in Section 5.

2   Literature Review

In the early research work on multiobjective scheduling, researchers started to solve this kind of problem by trying single-objective algorithms. Multiple objective values were converted into a single objective value or were optimized in a predefined order. Rajendran [7] proposed a local search algorithm to minimize the equal-weighted sum of normalized makespan, total flow time, and machine idletime. Neppali *et al.* [8] addressed the two-stage flow shop scheduling problem with the objective of minimizing the total time subject to obtaining the optimal makespan. Sridhar and Rajendran [9] dealt with the same problem as that in [7] but based their approach on the GA framework. Cavalieri and Gaiardelli [10] solved the job allocation and sequencing problem in the flow shop by the GA. They defined a non-linear aggregation function to convert makespan and total tardiness into a single value and used this value as the fitness.

Different from the early work, Ishibuchi and Murata [11] started a new viewpoint to solve the multiobjective scheduling problem. They did not focus on searching for a single optimal solution with respect to a predefined aggregated objective function but focused on searching for the set of Pareto optimal solutions. Their approach is a genetic local search (GLS) algorithm and also takes the linear weighted sum of multiple objective values as the fitness function. However, the feature of their algorithm is to randomly specify weight values whenever a pair of parent solutions is selected. This random assignment of weight values helps the algorithm to search toward different directions on the objective space and to obtain a set of diverse solutions. Later, Ishibuchi et al. [12] improved their multiobjective GLS algorithm by choosing only good individuals to do local search and assigning the appropriate local search direction by random weight assignment and tournament selection.

Entering the new century, there was a rapid growth of multiobjective evolutionary algorithms (MOEA). Well-known MOEAs including NSGA-II [13], PESA-II [14], and SPEA2 [15] were developed. The birth of these MOEAs led to the fast progress of the research on multiobjective scheduling. Talbi et al. [16] conducted extensive experiments on testing different selection and diversity maintenance schemes. They indicated the importance of elitism and combined sharing. Chang et al. [17] proposed the gradual-priority weighting (GPW) GA for permutation flow shop scheduling. They adopted the linear weighted sum method to convert multiple objective values to the fitness value. Different from the MOGLS by Ishibuchi et al. [11], in GPWGA the weight vector of each generation is unique. By adjusting the weight vector periodically, GPWGA searches for the optimal solutions following the direction of one objective at the beginning and towards other objectives until the end. Later, Chang et al. [18] focused on minimizing makespan and total tardiness in the permutation flow shop by using a sub-population GA. The linear weighted sum method is kept. The difference is in that the population is divided into sub-populations and different weight vectors are assigned to the sub-populations.

To minimize makespan, total tardiness, and maximum tardiness, Arroyo and Armentano [19] proposed an MA. The GA part of their MA is similar to NSGA-II [13]. The main feature of their MA is the introduction of Pareto dominance to do the local search in a population-based (rather than individual-based) manner. A multi-objective simulated annealing (MOSA) algorithm was developed by Varadharajan and Rajendran [20] to minimize makespan and total flow time in the permutation flow shop. At first, two initial solutions are generated through sophisticated job-insertion-based heuristics with respect to each objective. Then, each of them is taken as the initial solution to the MOSA, whose main feature is the use of a probability function to select one of the two concerned objectives probabilistically. The probability function is varied in a way so that the entire objective space is covered uniformly. Following the work by Varadharajan and Rajendran, Pasupathy et al. [21] proposed the PGA-ALS algorithm based on the framework of MA. The GA part in PGA-ALS is also similar to NSGA-II [13]. An archive of non-dominated solutions is maintained, and each individual in the archive is subjected to the local search with insertion and swap as the neighborhood functions. Geiger [22] conducted a study of the problem structure of multiobjective permutation flow shop scheduling problems. He found that the Pareto optimal solutions are concentrated in the search space and form a "big-valley" structure just like in some single-objective optimization problems. Cheng

et al. [23] proposed an adaptive GLS algorithm, which adjusts the population size and local search intensity dynamically. Their local search adopts the insertion and swap neighborhood functions and controls the number of neighbors to be generated by a dominance-measured progress ratio. A comprehensive review and evaluation was done by Minella *et al*. [5]. They implemented twenty-three existing algorithms to solve the permutation flow shop scheduling problem regarding three two-criterion combinations. In their study, the MOSA [20] and the MOGLS [19] were regarded as the best performers.

3   The Proposed Memetic Algorithm

Before going into the details of the proposed MA, the steps of the entire algorithm are summarized as follows.

1. Generate the initial population randomly based on the encoding scheme. Assign the fitness to each individual by the NSGA-II mechanism [13]. Set $t = 1$.
2. Generate the population at generation $t+1$: Repeat the following sub-steps for $N_{POP}/2$ times, where N_{POP} is the population size.
   2.1 Select two parents by the two-tournament selection.
   2.2 Generate two offspring by the two-point linear order crossover. The swap mutation is then applied to each offspring in probability r_m.
   2.3 Select two best individuals among the two parents and two offspring and then replace the two parents.
3. If t is a multiple of T_{GEN}/T_{LS} (T_{GEN}: maximum number of generations; T_{LS}: the number of times of doing the local search), do the local search for the best $p_{LS}\%$ individuals in the population. Otherwise, go to Step 4.
4. If t is not greater than T_{GEN}, $t = t +1$; go to Step 2. Otherwise, stop.

3.1   Chromosome Encoding and Decoding

Each solution/schedule is encoded by a permutation of job indices. Given a chromosome, let $\pi(i)$ denote the job index at the ith position in the permutation, $p_{j,k}$ denote the processing time of job j on machine k, and $C_{j,k}$ denote the completion time of job j on machine k. Then, we have $C_{\pi(i),k} = \max\{C_{\pi(i-1),k}, C_{\pi(i),k-1}\} + p_{\pi(i),k}$. The values of makespan and total flow time are obtained by $C_{\pi(n),m}$ and $\sum_{i=1...n} C_{\pi(i),m}$, respectively.

3.2   Fitness Assignment and Mating Selection

After decoding a chromosome and calculating the makespan and total flow time of its corresponding schedule, the next step is to evaluate its solution quality among the population. There are various ways to evaluate chromosomes with respect to multiple objectives, and here we choose one of the most popular fitness assignment mechanisms, the mechanism of NSGA-II [13].

Mating selection refers to the action of selecting parents to produce the offspring (new individuals). We use the 2-tournament selection, which selects two individuals randomly and takes the better one as the parent.

3.3   Crossover and Mutation

Crossover and mutation are the means by which the parents produce the offspring. There are many kinds of crossover and mutation operators for the permutation-based encoding. We use the two-point linear order crossover (LOX) and the swap mutation. The two-point LOX first chooses two random cut points and passes the section enclosed by the cut points from one parent to one offspring. Next, the un-passed job indices are put in the unfilled positions in the order of their occurrence in the other parent. Given two parents 12345 and 54132, for example, assuming that the cut points are at positions two and four, the two offspring produced by LOX are 52341 and 24135. The swap mutation chooses two positions randomly and swaps their values.

3.4   Environmental Selection

Most environmental selection mechanisms follow the principle of "survival of the fittest" to raise the solution quality and add strategies like immigration (replacing current individuals with randomly-created individuals) to keep the population diversity. In our previous work [24], we proposed a 2/4-replacement mechanism and found that it provided good performance. Whenever two offspring are produced by two parents, we re-calculate the fitness for each of them. The best two individuals will replace the two parents in the population. Doing the replacement avoids the rapid growth of individuals with similar gene structures and the loss of population diversity. The 2/4-replacement mechanism implicitly implements the elitism strategy since it puts the best two individuals among the two parents and two offspring back to the population.

3.5   NEH-Based Local Search

The MA integrates the GA and the local search in order to obtain the advantages of both algorithms, the exploration ability of the GA and the exploitation ability of the local search. One key point in the local search procedure is the neighborhood function, which generates the neighboring solutions around the current solution. Since the local search aims at searching intensively in a (small) promising region on the solution space, the neighboring solutions are usually generated by making a small modification to the current solution. In the literature, the insertion (also called shift) function [11][16][19][20][23] and the swap function [12][21][22][23] are two popular neighborhood functions. In our preliminary experiments, however, we found that the size of the neighborhood built by the insertion and the swap functions might be too small to search for new good solutions. Hence, we propose an NEH-based neighborhood function.

The NEH heuristic [25] is a well-known heuristic for minimizing makespan in the permutation flow shop. It firstly sorts all jobs in non-increasing order of the total processing time. All n jobs are marked as unscheduled. Then, it follows this order and iteratively inserts unscheduled jobs one by one into the best partial sequence of scheduled jobs. Given the best partial sequence at iteration t, the NEH inserts the $(t+1)$-th unscheduled job into $t+1$ positions while keeping the relative orders of the previous t scheduled jobs unchanged. Finally, among the $t+1$ partial sequences, the one whose corresponding schedule has the shortest makespan is chosen as the best

partial sequence at iteration $t+1$. The inserted job is marked as scheduled. Repeat the above steps for n times, the NEH obtains a complete schedule with high quality.

Integrating the NEH heuristic and metaheuristics to solve permutation flow shop scheduling problems is usually through the use of the NEH to generate the initial solutions for the metaheuristics [9][20][21]. In this study, the NEH and the MA are integrated more tightly – we take the idea of the NEH to generate the neighborhood in the local search procedure. The integration is achieved through several modifications to the NEH. First, we do not start the NEH with an empty set of scheduled jobs. Since individuals evolve during execution of the MA, we think that most parts of the job sequences recorded on the individuals are correct. Besides, re-constructing the entire job sequence could be computationally expensive. Therefore, we do the NEH-based local search by extracting a set of jobs and re-inserting them back while maintaining the relative positions between non-extracted jobs. The number of extracted jobs is controlled by the parameter N_{NEH} (maximum variation size) and the evolution progress. Based on our experience, it is a good strategy to do exploration (using a large neighborhood) at the beginning of the evolution and to do exploitation (using a small neighborhood) near the end of the evolution. Thus, we define the number of extracted jobs n_{NEH} by $n_{NEH} = \min\{N_{NEH}\cdot(1 - gen/T_{GEN}), 1\}$, where gen is the current generation number and T_{GEN} is the maximum number of generations. Second, the order in which the extracted jobs are re-inserted back does not follow the longest-total-processing-time-first principle. Again, we expect that the MA can find good job sequences, and we can rely on the MA to determine the orders for re-inserting the extracted jobs. Third, we do not focus only on the makespan objective. To deal with multiple objectives, we randomly choose one of the two concerned objective functions and select the best partial sequence according to the chosen objective function. Fourth, at the last iteration of the NEH, not only one best solution is selected. All non-dominated solutions among the n solutions are stored. At the end of the local search procedure, the stored solutions that are non-dominated by any solution in the current population will replace the worst solutions in the population.

We use Fig. 1 to explain the proposed NEH-based local search. Given an individual in the population, we randomly choose two cut points and extract the segment of jobs enclosed by the cut points. (We assume that n_{NEH} is two.) Here, jobs 2 and 3 are extracted, and the relative orders of non-extracted jobs 1 and 4 will be maintained. Extracted jobs 2 and 3 will be re-inserted following their relative orders, i.e., job 2 and then job 3. At the first iteration, job 2 is inserted into three positions.
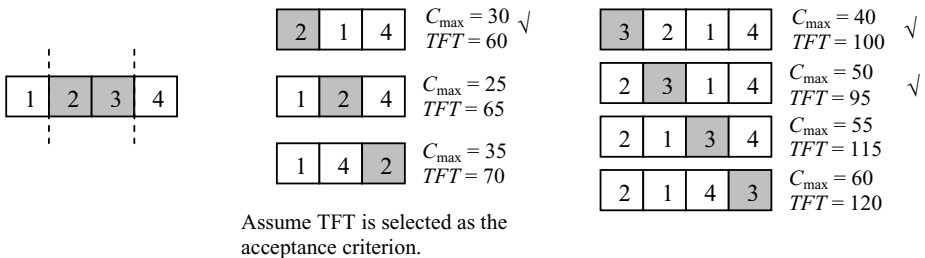

Fig. 1. An illustration of the NEH-based local search

Assuming that *TFT* is chosen as the acceptance criterion, the first sequence 214 is the best partial sequence. Then, we take the sequence 214 and insert job 3 into four positions. Since this is the last iteration (all extracted jobs are inserted), we collect all non-dominated solutions. In this example, sequences 3214 and 2314 will be stored.

3.6 Initial Population and Stopping Criterion

The individuals in the initial population are generated randomly. Every individual is a random permutation of n indices of jobs. The proposed MA stops when a maximum number (T_{GEN}) of generations is reached. The non-dominated individuals in the population at the last generation are taken as the output of the MA.

4 Experiments and Results

4.1 Benchmark Problem Instances and Algorithms

Problem instances in our experiments are taken from Taillard's benchmark data set [27], which is one of the most commonly adopted benchmark data sets for shop scheduling. We take forty problem instances, TA11−30 and TA41−60. The problem scale ($n \times m$) is 20×10, 20×20, 50×10, and 50×20 for instances TA11−20, TA21−30, TA41−50, and TA51−60, respectively. They are chosen since they were used in several existing work [5][20][21][26], and the net sets of non-dominated solutions for these instances are available either in the published papers or on the author's website. To verify the performance of the proposed MA, we compare the net sets of non-dominated solutions generated by our MA with those generated by the four above-mentioned studies.

In [5], the net sets of non-dominated solutions were generated by twenty-three existing algorithms, including algorithms proposed in [13], [14], [15], [19], and [22]. In [20], the net sets were generated by their proposed MOSA and four existing algorithms [11][17][28][29]. The net sets in [21] were generated by the proposed PGA-ALS and three existing algorithms [11][17][28]. Finally, the net sets in [26] were generated by the proposed DPSO and the four algorithms in [21]. Thanks to the provision of the net sets of non-dominated solutions in these studies, we can evaluate the proposed MA by comparing with twenty-seven algorithms ([11], [17], [26], [29], and twenty-three algorithms in [5]) conveniently.

4.2 Performance Metrics

Three performance metrics are taken to do performance evaluation. The first two metrics, I_{ε} and I_H, are used following the instructions in [5]. Given two sets, A and B, of solutions, the binary epsilon indicator $I_{\varepsilon}(A, B)$ equals the minimum factor ε such that any solution in B is ε-dominated by at least one solution in A. (For more details of ε-dominance, see [30].) In the experiment, we calculate the $I_{\varepsilon}(A, R)$ value for the net set of non-dominated solutions obtained by each tested algorithm A, where R is the net set of non-dominated solutions obtained by all tested algorithms. The objective values of each solution x are normalized into values in the interval [1, 2] by $f_j'(x) =$

$(f_j(x) - f_j^{\min})/(f_j^{\max} - f_j^{\min}) + 1$, where $f_j(x)$ denotes the jth objective value of x and $f_j^{\min}/f_j^{\max}$ denotes the minimum/maximum of the jth objective value among the solutions in the union of the net sets of non-dominated solutions obtained by all tested algorithms. The normalized I_ε indicator is defined by $I_\varepsilon(A, R) = \max_{y \in R} \min_{x \in A} \max_{1 \leq j \leq 2} \{f_j'(x)/f_j'(y)\}$. It will take values between 1 and 2, and a small I_ε value indicates better solution quality. The second metric, the hypervolume indicator I_H, measures the area covered by a set of solutions in the case of two objectives. We calculate the I_H value for the net set of non-dominated solutions obtained by each tested algorithm. The I_H values are calculated based on the objective values normalized into [0, 1]. The worst objective values are set by 1.2, and hence the maximum I_H value is $1.2^2 = 1.44$. The third metric I_{PF} is the fraction of Pareto front that is covered by the net set of non-dominated solutions (A) obtained by each tested algorithm. Since the true Pareto front is not known, we use the reference set R instead. Thus, the third metric is defined by $I_{PF}(A) = |A \cap R|/|R|$. A larger value for the I_H and I_{PF} indicators means better solution quality.

4.3 Parameter Setting

There are six parameters in the proposed MA: The population size (N_{POP}), the maximum number of generations (T_{GEN}), the percentage of population to do the local search (p_{LS}), the number of times of doing the local search (T_{LS}), the mutation rate (r_m), and the maximum variation size in the NEH-based neighborhood function (N_{NEH}). We pick up two problem instances randomly from each of the four problem categories with different scales. The four instances with 20 jobs ($n=20$) form the data set S_1, and the other four instances with 50 jobs ($n=50$) form the data set S_2. We test various combinations of parameter values on data sets S_1 and S_2. The best parameter settings for S_1 and S_2 are applied to solve instances TA11−30 and TA41−60, respectively.

We set the mutation rate by 0.05 following the past experience. In the proposed MA, the variation size in the NEH-based neighborhood function is initialized by N_{NEH} and decreases gradually during the execution. We call it the D-strategy for its dynamic adjustment of the variation size. To see the effect of the D-strategy, we also test another strategy, which fixes the variation size as N_{NEH} throughout the execution of the MA. We call it the S-strategy for its static nature. The tested values for the parameters and the strategy are summarized in Table 1.

Table 1. The tested values for the parameters in the proposed MA

Problem set	$N_{POP} \times T_{GEN}$	$p_{LS}\%$	T_{LS}	N_{NEH}	Strategy
S_1 ($n=20$)	{50×4000, 100×2000, **200×1000**}	{**10%**, 15%, 20%}	{50, **100**, 200}	{2, 6, **10**}	{S, **D**}
S_2 ($n=50$)	{100×5000, **200×2500**, 250×2000}	{10%, 15%, **20%**}	{50, 100, **200**}	{2, 5, **10**}	{S, **D**}

For each data set, we select the best parameter setting among the $3 \times 3 \times 3 \times 3 \times 2 = 162$ settings according to the I_ε and I_H values. The value of each parameter in the best setting is marked in boldface in Table 1. From the experimental results on either data

set, we also observe that among the best ten parameter settings at least seven settings use the D-strategy. First, it confirms that the D-strategy has better performance than the S-strategy. Second, since each strategy is tested with an equal number of combinations (162/2=81) of four parameters, the observation reveals that it is easier to find a good setting of parameter values when the D-strategy is used.

4.4 Performance Comparison

After determining the best parameter settings, the proposed MA is applied to solve each of the forty problem instances for twenty runs. For each instance, the net set of non-dominated solutions over twenty runs is collected and is then compared with the net set obtained by the benchmark algorithms. The I_ε, I_H, and I_{PF} values are calculated for each tested algorithm for each instance. The average values of I_ε, I_H, and I_{PF} metrics for each problem category are reported in Table 2, 3, and 4, respectively. In these tables, a cell is marked by "−" if the net set of non-dominated solutions of the corresponding problem category is not available.

From the experimental results, we can see that on average the proposed MA is the best performer. The Combo23 algorithm, which collects the net set of non-dominated solutions obtained by twenty-three algorithms, is the second one. Although we do not compare our MA with the other three benchmark algorithms using all problem categories due to the lack of data, our MA significantly outperforms them in one or two problem categories. When small-scale problem instances (TA11−30) are solved, the best two algorithms have close performance. The proposed MA slightly outperforms Combo23 on TA21−30, whereas Combo23 slightly outperforms the proposed MA on TA11−20. When large-scale problems instances (TA41−60) are solved, the proposed MA shows much better performance than Combo23 does. In Table 2, the average I_ε value of Combo23 is around 1.1, which means that the deviation percentage of the normalized objective values of the solutions provided by Combo23 from the normalized objectives values of the solutions in the reference set is up to 10%. By contrast, the average I_ε value of the proposed MA is less than 1.02, which indicates that the maximum deviation percentage is less than 2%.

Table 2. The average I_ε values of tested algorithms for each category of problem instances

	TA11−20	TA21−30	TA41−50	TA51−60	Average
	20×10	20×20	50×10	50×20	
MOSA [20]	−	1.101	−	1.212	−
PGA-ALS [21]	−	−	1.245	1.233	−
DPSO [26]	−	−	1.320	−	−
Combo23 [5]	1.012	1.027	1.093	1.115	1.062
Proposed MA	1.021	1.019	1.015	1.005	1.015

The I_H results in Table 3 are consistent with the I_ε results. When large-scale problems are solved, the average I_H values of the proposed MA are greater than those of Combo23 by around 10%. The large improvement of the proposed MA on the I_ε and I_H values can be explained by the average I_{PF} values in Table 4. When the problem instances TA41−50 are solved, our MA finds more than 90% of the solutions

Table 3. The average I_H values of tested algorithms for each category of problem instances

	TA11−20 20×10	TA21−30 20×20	TA41−50 50×10	TA51−60 50×20	Average
MOSA [20]	–	1.137	–	1.012	–
PGA-ALS [21]	–	–	0.959	0.971	–
DPSO [26]	–	–	0.845	–	–
Combo23 [5]	1.176	1.214	1.224	1.199	1.203
Proposed MA	1.173	1.218	1.343	1.353	1.272

Table 4. The average I_{PF} values of tested algorithms for each category of problem instances

	TA11−20 20×10	TA21−30 20×20	TA41−50 50×10	TA51−60 50×20	Average
MOSA [20]	–	21.4 %	–	0 %	–
PGA-ALS [21]	–	–	0 %	0 %	–
DPSO [26]	–	–	0 %	–	–
Combo23 [5]	90.9 %	85.3 %	8.9 %	3.5 %	47.2 %
Proposed MA	86.6 %	86.7 %	91.1 %	96.5 %	90.2 %

in the reference set, but Combo23 only finds less than 10%. When the problem instances TA51−60 are solved, our MA finds more than 95% of the solutions in the reference set, but Combo23 finds less than 5%. Due to the limitation of space, the complete list of solutions in the reference set is not provided here. Readers who are interested in comparing their algorithms with ours are welcome to contact the first author to get the electronic file of the list. The proposed MA is implemented in C++ with Microsoft Visual Studio 2005. The test is conduced on personal computers with Intel 3.0 GHz CPU and 2 GB RAM. The average computation times for solving problem categories TA11−30 and TA51−60 are 21 s and 49 s, respectively.

Table 5. Average deviation percentage from the best known solutions

	TA11−20 20×10	TA21−30 20×20	TA41−50 50×10	TA51−60 50×20	Average
C_{max}	0.3 %	0.3 %	1.0 %	1.9 %	0.88 %
TFT	0 %	0 %	0.7 %	0.6 %	0.33 %

In Table 5, we also provide the average deviation percentage of the minimum C_{max} and *TFT* values of the solutions obtained by the proposed MA from the best known solutions reported in [31]−[33]. The average deviation percentage for C_{max} is only 0.3% for small-scale instances and less than 2% for large-scale instances. For the *TFT* values, our MA finds the best known solutions for all small-scale instances. The average deviation percentage for large-scale instances is at most 0.7%.

In summary, we compare the solutions obtained by the proposed MA with those obtained by twenty-seven multiobjective algorithms using forty problem instances. Our MA provides close performance for twenty small-scale instances and much better performance for twenty large-scale instances. Most of the twenty-seven benchmark algorithms focused on the fitness assignment and/or control of searching directions but not the way to generate new solutions. In this work, we try another research

direction. We introduce an NEH-based neighbourhood function into the local search procedure in the MA. The size of the neighbourhood is adjusted dynamically during the execution of the MA. The experimental results confirm that the proposed dynamic NEH-based local search is useful to find high-quality solutions, especially when dealing with the large-scale and thus large-solution-space problem instances. Furthermore, the average deviation percentage of $C_{\max}$ and *TFT* values provided by our MA from the best known solutions is no more than 2%. Note that these best known solutions are generated by algorithms dedicated to one objective only. By contrast, our MA shows the ability to minimize two objectives simultaneously.

5   Conclusions

In this study, we propose an MA to minimize the makespan and total flow time in the permutation flow shop. The aim is to find the set of Pareto optimal solutions. We design a neighborhood function following the idea of the NEH procedure and use this neighborhood function in the local search procedure in our MA. We also adjust the size of neighborhood dynamically. Through experiments on forty widely-used benchmark problem instances, the results show that (1) the dynamic adjustment of the size of neighborhood makes the parameter tuning process easier; (2) the proposed MA provides close performance for small-scale problem instances and significantly better performance for large-scale problem instances, compared with twenty-seven existing algorithms in terms of three performance metrics (epsilon indicator, hypervolume, and the fraction of reached non-dominated solutions); (3) more than 90% of the net sets of non-dominated solutions for the large-scale problem instances are updated by the proposed MA.

   This study will be extended following three directions: (1) We want to develop an adaptive mechanism to set the values of algorithm parameters and reduce the workload of parameter tuning. (2) More advanced fitness assignment mechanisms will be tried in our algorithm. (3) A comprehensive experiments on the whole set of Taillard's problem instances is to be conducted.

Acknowledgments

References

1. Framinan, J.M., Leisten, R., Ruiz-Usano, R.: Comparison of Heuristics for Flowtime Minimization in Permutation Flowshops. Computers & Operations Research 32, 1237–1254 (2005)
2. Ruiz, R., Maroto, C.: A Comprehensive Review and Evaluation of Permutation Flowshop Heuristics. European Journal of Operational Research 165, 479–494 (2005)

3. Deb, K.: Multiobjective Optimization Using Evolutionary Algorithms. Wiley, U.K. (2001)
4. Coello, C.A.C., Van Veldhuizen, D.A., Lamont, G.B.: Evolutionary Algorithms for Solving Multi-Objective Problems. Kluwer Academic Publishers, Dordrecht (2007)
5. Minella, G., Ruiz, R., Ciavotta, M.: A Review and Evaluation of Multi-objective Algorithms for the Flowshop Scheduling Problem. INFORMS Journal of Computing 20, 451–471 (2008)
6. Lei, D.: Multi-objective Production Scheduling: A Survey. International Journal of Advanced Manufacturing Technology (in press)
7. Rajendran, C.: Heuristics for Scheduling in Flowshop with Multiple Objectives. European Journal of Operational Research 83, 540–555 (1995)
8. Neppalli, V.R., Chen, C.L., Gupta, J.N.D.: Genetic Algorithms for the Two-stage Bicriteria Flowshop Problem. European Journal of Operational Research 95, 356–373 (1996)
9. Sridhar, J., Rajendran, C.: Scheduling in Flowshop and Cellular Manufacturing Systems with Multiple Objectives—A Genetic Algorithmic Approach. Production Planning & Control 7, 374–382 (1996)
10. Cavalieri, S., Gaiardelli, P.: Hybrid Genetic Algorithms for a Multiple-objective Scheduling Problem. Journal of Intelligent Manufacturing 9, 361–367 (1998)
11. Ishibuchi, H., Murata, H.: A Multi-objective Genetic Local Search Algorithm and Its Application to Flowshop Scheduling. IEEE Transactions on Systems, Man, and Cybernetics – Part C, 392–403 (1998)
12. Ishibuchi, H., Yoshida, T., Murata, T.: Balance Between Genetic Search and Local Search in Memetic Algorithms for Multiobjective Permutation Flowshop Scheduling. IEEE Transactions on Evolutionary Computation 7, 204–223 (2003)
13. Deb, K., Pratap, A., Agarwal, S., Meyarivan, T.: A Fast and Elitist Multiobjective Genetic Algorithm: NSGA-II. IEEE Transactions on Evolutionary Computation 6, 182–197 (2002)
14. Corne, D.W., Jerram, N.R., Knowles, J.D., Oates, M.J.: PESA-II: Region-based Selection in Evolutionary Multiobjective Optimization. In: Spector, et al. (eds.) GECCO 2001, pp. 283–290 (2001)
15. Zitzler, E., Laumanns, M., Thiele, L.: SPEA2: Improving the Strength Pareto Evolutionary Algorithm. Technical Report 103, Computer Engineering and Network Laboratory (TIK), ETH, Zurich, Switzerland (2001)
16. Talbi, E.-G., Rahoual, M., Mabed, M.H., Dhaenens, C.: A hybrid evolutionary approach for multicriteria optimization problems: Application to the flow shop. In: Zitzler, E., Deb, K., Thiele, L., Coello Coello, C.A., Corne, D.W. (eds.) EMO 2001. LNCS, vol. 1993, pp. 416–428. Springer, Heidelberg (2001)
17. Chang, P.C., Hsieh, J.C., Lin, S.G.: The Development of Gradual-priority Weighting Approach for the Multi-objective Flowshop Scheduling Problem. International Journal of Production Economics 79, 171–183 (2002)
18. Chang, P.C., Chen, S.H., Liu, C.H.: Sub-population Genetic Algorithm with Mining Gene Structures for Multiobjective Flowshop Scheduling Problems. Expert Systems with Applications 33, 762–771 (2007)
19. Arroyo, J.E.C., Armentano, V.A.: Genetic Local Search for Multi-objective Flowshop Scheduling Problems. European Journal of Operational Research 167, 717–738 (2005)
20. Varadharajan, T.K., Rajendran, C.: A Multi-objective Simulated-annealing Algorithm for Scheduling in Flowshops to Minimize the Makespan and Total Flowtime of Jobs. European Journal of Operational Research 167, 772–795 (2005)

21. Pasupathy, T., Rajendran, C., Suresh, R.K.: A Multi-objective Genetic Algorithm for Scheduling in Flow Shops to Minimize the Makespan and Total Flow Time of Jobs. International Journal of Advanced Manufacturing Technology 27, 804–815 (2006)
22. Geiger, M.J.: On Operators and Search Space Topology in Multi-objective Flow Shop Scheduling. European Journal of Operational Research 181, 195–206 (2007)
23. Cheng, H.C., Chiang, T.C., Fu, L.C.: Multiobjective Permutation Flowshop Scheduling by an Adaptive Genetic Local Search Algorithm. In: IEEE Congress on Evolutionary Computation, pp. 1596–1602 (2008)
24. Chiang, T.C., Fu, L.C.: A Rule-centric Memetic Algorithm to Minimize the Number of Tardy Jobs in the Job Shop. International Journal of Production Research 46, 6913–6931 (2008)
25. Nawaz, M., Enscore Jr., E.E., Ham, I.: A Heuristic Algorithm for the m-machine, n-job Flowshop Sequencing Problem. OMEGA 11, 91–95 (1983)
26. Guo, W., Chen, G., Huang, M., Chen, S.: A Discrete Particle Swarm Optimization Algorithm for the Multiobjective Permutation Flowshop Sequencing Problem. In: Cao (ed.) Fuzzy Information and Engineering, vol. 40, pp. 323–331 (2007)
27. Taillard, E.: Benchmarks for Basic Scheduling Problems. European Journal of Operational Research 64, 278–285 (1993)
28. Bagch, T.P.: Multiobjective Scheduling by Genetic Algorithms. Kluwer, Boston (1999)
29. Framinan, J.M., Leisten, R., Ruiz-Usano, R.: Efficient Heuristics for Flowshop Sequencing with the Objectives of Makespan and Flowtime Minimization. European Journal of Operational Research 141, 559–569 (2002)
30. Laumanns, M., Thiele, L., Deb, K., Zitzler, E.: Combining Convergence and Diversity in Evolutionary Multiobjective Optimization. Evolutionary Computation 10, 263–282 (2002)
31. Zobolas, G.I., Tarantilis, C.D., Ioannou, G.: Minimizing Makespan in Permutation Flow Shop Scheduling Problems Using a Hybrid Metaheuristic Algorithm. Computers & Operations Research 36, 1249–1267 (2009)
32. Zhang, Y., Li, X., Wang, Q.: Hybrid Genetic Algorithm for Permutation Flowshop Scheduling Problems with Total Flowtime Minimization. European Journal of Operational Research 196, 869–876 (2009)
33. Tseng, L.Y., Lin, Y.-T.: A Hybrid Genetic Local Search Algorithm for the Permutation Flowshop Scheduling Problem. European Journal of Operational Research 198, 84–92 (2009)

Intelligent Nonlinear Friction Compensation Using Friction Observer and Backstepping Control

Seong Ik Han[1], Chan Se Jeong[2], Sung Hee Park[2], Young Man Jeong[2], Chang Don Lee[2], and Soon Yong Yang[3]

[1] Dept. of Electrical Automation, Suncheon First College, 540-744,Cheonnam, Korea
[2] Graduate School of Mechanical and Automotive Engineering, University of Ulsan, Ulsan 680-749, Korea
[3] School of Mechanical and Automotive Eng., Ulsan University, 680-749, Ulsan, Korea
snikhan@gmail.com, worwordl81@nate.com,
sunghee876@mail.ulsan.ac.kr, messria@nate.com,
changdon@paran.com, soonyy@ulsan.ac.kr

Abstract. In this article, a robust nonlinear friction control strategy is developed using friction observer and recurrent fuzzy neural network. The adaptive dynamic friction observer based on the LuGre friction model is proposed to estimates the friction parameters and a directly immeasurable friction state variable. A RFNN approximator and reconstructed error compensator is also designed to give additional robustness to the control system due to the presence of the friction model uncertainty. A proposed composite control scheme with basic basckstepping controller is applied to the position tracking control of the servo mechanical system.

Keywords: Nonlinear friction, LuGre friction model, Backstepping control, Friction observer, RFNN.

1 Introduction

A friction frequently acts on the main obstacle in diverse servo control systems since friction properties are varied according to several conditions such as time, temperature and relative velocity, and some properties cannot be completely understood. Thus, the influence of friction on the control system must be identified and adequately compensated in order to improve the transient performance and reduce the steady-state tracking errors. In recent years, the dynamic friction models [1-2] have studied that capture dynamic friction phenomena in the presliding range such as memory hysteresis, etc.

The LuGre model, suggested by Canudas de Wit et al. [1], is the friction model which can represent several useful and important friction properties by only a few parameters. Until now, most of the friction compensation methods using friction observer have tried to compensate the friction by combing basic controller with friction observer based on fixed friction model. Even though this method has provided many useful results, more improved result of friction compensation can be expected if

a robust compensator is supplemented in control system in addition to the friction observer. This uncertainty of the system can be approximated well by a fuzzy or neural networks scheme.

In recent days, much attention in identification and control of complex nonlinear dynamic systems has been increased using fuzzy logic and neural networks [3-4]. The fuzzy neural network (FNN) combines the merits both the low-level learning of neural network and the high-level human knowledge representation and thinking of fuzzy theory. On the other hand, the recurrent fuzzy neural network (RFNN) [5-6] is a dynamic mapping and has a good control performance in the presence of uncertainty such as parameter variations of the system, external load, and unmodeled dynamics compared to the feedforward FNN.

In this article, we design the adaptive dynamic friction observer based on the LuGre friction model with the backstepping control. Next, for a friction uncertainty, a RFNN approximator and a reconstructed error compensator are also designed to guarantee robustness of the control system and complement the friction observation error. The performance of a proposed control scheme will be evaluated by experiment for the servo mechanical system with nonlinear dynamic friction.

2   Design Controller and Observer

2.1   Design Backstepping Contoller and Friction Observer

The dynamic model for the servo system including the nonlinear friction is

$$J\ddot{\theta} + T_f = u \tag{1}$$

where J is a moment of inertia, θ is the angular position, and T_f is the nonlinear dynamic friction. The well-known LuGre friction model [1] is represented as

$$\dot{z} = \dot{\theta} - f(\theta,\dot{\theta})z \tag{2}$$

where

$$f(\theta,\dot{\theta}) = \alpha(\dot{\theta},\dot{\theta})\left|\dot{\theta}\right| \tag{3}$$

$$\alpha(\theta,\dot{\theta}) = \sigma_0 / T_c + (T_s - T_c)e^{-(\dot{\theta}/\dot{\theta}_s)^2} \tag{4}$$

and T_c is Coulomb friction, T_s is stiction level, $\dot{\theta}_s$ is Stribeck angular velocity. The function $\alpha(\cdot)$ in the LuGre model has no terms which explicitly account for position dependence of the friction force. The dynamic friction term excluding viscous friction torque is described by

$$T_f = \sigma_0 z + \sigma_1 \dot{z} + \sigma_2 \dot{\theta} \tag{5}$$

where σ_0 is the stiffness of the elastic bristle, σ_1 is a damping coefficient in elastic range and σ_2 is the viscous friction coefficient. The new states are defined as follows in order to design the backstepping controller:

$$z_1 = \theta_d - \theta \tag{6}$$

$$z_2 = \dot{\theta} + \alpha \tag{7}$$

where θ_d is the desired command input. From Eq. (8), the derivative of z_1 can be described as

$$\dot{z}_1 = \dot{\theta}_d - z_2 + \alpha . \tag{8}$$

In order to select $\alpha_1(z_1)$ to guarantee stability for error dynamics in Eq. (7), the Lyapunov function is defined as

$$V_1 = \frac{1}{2} z_1^2 . \tag{9}$$

Differentiating Eq. (9) and if we select $\alpha = -k_1 z_1 - \dot{\theta}_d$, then

$$\dot{V}_1 = -k_1 z_1^2 - z_1 z_2 \tag{10}$$

where k_1 is a positive constant. Define the second Lyapunov function as follows:

$$V_2 = V_1 + \frac{1}{2} z_2^2 . \tag{11}$$

Differenting Eq. (15), it is given as

$$\dot{V}_2 = -k_1 z_1^2 + z_2[(k_1^2 - 1)z_1 + k_1 z_2 + b_p(-\hat{\sigma}_0 \hat{z} - \hat{\sigma}_1 \dot{\hat{z}} - \hat{\sigma}_2 x_2 - T_d + u) - \ddot{\theta}_d]$$
$$+ b_p z_2[-\sigma_0 \tilde{z} - \hat{z} \tilde{\sigma}_0 - \sigma_1 \dot{\tilde{z}} - \dot{\hat{z}} \tilde{\sigma}_1 - \dot{\theta} \tilde{\sigma}_2] \tag{12}$$

where $\tilde{z} = z - \hat{z}$, $\dot{\tilde{z}} = \dot{z} - \dot{\hat{z}}$, $\tilde{\sigma}_0 = \sigma_0 - \hat{\sigma}_0$, $\tilde{\sigma}_1 = \sigma_1 - \hat{\sigma}_1$, $\tilde{\sigma}_2 = \sigma_2 - \hat{\sigma}_2$ and notation of hat denotes the estimation value. In Eq. (12), the following control input can be chosen as

$$u = J[\ddot{\theta}_d + (1 - k_1^2)z_1 - (k_1 + k_2)z_2] + \hat{\sigma}_0 \hat{z} + \hat{\sigma}_1 \dot{\hat{z}} + \hat{\sigma}_2 \dot{\theta} \tag{13}$$

where k_2 is a positive constant. Since, however, the state variable z cannot be measured directly, we suggest the friction state observer to estimate z as follows:

$$\dot{\hat{z}} = \dot{\theta} - \hat{\sigma}_0 f(\dot{\theta}) \hat{z} . \tag{14}$$

Thus, the estimated friction torque is derived as follows:

$$\hat{T}_f = \hat{\sigma}_0 \hat{z} + \hat{\sigma}_1 \dot{\hat{z}} + \hat{\sigma}_2 \dot{\theta} . \tag{15}$$

Next, the new Lyapunov function is defined as

$$V_3 = V_2 + \frac{1}{2} \tilde{z}^2 + \frac{1}{2\eta_0} \tilde{\sigma}_0^2 + \frac{1}{2\eta_1} \tilde{\sigma}_1^2 + \frac{1}{2\eta_2} \tilde{\sigma}_2^2 . \tag{16}$$

We take the time derivative of Eq. (16) and the following expression is obtained

$$\dot{V}_3 = -k_1 z_1^2 - k_2 z_2^2$$

$$+ \tilde{\sigma}_0 (-b_p z_2 \hat{z} - \frac{1}{\eta_0} \dot{\hat{\sigma}}_0) + \tilde{\sigma}_1 (-b_p z_2 \dot{\hat{z}} - \frac{1}{\eta_1} \dot{\hat{\sigma}}_1) + \tilde{\sigma}_2 (-b_p z_2 \dot{\theta} - \frac{1}{\eta_2} \dot{\hat{\sigma}}_2)$$

$$- \sigma_0 f(\theta,\dot{\theta}) \tilde{z}^2 + b_p z_2 \sigma_0 (-1 + \sigma_1 f(\theta,\dot{\theta})) \tilde{z} - \hat{z} f(\dot{\theta}) \tilde{z} \tilde{\sigma}_0 + b_p z_2 \sigma_1 \hat{z} f(\theta,\dot{\theta}) \tilde{\sigma}_0 \quad (17)$$

In order to estimate the friction parameters of the LuGre model, the following observers are proposed

$$\dot{\hat{\sigma}}_0 = -\eta_0 b_p z_2 \hat{z}, \quad (18)$$

$$\dot{\hat{\sigma}}_1 = -\eta_1 b_p z_2 \dot{\hat{z}}, \quad (19)$$

$$\dot{\hat{\sigma}}_2 = -\eta_2 b_p z_2 \dot{\theta}. \quad (20)$$

Introducing Eq. (18), (19), (20) into Eq. (17), Eq. (17) can be also represented as the compact matrix form as follows:

$$\dot{V}_3 = -\boldsymbol{\Phi}^T \boldsymbol{M} \boldsymbol{\Phi} \leq 0 \quad (21)$$

where $\boldsymbol{\Phi} = [z_1 \quad z_2 \quad \tilde{z} \quad \tilde{\sigma}_0]^T$ and

$$\boldsymbol{M} = \begin{bmatrix} k_1 & 0 & 0 & 0 \\ 0 & k_2 & b_p \sigma_0 (1 - \sigma_1 f(\theta,\dot{\theta})) & -b_p \sigma_1 \hat{z} f(\theta,\dot{\theta}) \\ 0 & 0 & \sigma_0 f(\dot{\theta}) & \hat{z} f(\theta,\dot{\theta}) \\ 0 & 0 & 0 & 0 \end{bmatrix} \geq 0.$$

2.2 Design RFNN Uncertainty Approximator

An uncertainty of friction parameters is considered in order to design the robust controller. The dynamics of Eq. (10) considering the uncertainty can be rewritten as

$$J\ddot{\theta} + \sigma_0 z + \sigma_1 \dot{z} + \sigma_2 \dot{\theta} = J\ddot{\theta} + (\overline{\sigma}_1 + \overline{\sigma}_2)\dot{\theta} + (\overline{\sigma}_0 - \overline{\sigma}_1 \overline{\alpha}(\theta,\dot{\theta})|\dot{\theta}|z) + T_d = u \quad (22)$$

where $T_d(\theta,\dot{\theta},\ddot{\theta},z) = \Delta(\sigma_1 + \sigma_2)\dot{\theta} + \Delta(\sigma_0 - \sigma_1 \alpha(\theta,\dot{\theta})|\dot{\theta}|z)$ is the total lumped uncertainty and upper bar notations represent their nominal values. Now, since the value of T_d cannot be known exactly, the recurrent fuzzy neural network based estimator is adopted to approximate the value of T_d. The signal propagation and the basic function in each layer of the RFNN depicted in Fig. 1 are introduced as follows:

Layer 1- Input layer: The net inputs and the net output are represented as

$$O_i^I(t) = u_i^I w_i^I O_i^I(t-T), \quad i = 1,2 \quad (23)$$

where u_i^I represent the inputs, w_i^I is the recurrent weight and T denotes the time delay.

Layer 2-Membership Layer:

$$net_j^{II}(t) = -[O_i^I(t) + w_{ij}^{II} O_j^{II}(t-T) - m_{ij}]^2 / (\sigma_{ij})^2 \qquad (24)$$

$$O_j^{II} = exp(net_j^{II}), \; j = 1, \dots, n \qquad (25)$$

where m_{ij} and σ_{ij} are, respectively, the mean and the standard deviation of the Gaussian function in the jth term of the ith term input linguistic variable u_{ij}^{II} to the node of the layer II.

Layer 3 – Rule Layer:

$$net_k^{III} = \Pi_j [O_j^{II}(t) + w_{ok}^{IV} O_o^{IV}(t-T)], \quad k = 1, \dots, n \qquad (26)$$

where u_j^{III} represents the jth input to the node of layer III and w_{ok}^{IV}, the recurrent weights.

Layer 4, Output layer:

$$O_o^{IV} = \sum_k w_{ko}^{IV} u_{ko}^{IV} = W_{ko}^T U_{ko} \qquad (27)$$

where the connecting weight w_{ko}^{IV} is the output action strength of the oth output associated with the kth rule, u_{ko}^{IV} represents the kth input to the node of layer IV, O_o^{IV} is the output of the fuzzy neural network.
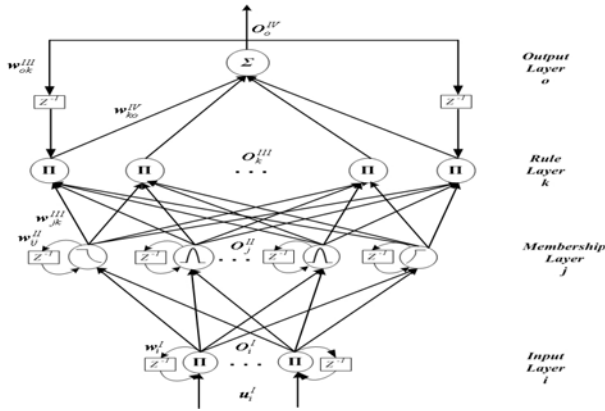

Fig. 1. Structure of the RFNN

The lumped friction uncertainty $T_d(\theta, \dot{\theta}, \ddot{\theta}, z)$ is bounded by

$$|T_d(\theta, \dot{\theta}, \ddot{\theta}, z)| = T_{dm}(\theta, \dot{\theta}, \ddot{\theta}) \qquad (28)$$

where $T_{dm}(\theta, \dot{\theta}, \ddot{\theta})$ is the bounding function of $T_d(\theta, \dot{\theta}, \ddot{\theta}, z)$ and there exists the following function approximation

$$T_{dm}(\theta, \dot{\theta}, \ddot{\theta}) = W_{ko}^{*T} U_{ko} + \varepsilon \qquad (29)$$

where W_{ko}^{*T} is the optimal weighting vector, and $|\varepsilon| \leq E$. Then, consider following the control input

$$u = J[\ddot{\theta}_d + (1 - k_1^2)z_1 - (k_1 + k_2)z_2] + \hat{\sigma}_0 \hat{z} + \hat{\sigma}_1 \dot{\hat{z}} + \hat{\sigma}_2 \dot{\theta} + \hat{T}_{dm} + \hat{T}_{us} \qquad (30)$$

where $\hat{T}_{dm} = O_o^{IV} = W_{ko}^T U_{ko}$ is the RFNN approximation of function bound T_{dm} and $\hat{T}_{us}$ is designed to compensate the difference between the RFNN approximation and the lumped friction uncertainty. Define the following Lyapunov function

$$V_4 = V_3 + \frac{1}{2\eta_w}\tilde{W}_{ko}^T \tilde{W}_{ko} . \qquad (31)$$

where η_w is a positive constant. The time derivative of Eq. (31) is

$$\dot{V}_4 = \dot{V}_3 + \frac{1}{\eta_w}\tilde{W}_{ko}^T \dot{\tilde{W}}_{ko}$$

$$= -\boldsymbol{\Phi}^T \boldsymbol{M} \boldsymbol{\Phi} + \tilde{W}_{ko}^T(-b_p z_2 \boldsymbol{U}_k + \frac{1}{\eta_w}\dot{\tilde{W}}_{ko}) - b_p z_2 \varepsilon + b_p z_2 \hat{T}_{us} \qquad (32)$$

Since W_{ko}^* is constant vector, the following adaptation law and compensation law are chosen as

$$\dot{W}_{ko} = -\dot{\tilde{W}}_{ko} = -\eta_w b_p z_2 \boldsymbol{U}_k \qquad (33)$$

$$\hat{T}_{us} = -\hat{E}\,\mathrm{sgn}(z_2) \qquad (34)$$

where $\hat{E}$ is the estimation bound value of the approximation error. Then, the estimation error is defined as

$$\tilde{E} = E - \hat{E} . \qquad (35)$$

Define the following Lyapunov function

$$V_5 = V_4 + \frac{1}{2\eta_E}\tilde{E}^2 \qquad (36)$$

where η_E is a positive constant. Differentiating Eq. (36) with respect to time and using the previous results, it can be obtained that

$$\dot{V}_5 = -\boldsymbol{\Phi}^T \boldsymbol{M} \boldsymbol{\Phi} - b_p z_2 \varepsilon - b_p \hat{E}|z_2| + \frac{1}{\eta_E}(E - \hat{E})\dot{\tilde{E}} \qquad (37)$$

If the estimation law is chosen as

$$\dot{\hat{E}} = -\dot{\tilde{E}} = -\eta_E b_p |z_2| \qquad (38)$$

then Eq. (37) can be rewritten as

$$\dot{V}_5 = -\boldsymbol{\Phi}^T \boldsymbol{M} \boldsymbol{\Phi} - b_p z_2 \varepsilon - b_p E|z_2| \leq -b_p |z_2||\varepsilon| - b_p E|z_2| = -b_p |z_2|(|\varepsilon| + E) \leq 0 \qquad (39)$$

Since $\dot{V}_4 \leq 0$, the asymptotic stability of $z_1 = 0$, $z_2 = 0$, $\tilde{z} = 0$ and $\tilde{\sigma}_0 = 0$ follows. Also $z_1 \to 0$, $z_2 \to 0$, $\tilde{z} \to 0$ and $\tilde{\sigma}_0 \to 0$ as $t \to \infty$ by Barbalat lemma [7].

Similar online parameter training methodology to train the RFNN via the gradient descent method is well found in the reference literatures [5-6] and we omit it here.

3   Experimental Results and Discussion

Some experiments are also executed to verify the effectiveness of the proposed Back_RFNN control system like simulation. The control algorithms are programmed in 'Turbo-C' language in DOS-mode and the control signals are transmitted into the DC
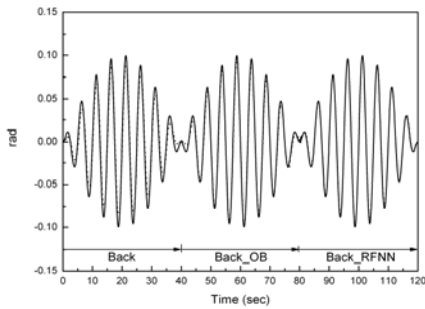
(a) Position tracking results   (b) Position tracking error
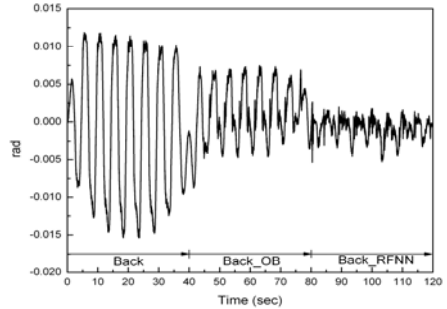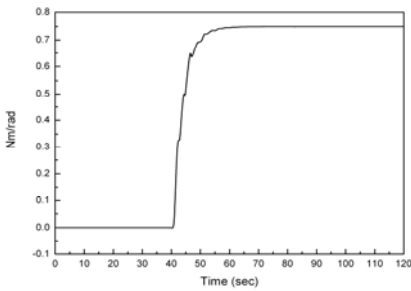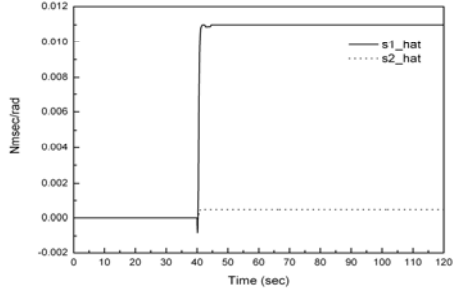
(c) Estimated result of σ_0 (d) Estimated results of σ_1 and σ_2
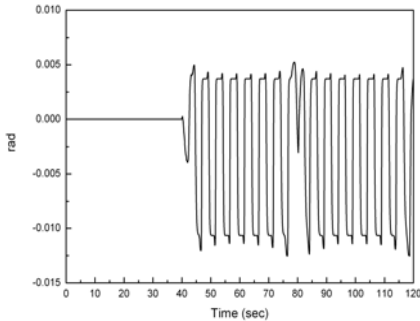
(e) Estimated result of the friction state z (f) Estimated result of the friction torque

Fig. 2. Experimental results for a sine input command with varied amplitude

motor drive through the DR8330 data acquisition board. In the experiment, to investigate the effectiveness the proposed control system, three control schemes are designed; Back, Back_OB, and Back_RFNN system. $\theta_d = 0.1\sin(2\pi \times 0.2t) \times \sin(2\pi \times 0.0125t)$ is chosen as the position command input. The design parameters of backstepping control are set as $k_1 = 150$ and $k_2 = 200$.
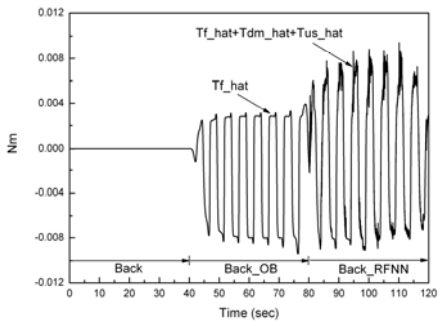
In Fig. 2, the position tracking performance of the Back_RFNN is superior to the other control systems since it has the ability of uncertainty estimation as well as the function of the friction parameter estimation. Thus, it can be known that in the proposed friction compensation scheme, the real friction torque containing the other uncertainty such as the inertia parameter is sufficiently much estimated than the ordinary friction compensation system with the friction observer only. Therefore, it is concluded that the robust compensation scheme must be necessarily considered in order to obtain more precise position tracking performance against the uncertainty

4 Conclusions

In this paper, the robust friction compensation scheme using the friction parameter observer, RFNN and adaptive error compensator is proposed. The nonlinear friction parameters of the servo system are estimated by the proposed friction parameter observer. Next, in order to add the robustness to the friction compensation scheme, the online RFNN and the reconstructed error compensator are also developed to approximate the friction uncertainty. We show that the proposed control scheme guarantees the robustness to the friction uncertainty and the precision position tracking performance in servo mechanical system.

References

1. Canudas de Wit, C., Olsson, H., Astrom, K., Lischinsky, P.: A new model for control of sy stems with friction. IEEE Trans. Automat. Control 40, 419–425 (1995)
2. Al-Bender, F., Lampaert, V., Swever, J.: The generalized Maxwell-slip model: a novel mo del for friction simulation and compensation. IEEE Trans. Automat. Control 50, 1883–188 7 (2005)
3. Leu, Y.G., Lee, T.T., Wang, W.Y.: Online turning of fuzzy neural network for adaptive cont rol of nonlinear dynamical systems. IEEE Trans. Syst., Man, Cybern. 27, 1034–1043 (1997)
4. Lin, F.J., Hwang, W.J., Wai, R.J.: A supervisory fuzzy neural network control system for t racking periodic inputs. IEEE Trans. Fuzzy Syst. 7, 41–52 (1997)
5. Lin, F.J., Wai, R.J.: Robust recurrent fuzzy neural network control for linear synchronous motor drive system. Neurocomputing 50, 365–390 (2003)
6. Lin, C.H.: Adaptive recurrent fuzzy neural network control for synchronous reluctance mot or servo drive. IEE Proc. Electr. Power Appl. 151, 711–724 (2004)
7. Narendra, K.S., Annaswamy, A.M.: Stable adaptive systems. Prentice-Hall, Englewood Cli ffs (1989)

Multi-UCAV Cooperative Path Planning Using Improved Coevolutionary Multi-Ant-Colony Algorithm

Fei Su, Yuan Li, Hui Peng, and Lincheng Shen

College of Mechatronic Engineering and Automation,
National University of Defense Technology, Changsha 410073, China
`minehome@yeah.net`

Abstract. Teams of unmanned combat aerial vehicles (UCAVs) are well suited to perform cooperative mission in hostile environment, and cooperative path planning holds great attention for improving the efficiency of multi-UCAV combating. In this paper, a mathematical formulation for cooperative path planning problem is presented based on the analysis of typical constraints in the scenario. Different from previous studies, the formulation introduces cooperation coefficient to estimate how the UCAV flight paths fulfill the cooperative constraints. Then a coevolutionary multi-ant-colony algorithm is designed and implemented to solve the above-mentioned problem, based on multi-ant-colony algorithm and coevolutionary strategy. The state transition rule and pheromone updating strategy is modified to increase the algorithm performance. Finally, the proposed method is validated to be effective and feasible to solve the cooperative constraints efficiently, and is effective for the multi-UCAV cooperative path planning problem.

Keywords: Unmanned combat aerial vehicle (UCAV), coevolutionary strategy, multi-colony-ant algorithm, cooperative path planning.

1 Introduction

The use of multiple unmanned combat aerial vehicles (UCAVs) to perform cooperative missions such as suppression of enemy air defenses (SEAD) has received a growing attention in the last decade. Multi-UCAV cooperative path planning is the key to take full advantage of multi-UCAV combating cooperatively, which is of significant theoretical value and great practical value. The objective of cooperative path planning is to design optimum flight path for each UCAV, which allows UCAVs to arrive at their targets safely and satisfy the specific cooperative requirements in cooperative mission.

The algorithm for multi-UCAV cooperative path planning is one of the challenging areas in UCAV studies. Motivated by recent advances in heuristic intelligent algorithm, many path planning algorithms have been presented to solve the UCAV path planning problem. In [1, 2], the evolutionary computation based approach is proposed. In [3] sparse A* algorithm is taken to solve TA/TF (Terrain Avoidance/ Terrain Follow) of aerial vehicles. A voronoi graph based approach to obtain flight

path is presented in [4]. Distance transformation algorithm is also a useful method for UCAV path planning [5]. While in these previous studies, the typical cooperative constraints in multi-UCAV cooperative mission are not considered thoroughly, such as the space coordinate constraints, precedence and time window constraints [6,7], so the planned path can not meet the needs of multi-UCAV cooperative missions well.

In this study, we presented a criterion of cooperative path estimation on how the UCAV flight paths satisfy the typical cooperative constraints, based on cooperative coefficient. Then a cooperative path planning algorithm for multiple UCAVs was proposed, which is based on coevolutionary strategy [8] and multi-ant-colony algorithm [9]. Our approach enables multi-UCAV to achieve their targets more effectively without violating constraints.

The remainder of the paper is organized as follows. Section 2 analyzes the typical constraints in multi-UCAV cooperative path planning, and presents a criterion of cooperative path estimation in terms of cooperative coefficient. The proposed coevolutionary multi-ant-colony algorithm is described in detail in section3. Section 4 carries out the simulation and section 5 draws conclusion.

2 Problem Formulation

The multi-UCAV cooperative path planning problem can be described as follows: assume there are N_M enemy threats $\{M_i \mid i = 1, 2, \cdots, N_M\}$ in the $L_W \times L_H$ area E, N_V UCAV$\{V_j \mid j = 1, 2, \cdots, N_V\}$ attack their preassigned targets $\{T_j \mid j = 1, 2, \cdots, N_V\}$, the starting position of each UCAV is $\{S_j \mid j = 1, 2, \cdots, N_V\}$. The objective of multi-UCAV cooperative path planning is that UCAVs must arrive at their target areas without violating the cooperative constraints (in terms of space constraints, task precedence and timing constraints), and the cost for UCAVs to arrive at their targets to perform specific tasks must be minimized. The flight path r_j of the jth UCAV V_j is made up of a series of navigation points $\{np_j(k) = (x_j(k), y_j(k)) \mid k = 0, 1, \cdots, n_j\}$ at time points denoted by $t_k, k = 0, 1, \cdots, n_j$, and $t_{k+1} - t_k = \Delta t, k = 0, 1, \cdots, n_j - 1$, where Δt is the interval between t_k and t_{k+1}.

2.1 Battlefield Entities Model

Since the entities in battlefield are composed of UCAVs and enemy threats, two models are determinately needed for UCAV and threat respectively.

(1) *UCAV Model.* It is assumed that all UCAVs have same flight performance. Let L_{max} be the maximum voyage of UCAV, $\Delta\theta_{max}$ is the maximum turning angle, $[v_{min}, v_{max}]$ denotes its speed range. For UCAV V_j, a kinematic model can be given by the Eq.(1), in which $(x_j(k), y_j(k))$ is the position of V_j at time t_k, $\theta_j(k)$ denotes its heading angle, $\Delta\theta_j(k)$ is the heading angle change, and $v_j(k)$ is the speed of V_j at time t_k.

$$
\begin{aligned}
&x_j(k+1) = x_j(k) + v_j(k) \cdot \Delta t \cdot \cos\theta_j(k+1) \\
&y_j(k+1) = y_j(k) + v_j(k) \cdot \Delta t \cdot \sin\theta_j(k+1) \\
&\theta_j(k+1) = \theta_j(k) + \Delta\theta_j(k) \\
&0 \le \Delta\theta_j(k) \le \Delta\theta_{\max} \\
&v_{\min} \le v_j(k) \le v_{\max}
\end{aligned} \tag{1}
$$

(2) *Threat Model*. We consider that there are two kinds of threats in battlefield, which are anti-aircraft radars and SAMs (Surface to Air Missile). To simplify the representation of the problem, the simplified threats model [10] is used.

If the ith threat M_i is an anti-aircraft radar, its danger level to V_j at time t_k is:

$$
\omega_{i,j}(k) = \begin{cases} 0, d_{i,j}(k) > R_{\max}^{(i)} \cup d_{i,j}(k) < R_{\min}^{(i)} \\ \dfrac{\omega_{\mathrm{m}}^{(j)} \cdot K_{\mathrm{RCS}} \cdot \gamma_v}{(d_{i,j}(k))^4}, R_{\min}^{(i)} \le d_{i,j}(k) \le R_{\max}^{(i)} \end{cases} \tag{2}
$$

If M_i is an SAM, its danger level to V_j at time t_k, can be given as follows:

$$
\omega_{i,j}(k) = \begin{cases} 0, d_{i,j}(k) > R_{\max}^{(i)} \cup d_{i,j}(k) < R_{\min}^{(i)} \\ \dfrac{\omega_{\mathrm{m}}^{(i)} \cdot K_{\mathrm{RCS}} \cdot (R_{\max}^{(i)} - d_{i,j}(k))^\vartheta \cdot \gamma_v}{(R_{\max}^{(i)} - R_{\min}^{(i)})^\vartheta}, R_{\min}^{(i)} \le d_{i,j}(k) \le R_{\max}^{(i)} \end{cases} \tag{3}
$$

where $R_{\min}^{(i)}$ and $R_{\max}^{(i)}$ denote the minimum and maximum detection/attack range of M_i, $d_{i,j}(k)$ is the distance between M_i and V_j, $\omega_{\mathrm{m}}^{(i)}$ is the grade parameter of M_i, and K_{RCS} is a parameter reflecting the RCS characteristic of UCAV. Among them, $\omega_{\mathrm{m}}^{(i)}$ depends on the location, type and capability of M_i, γ_v is the attenuation coefficient of danger level with the increasing of UCAV speed, it can be defined as

$$
\gamma_v = 1 + (v_{\max} - v_{\tan}^{(i,j)}(k)) \cdot (v_{\max} - v_{\min})^{-1} \tag{4}
$$

where $v_{\tan}^{(i,j)}(k)$ denotes the tangential speed of V_j relative to M_i. For the threats set $\{M_i \mid i = 1, 2, \cdots, N_M\}$, its level of danger to UCAV V_j at time t_k, which denoted by $\omega_j(k)$, can be given based on Eqs. (2) ~ (4).

$$
\omega_j(k) = 1 + \sum_{i=1}^{N_M} \omega_{i,j}(k) \tag{5}
$$

2.2   Cooperative Constraints

The typical cooperative constraints of multi-UCAV cooperative path planning problem cover collision avoidance constraints, task precedence constraints, and timing constraints.

For collision avoidance constraint, each UCAV must keep a safe distance d_{safe} with other UCAVs at any time $t_k, k = 0, 1, \cdots$, that is

$$
\| np_i(k) - np_j(k) \| \ge d_{\mathrm{safe}}, i, j = 1, 2, \cdots, N_V, i \ne j \tag{6}
$$

The target precedence constraint states that UCAVs must arrive at their pre-assigned targets in appointed precedence for some special purpose. Let $P(V_j)$ denotes the UCAVs set that must arrive their targets to accomplish tasks before V_j, $N(V_j)$ denotes the UCAVs set that must arrive targets after V_j has accomplished its task, let $AT_j, j = 1, 2, \cdots, N_V$ be the arrival time of each V_j at its target T_j, and the precedence constraints can be formulated as follows:

$$AT_m < AT_j < AT_n, V_m \in P(V_j), V_n \in N(V_j) \tag{7}$$

As for the timing constraints, on one side, a certain UCAV must arrive at its target within a given time window; on the other side, there should always be an expected arrival time interval $\Delta T_{j,m}^*$ between V_j and $V_m \in P(V_j)$, the closer the arrival time of V_j is near to the expected interval, the better the effect of V_j to perform task on its target will be. For each V_j, the time window can be given by following equations:

$$
\begin{aligned}
ET_j &< AT_j < LT_j \\
ET_j &= \max \left(\max_{V_m \in P(V_j)} (AT_m + dt_{\min}^{(j,m)}), \max_{V_n \in P(V_j)} (AT_n - dt_{\max}^{(n,j)}), ET_j^* \right) \\
LT_j &= \min \left(\min_{V_m \in P(V_j)} (AT_m + dt_{\max}^{(j,m)}), \min_{V_n \in P(V_j)} (AT_n - dt_{\min}^{(n,j)}), LT_j^* \right)
\end{aligned}
\tag{8}
$$

where $dt_{\min}^{(j,m)}$ and $dt_{\max}^{(j,m)}$ denote the minimum and maximum arrival time interval between V_j and $V_m \in P(V_j)$, respectively $[ET_j^*, LT_j^*]$ is the fixed arrival time window of V_j.

2.3 UCAV Flight Path Performance Requirements

For UCAV V_j, the integrated cost of its flight path r_j can be formulated as follows:

$$F^{(j)} = \sum_{k=1}^{n_j} f\left(np_j(k), np_j(k-1) \right) \tag{9}$$

$$f(np_j(k), np_j(k-1)) = c_f(v_j(k)) \cdot L(np_j(k), np_j(k-1)) \cdot \omega_j(k) \tag{10}$$

where $f(u_j(k), u_j(k-1))$ denotes the integrated cost of path segment $(np_j(k), np_j(k-1))$, which includes fuel consumption cost and threat cost. In Eq. (10) $c_f(v_j(k)) \propto v_j(k)$ is fuel consumption ratio coefficient with speed, $L(\bullet, \bullet)$ denotes the distance between two navigation points, and $\omega_j(k)$ reflects the threat cost.

In this study, the criterion based on cooperation coefficient was proposed to estimate the cooperation performance of the flight path. The cooperation coefficient describes how the flight paths of UCAVs set fulfill the target precedence and timing constraints. There are two requirements for cooperation coefficient. The first is that the performance of the path r_j is determined by its integrated cost when the arrival time AT_j of V_j satisfies the task precedence and timing constraints; the second is that

the performance of r_j should be punished when AT_j deviates the constraints, and the more AT_j deviates the constraints, the more the punishment will be. Let $\lambda_j(AT_j)$ denotes the cooperation coefficient of r_j, it can be given as follows:

$$\lambda_j(AT_j) = \lambda_j^{(1)}(AT_j) \cdot \prod_{k \neq j} \lambda_{j,k}^{(2)}(AT_j) \tag{11}$$

$$\lambda_j^{(1)}(AT_j) = \begin{cases} \dfrac{|AT_j - ET_j| + |AT_j - LT_j| + \sigma}{|LT_j - ET_j| + \sigma}, & \text{if } ET_j \leq LT_j \text{ and } \lambda_j^{(1)} < \lambda_{\max}^{(1)} \\ \lambda_{\max}^{(1)}, & \text{otherwise} \end{cases} \tag{12}$$

$$\lambda_{j,k}^{(2)}(AT_j) = \begin{cases} 1 + \dfrac{|(AT_j - AT_k) - \Delta T_{j,k}^*|}{\max(|dt_{\min}^{(j,k)} - \Delta T_{j,k}^*|, |dt_{\max}^{(j,k)} - \Delta T_{j,k}^*|)}, & \text{if } V_k \in P(V_j) \\ 1 + \dfrac{|(AT_k - AT_j) - \Delta T_{k,j}^*|}{\max(|dt_{\min}^{(k,j)} - \Delta T_{k,j}^*|, |dt_{\max}^{(k,j)} - \Delta T_{k,j}^*|)}, & \text{if } V_k \in N(V_j) \end{cases} \tag{13}$$

where σ is a small real constant, $\lambda_{\max}^{(1)} > 1$ denotes the maximum value of $\lambda_j^{(1)}(AT_j)$. Based on the cooperation coefficient given by Eq.(11) ~ (13), the cooperative performance criterion of flight path can be formulated as follows:

$$F_C^{(j)} = \begin{cases} F^{(j)}\lambda_j(AT_j), & \text{if } ET_j < LT_j \text{ or } AT_j = ET_j = LT_j \\ F_{\max}\lambda_j(AT_j), & \text{otherwise} \end{cases}, \quad j = 1, 2, \cdots, N_V \tag{14}$$

where $F_{\max}$ is a large constant.

So the cooperative path planning optimization model can be expressed as the minimization problem of Eq. (14) under the constraints (1) and (6) ~ (8).

3   Coevolutionary Multi-Ant-Colony Algorithm Based Cooperative Path Planning

Ant colony algorithm (ACA) [11] has been applied to solve flight path planning problem and obtained a lot of practical applications in recent years [12, 13, 14], but these previous studies mainly focus on single vehicle path planning problem. For multi-UCAV cooperative path planning, satisfaction of cooperative constraints between UCAVs is a key issue. To achieve effective multi-UCAV cooperative path planning, we design the coevolutionary multi-ant-colony algorithm.

3.1   Coevolutionary Multi-Ant-Colony Mechanism

In coevolutionary multi-ant-colony algorithm, ant colony AC is divided into several sub-colonies AC_j, $j = 1, 2, \cdots, N_V$, the jth sub-colony $AC_j = \{Ant_{j,l} \mid l = 1, 2, \cdots, N_S\}$ plans flight path for UCAV V_j, $Ant_{j,l}$ denotes the lth ant individual in AC_j, and N_S is

the ant number in each sub-colony. For AC_j, it has corresponding pheromone matrix $\boldsymbol{\tau}^{(j)} = \{\tau_{x,y}^{(j)} \mid (x,y) \in \boldsymbol{E}\}$ and coordinates with $AC_{i \neq j}$ using coevolution mechanism. Fig.1 shows an example of coevolutionary multi-ant-colony for $N_V = 3$.
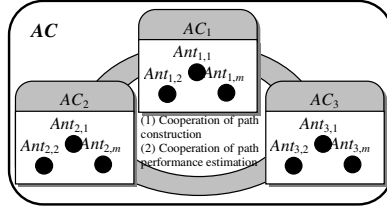

Fig. 1. The coevolutionary multi-ant-colony

In afore-mentioned mechanism, $Ant_{j,l}$ must coordinates with the best flight path obtained by other sub-colonies when it builds flight path for V_j and calculates cooperation coefficient based on Eqs.(11) ~ (13), so as to fulfills the collision avoidance constraints and get the cooperation performance criterion of its flight path.

3.2 State Transition Rule

In this paper, the state transition of ant individuals means that ant individuals move from current position to another under the constraints given by (1) and (6) ~ (8). Each ant generates a complete flight path by choosing the navigation point according to pseudo-random-proportional rule [11]. For AC_j, let $np_j(k)$ be the navigation point of $Ant_{j,l}$ at time t_k, $\boldsymbol{AL}_{j,l}(k)$ denotes the candidate navigation points set of $Ant_{j,l}$ at time t_k which satisfy the constraints given by Eq. (1) and Eq. (6), the state transition rule can be given as follows:

$$np_j(k+1) = \begin{cases} \arg \max_{np \in \boldsymbol{AL}_{j,l}(k)} (\tau_{np}^{(j)})^\alpha (\eta d_{np}^{(j,l)}(k))^\beta (\eta t_{np}^{(j,l)}(k))^\chi, & q \leq q_0 \\ NP, & \text{otherwise} \end{cases} \tag{15}$$

where q is a random number uniformly distributed in $[0,1]$, $q_0 \in [0,1]$ is a constant, and NP is a random variable selected according to the probability distribution given in Eq. (16).

$$p_{np_j(k),np_j(k+1)}^{(i,j)} = \begin{cases} \dfrac{(\tau_{np_j(k+1)}^{(i)})^\alpha \left(\eta d_{np_j(k+1)}^{(i,j,k)}\right)^\beta \left(\eta t_{np_j(k+1)}^{(i,j,k)}\right)^\chi}{\displaystyle\sum_{np \in \boldsymbol{AL}_{i,j}(k)} (\tau_{np}^{(i)})^\alpha \left(\eta d_{np}^{(i,j,k)}\right)^\beta \left(\eta t_{np}^{(i,j,k)}\right)^\chi}, & np_j(k+1) \in \boldsymbol{AL}_{i,j}(k) \\ 0, & \text{otherwise} \end{cases} \tag{16}$$

where α, β and χ are the parameter which determine the relative importance of pheromone versus heuristic info, $\eta d_{np}^{(j,l)}(k)$ and $\eta t_{np}^{(j,l)}(k)$ denote the distance and arrival time heuristic info, and $\tau_{np}^{(j)} \in \boldsymbol{\tau}^{(j)}$ is the pheromone at point $np \in \boldsymbol{E}$.

The distance heuristic info $\eta d_{np}^{(j,l)}(k)$ makes ants move to its target and avoid the area which is very dangerous, it can be defined as

$$\eta d_{np_j(k+1)}^{(j,l,k)} = \frac{\max\limits_{np \in AL_{j,J}(k)} (\widehat{F}_{np}^{(j,l,k)}) - \widehat{F}_{np_j(k+1)}^{(j,l,k)} + \lambda}{\max\limits_{np \in AL_{j,J}(k)} (\widehat{F}_{np}^{(j,l,k)}) - \min\limits_{np \in AL_{j,J}(k)} (\widehat{F}_{np}^{(j,l,k)}) + \lambda} \tag{17}$$

$$\widehat{F}_{np_j(k+1)}^{(j,l,k)} = f(np_j(k), np_j(k+1)) + g^{(j,l)}(np_j(k+1))$$

where λ is a small real constant, $f(np_j(k), np_j(k+1))$ is the integrated cost of path segment $(np_j(k), np_j(k+1))$ which can be calculated by Eq. (10), $g^{(j,l)}(np_j(k+1))$ is the integrated cost pre-estimation of path which have not been built by $Ant_{j,l}$ from $np_j(k+1)$ to T_j. Given the maximum speed $v_{\max}$ and time interval Δt, we can sample line $(np_j(k+1), T_j)$ at intervals of $v_{\max}\Delta t$ and get a series of sample points $\{e_{np_j(k+1)}^{(j,i)} \mid i = 0,1,\cdots,n_{np_j(k+1)}^{(j)}\}$. Based on Eq. (10), $g^{(j,l)}(np_j(k+1))$ can be defined as

$$g^{(j,l)}(np_j(k+1)) = \sum_{i=1}^{n_{np_j(k+1)}^{(j)}} f(e_{np_j(k+1)}^{(j,i-1)}, e_{np_j(k+1)}^{(j,i)}) \tag{18}$$

The arrival time heuristic info $\eta t_{np}^{(j,l)}(k)$ makes ants select the navigation points that fulfill the task precedence and timing constraints, it can be given as follows:

$$\eta t_{np_j(k+1)}^{(j,l,k)} = (\lambda_j^{(1)}(AT_j'))^{-1}$$
$$AT_j' = t_{k+1} + L(np_j(k+1), T_j) \cdot (v_{\max})^{-1} \tag{19}$$

where $\lambda_j^{(1)}(AT_j')$ can be calculated by Eq. (12), and AT_j' is the arrival time of V_j under the condition that it flies straight to T_j at its maximal speed.

3.3 Pheromone Updating Based on Local Diffusion Mechanism

In ant colony algorithm, ant individual easily converges to local best solutions, especially when the search space is vary large. For avoiding local best solution and the slow convergence speed, we introduce the pheromone local diffusion mechanism [15] to improve the original pheromone updating process of ant colony algorithm.

Define local diffusion coefficient as follows:

$$\varsigma(d) = e^{-d^2} \tag{20}$$

Let $N(np)$ denotes the set of neighboring points of np

$$N(np) = \{np' \mid L(np', np) \le ND_{\max}\} \tag{21}$$

$ND_{\max}$ is the maximum range of neighboring area of np. After each iteration of the sub-colony AC_j is performed, the algorithm will carry out the pheromone updating process, which includes local and global updating steps. First, when $Ant_{j,l}$ have built a flight path, if point np have been selected, then the pheromone level of $np' \in N(np)$ is changed in the following way

$$\tau_{np'}^{(j)} = \begin{cases} \tau_{np'}^{(j)} - \rho_{\text{local}}\varsigma(L(np',np))\tau_{np}^{(j)} + \rho_{\text{local}}\varsigma(L(np',np))\Delta\tau_{\text{local}}, \tau_{np'}^{(j)} \geq \tau_{\min} \\ \tau_{\min}, \qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad \text{otherwise} \end{cases} \qquad (22)$$

$$\Delta\tau_{\text{local}} = \begin{cases} (nF_{\text{nn}})^{-1}, \text{if } Ant_{j,l} \text{ built a complete path} \\ 0, \qquad\quad \text{otherwise} \end{cases}$$

where $\rho_{\text{local}} \in (0,1)$ denotes the local evaporation coefficient, n is the navigation point number of the flight path built by $Ant_{j,l}$, and F_{nn} is a constant related with the initial solution. And then, from all feasible flight paths built by sub-colony $\boldsymbol{AC}_j$, the algorithm finds the best path obtained so far by cooperative performance criterion shown in Eq. (14), and performs the global updating for pheromone matrix $\boldsymbol{\tau}^{(j)}$. The global updating rule is

$$\tau_{np'}^{(j)} = \begin{cases} (1-\rho_{\text{global}})\tau_{np'}^{(j)} + \rho_{\text{global}}\varsigma(L(np',np))\Delta\tau_{\text{global}}^{(np',j)}, \tau_{np'}^{(j)} \geq \tau_{\min} \\ \tau_{\min}, \qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad \text{otherwise} \end{cases} \qquad (23)$$

$$\Delta\tau_{\text{global}}^{(np',j)} = \begin{cases} (F_{\text{gb}}^{(j)})^{-1}, \text{if } np' \in N(np) \text{ and } np \in \text{global-best-path} \\ (F_{\text{ib}}^{(j)})^{-1}, \text{if } np' \in N(np) \text{ and } np \in \text{iteration-best-path} \\ 0, \qquad\quad \text{otherwise} \end{cases}$$

where $\rho_{\text{global}} \in (0,1)$ denotes the global evaporation coefficient, $F_{\text{ib}}^{(j)}$ and $F_{\text{ib}}^{(j)}$ are the cooperative performance criterion of global-best and iteration-best path obtained so far by $\boldsymbol{AC}_j$.

3.4   Algorithm Procedure

In multi-ant-colony algorithm for cooperative path planning, each sub-colony $\boldsymbol{AC}_j$ performs the same algorithm procedure. The algorithm procedure is given below:

Algorithm Multi-Ant-Colony Algorithm for Cooperative Path Planning
---
1:    Initialize sub-colony $\boldsymbol{AC}_j$ and pheromone matrix $\boldsymbol{\tau}^{(j)}$
2:    loop
3:        for $l := 1$ to N_S do
4:            loop
5:                Individual $Ant_{j,l}$ selects a navigation point to move according to state transition rules shown in Eqs.(15) ~ (19)
6:            until $Ant_{j,l}$ finishes its flight path
7:            Smooth the flight path built by $Ant_{j,l}$
8:            Compute cooperative performance criterion of the path built by $Ant_{j,l}$
9:            $Ant_{j,l}$ applies a local pheromone updating rule
10:        end for
11:        Select the iteration-best path built by $\boldsymbol{AC}_j$ to compare with the global-best path obtained so far, if the iteration-best is better than the global-best, then replace the global-best with the iteration-best. If there is new global-best path obtained, then send it to $\boldsymbol{AC}_{i\neq j}$
12:    until End_condition

The method of smoothing flight path used in step 7 can be described as follows: for each navigation point in flight path built by $Ant_{j,l}$, adjust it to the line between its previous and next navigation point, if the new point makes the path performance better than the original, then replace the navigation point with the adjusted one.

4   Results and Discussion

The performance of the proposed cooperative path planning algorithm is analyzed in this section using simulation. We consider the following assumptions: there are 10 threats $\{M_i, i = 0, 2, \cdots, 9\}$ generated randomly within a 60km×60km area, the area is divided into 500×500 grids; $M_0 \sim M_3$ are anti-aircraft radars which grade parameters $\omega_m^{(i)} = 1$, $M_4 \sim M_6$ are anti-aircraft radars which grade parameters are 2, $M_7 \sim M_9$ are SAMs which grade parameters are 3, and $R_{\min}^{(i)} = 0, i = 0, 1, \cdots, 9$; 3 UCAVs $\{V_0, V_1, V_2\}$ are used to attack 3 predefined targets $\{T_0, T_1, T_2\}$, the velocity range of UCAV is [120m/s, 200m/s] and $\Delta\theta_{\max} = 30°$; let $\Delta t = 3s$, and $d_{safe} = 3.6km$. The task precedence and timing constraints are shown in Tab.1 .The environment in which UCAVs are performing their tasks is shown in Fig.2 (a).

The parameters used in algorithm are set as follows: $N_S = 10$, $\alpha = \beta = 1$, $\chi = 5$, $ND_{\max} = 5grids$, and maximum interation number is 200. Fig.2 (b) shows the total paths of 3UCAVs in one typical simulation.

Tab.2 shows the arrival times of 3UCAVs, Fig.3 shows the distance between different UCAVs during the flight process. It can be seen that the flight path of each UCAV satisfies typical cooperative constraints well. Fig.4 shows the convergence trace of cooperative performance criterion at each interation, Fig.4 (a) describes the sum of cooperative performance criterion of all UCAVs, and Fig.4 (b) describes the cooperative performance criterion of each UCAV. From the simulation results, we can see that the proposed algorithm can effectively generate cooperative flight paths for multi-UCAV under typical cooperative constraints.
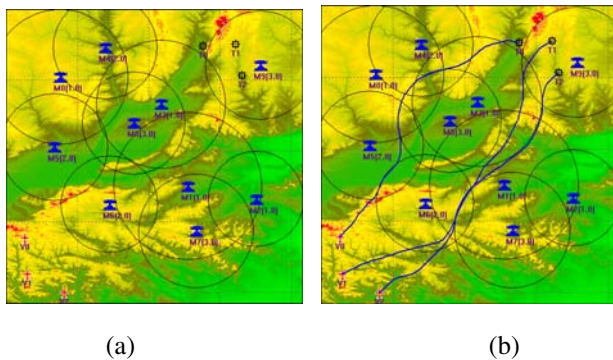

(a)   (b)

Fig. 2. Multi-UCAV Cooperative Path Planning Simulation

Table 1. The precedence and timing constraintsf headings

UCAV	$P(V_i)$	Timing Constraints		
		$[ET_j^*, LT_j^*]$ / $\triangle$t	$[dt_{min}^{(j,m)}, dt_{max}^{(j,m)}]$ / $\triangle$t	$\Delta T_{j,m}^*$ / $\triangle$t
V_0	/	[0, 150]	/	/
V_1	V_0, V_2	[0, 150]	[10,15] after V_0 and V_2	10 after V_0 and V_2
V_2	V_0	[0, 150]	[1,5] after V_0	2 after V_0

Table 2. The arrival times of multi-UCAV

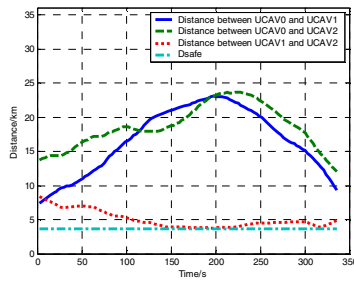| UCAV	Arrival time / $\triangle$t	Satisfy task precedence	Satisfy timing constraints
V_0	105	Yes	Yes
V_1	116	Yes	Yes
V_2	106	Yes	Yes

Fig. 3. Distance between different UCAVs

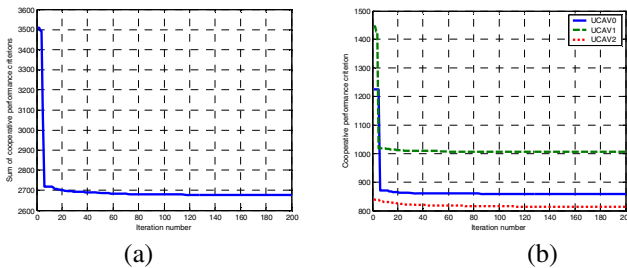(a)   (b)

Fig. 4. Convergence of cooperative performance criterion

To illustrate the efficacy of our algorithm, we compare its performance with standard ACO and genetic algorithm (GA). The population size in GA is 30, and the maximum interation number is 200, 10 rounds of computer simulations are conducted for each algorithm. Fig.5 (a) and Fig.5 (b) show the convergence traces of the sum of

cooperative performance criterion for all UCAVs in one typical simulation, and the statistics of comparison results are shown in Tab.3. From the comparison results, we can find that the proposed algorithm can avoid local best solutions more efficiently and converges to better solutions more quickly.
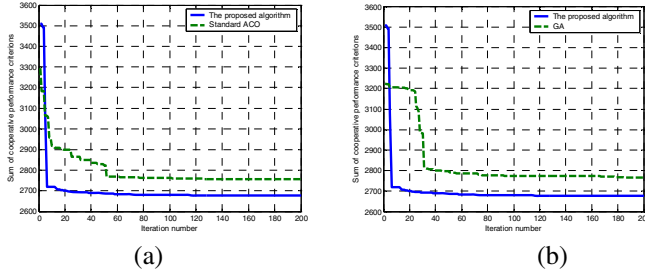

(a)   (b)

Fig. 5. Comparison of the convergence performances for several algorithms

Table 3. The statistics of comparison results

	V_0	V_1	V_2	Sum of criterion	Satisfy constraints
Our algorithm	858.5	1005.7	812.8	2677.0	Yes
Standard ACO	903.5	1033.3	818.3	2755.1	Yes
GA	906.6	1044.8	814.8	2766.2	Yes

5   Conclusion

In this paper, a cooperative performance criterion based on cooperation coefficient is presented to describe how the flight path of each UCAV fulfills the cooperative constraints. Then, an improved coevolutionary multi-ant-colony algorithm is designed and implemented to solve the afore-mentioned problem. The efficiency of the proposed algorithm is verified by simulations. The simulation results demonstrate that this proposed method do well in multi-UCAV cooperative path planning. Further, based on the proposed algorithm, we can implement parallel algorithm for real-time cooperative path planning. This work can make the algorithm more practical for the applications of multi-UCAV and is an important direction for our research.

References

1. Hocaoglu, C., Sanderson, C.: Planning Multiple Paths with Evolutionary Speciation. IEEE Transaction on Evolutionary Computation 5(3), 169–191 (2001)
2. Nikolos, I.K., et al.: Evolutionary Algorithm based Offline/Online Path Planner for UAV Navigation. IEEE Transactions on System. Man, and Cybernetics-Part B: Cybernetics 33(6), 898–912 (2003)
3. Szczerba, R.J., et al.: Robust Algorithm for Real-time Route Planning. IEEE Transactions on Aerospace and Electronic System 36(3), 869–878 (2000)

4. Chandler, P.R., Pachter, M., Rasmussen, S.J.: UAV Cooperative Control. In: Proceedings of the American Control Conference, pp. 50–55 (2001)
5. Horing, J.H., Li, J.T.Y.: Vehicle Path Planning by using Adaptive Constrainted Distance Transformation. Pattern Recognition 35, 1327–1337 (2002)
6. McLain, T.W., Bread, R.W.: Coordination Variables, Coordination Functions, and Cooperative Timing Missions. AIAA Journal of Guidance, Dynamics, and Control 28(1), 150–161 (2005)
7. Shima, T., Rasmussen, S.J., Sparks, A.G., Passino, K.M.: Multiple Task Assignments for Cooperating Uninhabited Aerial Vehicles using Genetic Algorithms. Computers & Operations Research 33(11), 3252–3269 (2006)
8. Potter, M.A., DeJong, K.A.: Cooperative Coevolution: an Architecture for Evolving Coadapted Subcomponents. Evolutionary Computation 8(1), 1–29 (2000)
9. Talbi, E.G., Roux, O., Fonlupt, C.: Parallel Ant Colonies for Combinatorial Optimization Problems. Parallel and Distributed Processing, 239–247 (1999)
10. Bread, R.W., McLain, T.W., Goodrish, M., Anderson, E.P.: Coordinated Target Assignment and Intercept for Unmanned Air Vehicles. IEEE Transactions on Robotics and Automation 18(3), 911–922 (2002)
11. Dorigo, M., Gambardella, L.M.: Ant Colony System: A Cooperative Learning Approach to the Traveling Salesman Problem. IEEE Transactions on Evolutionary Computation 1(1), 53–66 (1997)
12. Hsiao, Y.T., Chuang, C.L., Chien, C.C.: Ant Colony Optimization for Best Path Planning. In: Proceedings of the International Symposium on Communications and Information Technology, pp. 109–113 (2004)
13. Krenzke, T.: Ant Colony Optimization for Agile Motion Planning. M.S. Thesis. Massachusetts Institute of Technology (2006)
14. Chen, M., Wu, Q.X., Jiang, C.S.: A Modified Ant Optimization Algorithm for Path Planning of UCAV. Applied Soft Computing 8, 1712–1718 (2008)
15. Sauter, J.A., et al.: Performance of Digital Pheromones for Swarming Vehicle Control. In: Proceedings of Autonomous Agents and Multi-Agent Systems, pp. 903–910 (2005)

Adaptive Control Using Neural Network for Command Following of Tilt-Rotor Airplane in 0° -Tilt Angle Mode

Jae Hyoung Im[1] and Cheolkeun Ha[2]

[1] Research Assistant, Department of Aerospace Engineering, University of Ulsan, San 29, Mugeo-Dong Nam-Gu Ulsan-Kwangyeoksi, Ulsan, Korea, 682-749
`ttyroon@hotmail.com`
[2] Professor, Department of Aerospace Engineering, University of Ulsan
`cha@mail.ulsan.ac.kr`

Abstract. This paper deals with an autonomous flight algorithm design problem for the tilt-rotor airplane under development by Korea Aerospace Research Institute for simulation study. The objective of this paper is to design a guidance and control algorithm to follow the given command precisely. The approach to this objective is that model-based inversion is applied to the highly nonlinear tilt-rotor dynamics at fixed-wing mode (nacelle angle=0 deg), and then the classical controller is designed to satisfy overall system stabilization and precise command following performance. Especially, model uncertainties due to the tilt-rotor model itself and inversion process are adaptively compensated for in a simple neural network for performance robustness. The designed algorithm is evaluated from the nonlinear airplane simulation in fixed-wing mode to analyze the command following performance for given trajectory. The simulation results show that the command following performance is satisfactory and control responses are within control limits without saturation.

Keywords: KARI tilt-rotor airplane, command following, approximate model-based inversion, adaptive control, neural network.

1 Introduction

A typical tilt-rotor airplane [1] has three flight modes such as rotary-wing mode, fixed-wing mode and transition mode between rotary-wing mode and fixed-wing mode. This type of airplane has some useful characteristics: vertical takeoff and landing of helicopter, and fast speed of airplane with fixed-wing. In general, the tilt-rotor airplane has the dynamic characteristics of highly nonlinear, time-varying system with uncertainties of aerodynamics. In order to cope with the difficulties in controlling the airplane, this paper proposes a feedback linearization-based model reference adaptive control method [2]. This technique generates a linearized model of the nonlinear tilt-rotor airplane through dynamic model inversion. This technique has

some merits: any feedback control techniques available to the linear model are applicable to the technique, the controller resulted from this technique is of simple structure, and the gain-scheduling is not needed any more. However, this technique requires perfect knowledge of the dynamic system of the tilt-rotor airplane, which is impossible in practical sense. Recently, there has been developed an adaptive technique for compensating for model uncertainties originated from the dynamic inversion using neural network. In this approach the neural network-based adaptive control can eliminate the model error online so that the control effectiveness is enhanced and the requirement of exact knowledge of the vehicle dynamics is mitigated [3,4].

In this paper the tilt-rotor airplane that has been developed at Korea Aerospace Research Institute (KARI) for simulation study is utilized to study a flight control design technique of neural network-based adaptive control, where the flight mode of the airplane is at 0°-tile angle mode (Fixed-Wing Mode). Main contribution of this paper is that the proposed model reference adaptive control based on neural network is extended to the guidance control, and is applied to the tilt-rotor airplane developed at KARI. In section 2, dynamic characteristics of the KARI tilt-rotor airplane is explained. The adaptive control technique and its application to the tilt-rotor airplane are discussed in section 3. In section 4, this technique is illustrated in the 6-DOF nonlinear tilt-rotor airplane model simulation to show the effectiveness of performance and robustness of this technique, followed by conclusion in section 5.

2   Dynamic Model of the Tilt-Rotor Airplane

The airplane under development at KARI in Korea is a tilt-rotor airplane similar to Bell 301 in Fig. 1 [1]. This tilt-rotor airplane has 6-DOF dynamics of rigid airplane states and two rotor blades dynamics of each 3 flapping motion. This airplane uses different controls of the longitudinal and lateral motion for the fixed-wing mode and the rotary-wing mode, and one flight control system with a mixer is utilized for overall flight regime. In the fixed-wing mode the flight controls are similar to a typical airplane as shown in Fig. 1. In the rotary-wing mode the flight controls are cyclic pitch and collective pitch for two rotor systems and in the transition mode the two controls using a mixer logic are utilized at the varied nacelle angle.

Fig. 1. A Typical Tilt-Rotor Aircraft and Controls

3   Adaptive Control Based on Dynamic Model Inversion

In this paper autonomous flight controller for the tilt-rotor aircraft in fixed-wing mode is designed. In the design approach the feedback linearization associated with state-input level of the model is combined with the model reference adaptive control technique [5]. Especially, approximate dynamic model inversion is considered in this approach because the tilt-rotor aircraft dynamics are so complicate that the exact model information is usually impossible to be known. In the design process the model error and the model uncertainties are produced from the model inversion so that those models are cancelled by using a simple neural network. The final objective of this control system to be designed is to let the tilt-rotor aircraft in fixed-wing mode follow the guidance command coming out of the waypoint generator shown in Fig. 2. When the guidance command is given to the outer loop logic, the inner loop logic is initiated from the outer loop command in Fig. 2 where inner loop commands are pitch rate command, roll and yaw rate command. The inner loop logic generates the outputs of elevator, aileron and ruder inputs to the tilt-rotor aircraft. Note that the throttle input to the tilt-rotor aircraft is generated in the outer loop logic. Also the inner loop has stability augmentation logic for attitude stabilization of the tilt-rotor aircraft [6].

   In the following sections, the outer loop logic and the inner loop logic are separately designed using the dynamic model inversion, and the compensation of the model uncertainties using neural network introduced in this paper. Note that the inner loop logic and the throttle input from the outer loop logic are based on approximate model inversion because of the lack of the model information.
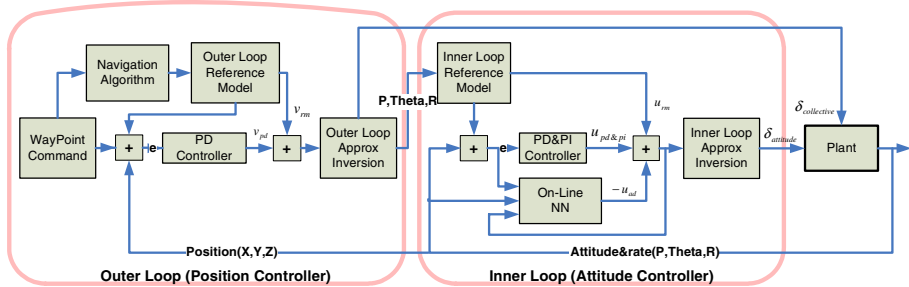

Fig. 2. Autonomous Flight Guidance and Control Algorithm in the Fixed-Wing Mode

3.1   Outer-Loop Control Logic Design

In this section the outer loop logic is designed for following commands available from the navigation logic shown in Fig. 2. The commands from the navigation logic are the positions and the velocity ($\dot{X}, \dot{Y}, \dot{Z}$) in the inertial frame. However the information to be directly used in the outer loop logic is the commands obtained from the coordinate transformation of the inertial frame to the body frame, from which the velocity (V_T), the angle of attack (α), the sideslip angle (β) are obtained. Once the components of the velocity (V_T), the angle of attack (α), the sideslip angle (β) are obtained, the

dynamic model inversion based on approximate information of the plant model is applied to designing the outer loop logic. The pseudo-controls in this logic are those for the velocity, the angle of attack, and the sideslip angle. For the pseudo-control design, the proportional-integral controller is introduced to the outer loop logic as defined in Eq. (1).

$$v_\chi = K_P \left(\chi_c - \chi \right) + K_I \int \left(\chi_c - \chi \right) \text{ for } \chi = V_T, \alpha, \beta . \tag{1}$$

From the force dynamics of the KARI tilt-rotor aircraft the angle of attack related equation is expressed in Eq. (2).

$$\dot{\alpha} = Q - \frac{Q_W}{\cos \beta} - P \cos \alpha \tan \beta - R \sin \alpha \tan \beta , \tag{2}$$

where the pitch rate Q_W is expressed in the wind frame, which is obtained from the relation to the accelerations available from the acceleration sensors in the Z-axis of the body frame. From Eq. (2) the pitch rate equation is obtained, and then applying to the feedback linearization technique realizes the pitch rate command to be required in the outer loop logic, to which the pseudo-control (v_α) is introduced,

$$Q_{des} = v_\alpha + \frac{Q_W}{\cos \beta} - P \cos \alpha \tan \beta + R \sin \alpha \tan \beta . \tag{3}$$

Next, the pitch angle command (θ_{des}) is obtained from the expression of Eq. (4). Combined with the attitude angles relation to the aircraft angular velocities, the pitch angle after integration of the angular rate is obtained as shown in Eq. (4)

$$\dot{\theta}_{des} = Q_{des} \cos \phi - r \sin \phi . \tag{4}$$

Furthermore the roll/yaw rates commands except the pitch angle are required in the inner loop logic. In order to have the yaw rate (R_{des}) of the tilt-rotor aircraft the force dynamics of the aircraft are utilized showing the sideslip angle dynamic equation in Eq. (5)

$$\dot{\beta} = R_W + P \sin \alpha - R \cos \alpha , \tag{5}$$

where the yaw rate R_w is expressed in the wind frame. Introducing the pseudo-control (v_β) to Eq. (5) produces the desired yaw rate command (R_{des}) shown in Eq. (6), where the angle of attack is within appropriate region during the maneuver of the aircraft in the fixed-wing mode.

$$R_{des} = \left(R_W - v_\beta + p \sin \alpha \right) \big/ \cos \alpha . \tag{6}$$

In the sequel, the roll rate command is designed as follows. The tilt-rotor aircraft in the fixed-wing mode is supposed to be in the turn coordination state ($\dot{\phi}_{des}$). The desired roll rate command (P_{des}) is obtained from the attitude angles and the aircraft angular velocity, as shown in Eq. (7)

$$P_{des} = \dot{\phi}_{des} - \dot{\psi}\sin\theta . \tag{7}$$

So far the output signals of pitch command in Eq. (4), roll/yaw rate commands in Eq. (6) and (7) from the outer loop logic are obtained. In order to follow the trajectory given from the navigation logic, the LOS (Line of Sight) guidance is used to track the trajectory, shown in Fig. 3.
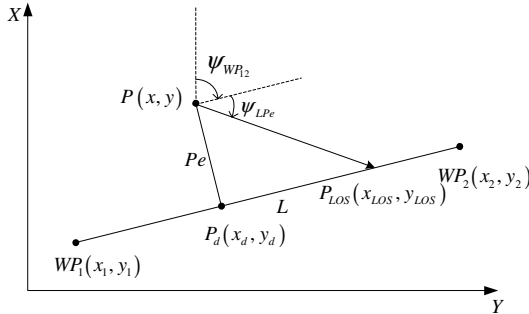

Fig. 3. LOS Guidance Logic

In Fig. 3 it is shown that the straight line connected from the waypoint 1 to the waypoint 2 makes the heading angle ($\psi_{WP_{12}}$) and the correcting heading angle (ψ_{LPe}) is generated to head up the tilt-rotor aircraft toward the desired track where the correcting heading angle (ψ_{LPe}) is defined from Fig. 3

$$\psi_{LPe} = \tan^{-1}\left(-\text{sgn}(P_e)\frac{K_{P_e} \cdot |Pe|}{L}\right), \tag{8}$$

where L denotes the distance from the P_d point to the point across the LOS vector on the track to follow from the waypoint 1 to the waypoint 2, as defined in Eq. (9).

$$L = K_u \cdot u . \tag{9}$$

Also K_u is a design parameter, and Pe is the shortest distance from the aircraft to the point on the track connecting two waypoints given in Fig. 3. Also K_{Pe} is a design parameter. Finally LOS guidance law is obtained from the information of the shortest distance P_e and the distance of L from the point P_d to the point P_{LOS} in Fig. 3, as shown in Eq. (10)

$$\psi_{cmd} = \psi_{WP_{12}} + \psi_{LPe} . \tag{10}$$

3.2 Inner-Loop Control Logic Design

In this section the inner loop logic in Fig. 2 is designed for the tilt-rotor aircraft in the fixed-wing mode. Inner loop logic in the aircraft is related with stabilization of the

attitude motion in longitudinal and lateral dynamics. In this design approach the inner-loop logic and the outer-loop logic are dynamically separated such that the inner-loop logic has the time constant less than the time constant of the outer loop logic. Hence the control bandwidth of the outer-loop logic is lower than that of the inner-loop logic. In the inner loop logic the approximate dynamic model inversion plays a role to stabilize the attitude motion of the tilt-rotor aircraft. The dynamic model inversion is based on the linear model obtained from a trim of the aircraft in the fixed-wing model, where this linear model is expressed in state space given in Eq. (11)

$$\dot{\Omega}_B = \hat{\Gamma}\, X_B + \hat{\Xi}\, \Lambda_\delta ,\tag{11}$$

where $\hat{\Gamma}$ and $\hat{\Xi}$ are depicted to be the matrices for state variables and control inputs with proper dimensions, and the associated variables are defined as follows;

$$\Lambda_\delta = \left\{\Delta\delta_e, \Delta\delta_a, \Delta\delta_p\right\}^T , \quad \Omega_B = \{p, q, r\}^T , \quad X_B = \{\Delta u, \Delta w, q, \Delta v, p, r, \Delta\delta_c\}^T .$$

From Eq. (11) the following control input to the tilt-rotor aircraft ($\hat{\Lambda}_\delta$) is obtained using the approximate dynamic inversion for the inner-loop logic shown in Eq. (12)

$$\hat{\Lambda}_\delta = \hat{\Xi}^{-1}\left\{\dot{\Omega}_B^c - \hat{\Gamma}\, X_B\right\},\tag{12}$$

where $\dot{\Omega}_B^c$ denotes the desired angular velocity. But these variables are not able to be directly obtained from the outer-loop logic. This would be discussed later in this paper. The pseudo-controls (U_j) for the inner-loop logic are defined to be in Eq. (13).

$$U_j = U_{c_j} + \dot{\omega}_{com} - U_{ad_j} , \quad j = p, \theta, r ,\tag{13}$$

where U_j denotes the proportional-derivative (PD) control (U_θ) for the longitudinal axis and the proportional-integral (PI) control (U_p, U_r) for the lateral axis, respectively. The pseudo-control (PI) in the roll axis can be expressed as follows:

$$U_p = K_P \cdot \tilde{P} + K_I \cdot \int_{t_0}^{t} \tilde{P}\, d\tau ,\tag{14}$$

where $\tilde{P} = P_c - P$ denotes the roll angular velocity error between the commanded and the actual roll rate. Also the desired commanded rate in Eq. (13) is expressed in Eq. (15).

$$\dot{\omega}_{com} = \begin{bmatrix} \dot{P}_c & \ddot{\Theta}_c & \dot{R}_c \end{bmatrix}^T .\tag{15}$$

However, the required commands to the inner-loop logic are defined as follows:

$$\dot{\Omega}_B^c = [\dot{p}_D \quad \dot{q}_D \quad \dot{r}_D]^T .\tag{16}$$

The commands defined in Eq. (16) are obtained from the relation between the Euler angle and the angular rate of the tilt-rotor aircraft, as shown in Eq. (17)

$$\dot{p}_D = U_p$$
$$\dot{q}_D = U_\theta / Cos(\Phi) + pqTan(\Phi) + pr + 2qrSin(\Phi)Tan(\Theta)$$
$$\quad + qqSin(\Phi)Tan(\Phi)Tan(\Theta) + rrCos(\Phi)Tan(\Theta)$$
$$\dot{r}_D = U_r$$
(17)

In the above, the model inversion is based on the approximate linear model so that there is the uncertain model error originated from the imperfect model inversion. Due to the uncertain model error in the controller design phase, the model error is going to be compensated for. This model error can be obtained from Eq. (18).

$$\delta = \dot{\Omega}_B - (\hat{\Gamma} X_B + \hat{\Xi} \hat{\Lambda}_\delta) .$$
(18)

For the longitudinal axis of the inner-loop logic in Fig. 2, Eq. (18) is expressed as shown in Eq. (19).

$$\ddot{\Theta} = U_\theta + \delta_2 \cdot Cos(\Phi) - \delta_3 \cdot Sin(\Phi) .$$
(19)

Substituting Eq. (13) to Eq. (19) gives the error equation for the longitudinal axis of the inner-loop logic, shown in Eq. (20).

$$\ddot{\tilde{\theta}} + K_D \cdot \dot{\tilde{\theta}} + K_P \tilde{\theta} = U_{ad_{(\theta)}} - \varepsilon_\theta ,$$
(20)

where the pitch angle error is defined as follows:

$$\tilde{\theta} = \Theta_C - \Theta , \quad \varepsilon_\theta = \delta_2 \cdot Cos (\Phi) - \delta_3 \cdot Sin (\Phi) .$$
(21)

The error dynamic equation is a typical 2-nd order differential equation so that the coefficients of the equation (K_D , K_p) can be determined from the following relation given in Eq. (22).

$$K_D = 2\xi\omega_n , \quad K_P = \omega_n^2 .$$
(22)

In this design process the relation of the coefficients (K_D , K_p) and the damping ratio (ξ) and the natural frequency (ω_n) is set to satisfy the handling quality shown in MIL-SPEC 8785C in which the overshoot and the settling time for unit step response is less than ±10% and 10sec, respectively. In this controller design, the overshoot less than 5% ($\varsigma = 0.8$) and the settling time less than 1.5sec ($\omega_n = 2.5$ rad/sec) are satisfied.

3.3   Compensation for Model Error in Neural Network

The approximate dynamic model inversion requires compensation for the uncertain model resulted from the inversion process. For the purpose of it a neural network is introduced to the inner loop control. The neural network describes the uncertain nonlinear model error as close as possible, and the weighted neural network tries to cancel the model error where the weights are updated online. In this study Sigma Pi-Neural Network (SP-NN) is adopted to compensate for the model error, as shown in Fig. 4, where SP-NN is described in Eq. (23)

$$\hat{u}_{AD(i)}(t) = \sum_{j=1}^{N} \hat{\alpha}_{ij}(t)\varphi'_{ij}(x,\dot{x},u) = \hat{\alpha}'_i(t)\phi'_i(x,\dot{x},u) \quad i = p,\theta,r, \tag{23}$$

where $\hat{\alpha}_i(t)$ is the weights of neural network and $\phi'_i(\cdot)$ is the basis function of the neural network. For instance, the weights of the neural network for the longitudinal motion of the tilt-rotor aircraft are determined to meet Lyapunov stability condition. For Lyapunov stability guaranteed in the longitudinal channel, the error equation in Eq. (20) is expressed as follows:

$$\dot{e}_\theta = A\,e + b\,(U_{AD_{(\theta)}} - \varepsilon_\theta), \tag{24}$$

where the error state vector (e_θ) and the associated matrices are defined to be as follows:

$$e_\theta = \begin{bmatrix} \tilde{\theta} & \dot{\tilde{\theta}} \end{bmatrix}^T, \quad A = \begin{bmatrix} 0 & 1 \\ -K_P & -K_D \end{bmatrix}, \quad b = \begin{bmatrix} 0 \\ 1 \end{bmatrix}. \tag{25}$$

The system matrix of A given in Eq. (25) should be Hurwitz according to Lyapunov stability theorem, and the weights on the neural network in Eq. (23) satisfies the following adaptation rule [4] as shown in Eq. (26)

$$\dot{\hat{\alpha}}_\theta(t) = \begin{cases} -\gamma_\theta e_\theta^T X_\theta b\phi'_\theta(x,\dot{x},u) & for\|e_\theta\|_2 > \varepsilon_o \\ 0 & for\|e_\theta\|_2 \le \varepsilon_o \end{cases}, \tag{26}$$

where the matrix ($X_\theta > 0$) satisfies the following condition

$$X_\theta A + A^T X_\theta = -I_2,$$

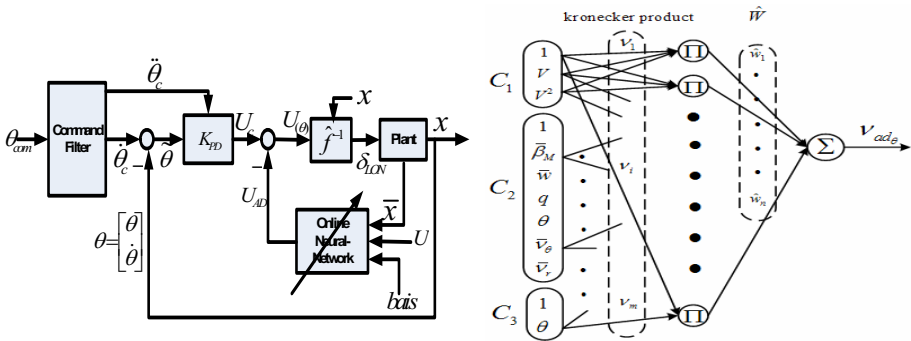and I_2 is a unit matrix. Note that the inner loop logic is depicted in Fig. 4.

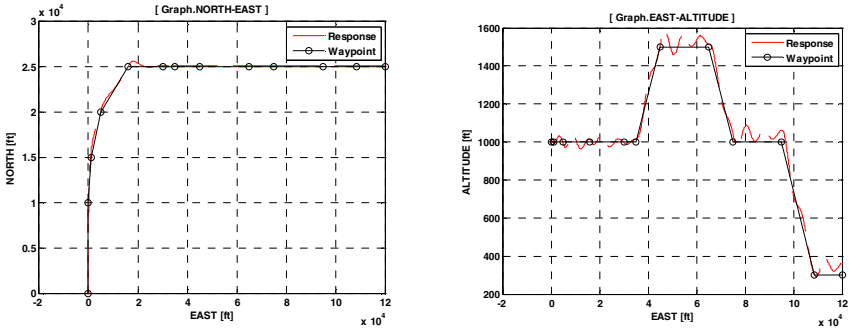**Fig. 4.** Inner Loop Control and SP-NN Structure
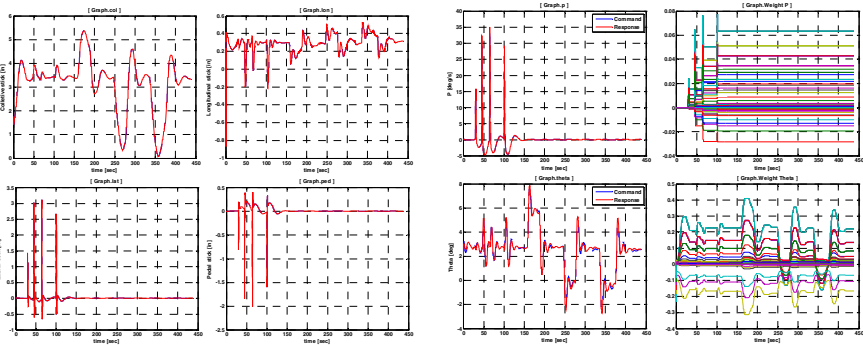
Fig. 5. 2-D Command Following Reponses

Fig. 6. Command Following Responses in Inner-Loop Control

4 Nonlinear Simulation for KARI Tilt-Rotor Aircraft

In this section the numerical simulation using 6-DOF nonlinear tilt-rotor aircraft developed by KARI for simulation purpose is conducted to evaluate the performance of the flight control system designed in section 3. The linear model for the dynamic model inversion is obtained from linearization at a trim of 200 Knots in forward flight, 1000 ft of altitude and 0° of nacelle tilt angle. The waypoints are depicted in Fig. 5 and the trajectory to be followed by the aircraft is generated by connecting the adjacent waypoints in straight line. The trajectory consists of two parts: one part of the trajectory from North to East with 90 deg turn with initial velocity of 200 Knots, and the other for landing from East. It is noted that the proportional gain $K_{P_{\psi}} = 0.5$ is selected for the roll command ϕ_{cmd} and the gains for LOS guidance are determined to be $K_u = 1.0$ and $K_{Pe} = 0.15$. The simulation results are shown in Fig. 5 and Fig. 6, respectively. In Fig. 5 the trajectory following performance in XY-plane of the inertial frame is shown that the following performance is excellent in the aid of LOS guidance in Eq. (8). The altitude response is satisfactory as shown in Fig. 5 even though there is some altitude error because the error is relatively small compared with

the scale of the altitude. It is observed that the altitude error is caused by the pitch command obtained from the transformation of the information to that in the body frame. In Fig. 6 the time responses of the pitch rate, the roll/yaw rate and those of the neural network weights are shown. It is observed that those rate responses are excellent and the weights of the neural network work well in such a way that the neural network for each three axes stop working when the error ($\|e_i\|$) is reached within the limits given in Eq. (26).

5   Conclusion

In this paper model reference adaptive control using neural network is introduced to design a command following controller for the KARI tilt-rotor aircraft. Approximate dynamic model inversion causes the nonlinear model error. The model error is compensated for using SP-NN. According to Lyapunov stability theorem the weights on the neural network are updated online. The command following controller for the aircraft is designed based on the dynamic characteristics of time scale separation into the inner-loop logic (fast time response) and the outer-loop logic (slow time response). To enhance the performance of following commands given from navigation logic, LOS guidance is incorporated into the outer-loop logic. From numerical simulation of the 6-DOF nonlinear aircraft, it is observed that performance of the command following is excellent and the neural network works well.

Acknowledgement

References

1. Harendra, P.B., Joglekar, M.J., Gaffey, T.M., Marr, R.L.: A Mathematical Model for Real Time Flight Simulation of the Bell Model 301 Tilt Rotor Research Aircraft. REPORT 301-099-001 (1973)
2. Baker, R.C., Charlie, B.: Nonlinear Unstable Systems. International Journal of Control 23(4), 123–145 (1989)
3. Troudet, T., Garg, S., Merrill, W.C.: Neural Network Application to Aircraft Control System Design. In: Proc. of AIAA Guidance, Navigation and Control Conference, vol. II, pp. 993–1009 (1991)
4. Kannan, S.K., Johnson, E.N.: Adaptive Trajectory Based Control for Autonomous Helicopters. In: AIAA Digital Avionics Conference, number 358, Irvine, CA (2002)
5. Yoon, H.S., Ha, C.K.: Robust Adaptive Nonlinear Control for Tilt-Rotor UAV. In: 2004 ICCAS International Conference on Control, Automation and Systems, Bangkok, Thailand, pp. 57–62 (2004)
6. Rysdyk, R.T., Calise, A.J.: Adaptive Model Inversion Flight Control for Tiltrotor Aircraft. American Helicopter Society 54th Annual Forum (1998)

INS/GPS Integration System with DCM Based Orientation Measurement

Ho Quoc Phuong Nguyen[1], Hee-Jun Kang[2,*], Young-Soo Suh[2], and Young-Shick Ro[2]

[1] Graduate School of Electrical Engineering, University of Ulsan, 680-749, Ulsan, South Korea
qtquocphuong@mail.ulsan.ac.kr
[2] School of Electrical Engineering, University of Ulsan, 680-749, Ulsan, South Korea
{hjkang,yssuh,ysro}@ulsan.ac.kr

Abstract. This paper works toward the development and implementation of a INS/GPS integration system for the land vehicle application. A developed INS system is introduced to keep measuring the position/orientation of the vehicle when the vehicle is passed through GPS signal shading area. A new orientation scheme is studied to full fill the measurement states of the integration system. Roll/pitch estimation compensating external acceleration is performed with inertial sensors and yaw angle is obtained with GPS information. And then, the orientation information is supplied to the linearized Kalman filter of error model. This process is shown to improve the performance of the integration system. The field test was performed along a non-flat contour with some dismissals of GPS on it.

Keywords: GPS/INS Integration, Orientation Measurement, Error Model.

1 Introduction

Recently, intelligent vehicle systems have been developed quickly in which information about vehicle dynamic states including position, velocity and orientation become necessary. INS/GPS integration system plays the most important role in supplying that information to other tasks such as navigation in urban environment, in tunnel, vehicle stability control, antilock brake system, side slip control and so on [8] [9]. For this purpose, low cost strapdown Inertial Navigation System (INS) aiding with Global Positioning System (GPS) is used.

The low-cost strapdown INS tends to have large accumulated error but acceptable in short time. Global Positioning System tends to have some sudden large time of outage but have no accumulated error. The integration system preserves the good point of long-time robust from GPS and short-time accuracy and self localization from the INS. Because of this complement property, the integration of INS/GPS is

---

widely used in practical [1]. On the same root, many works have been done to solve some aspects of this integration system such as adding vision and radar [3], considering the lever arm case between the GPS and inertial measurement unit [4], calibrating and initializing [5].

Among many integration methods with different complexity [6],[7], the dynamics error model which is derived on the local geographic frame is most applied [1]. In the system, INS mechanization help propagate from input sensors data to have the position, velocity and orientation as outputs. GPS supplies position and velocity to the linearized Kalman filter of error model as measurement. Here, measurement includes six information, three position components and three velocity components. The three remain information are orientation components include roll, pitch and yaw angle. Some other researchers try to use external sensors for this adding purpose such as inclinometers or GPS based attitude determination [11]. In this paper, a new stand-alone roll-pitch estimation scheme was implemented which uses the INS data itself to generate attitude for the error model's measurement. And the remaining yaw angle is obtained by switching between a magnetic compass outputs and GPS-based heading angle. With these three new ones, we have the full measurement for INS/GPS integration error model include nine components. With full measurement, the system can be preserved from accumulated errors and so it can run in a long time without restarting.

In order to estimate the attitude using inertial sensors, the difficulties to overcome is how to compensate the external acceleration which is mentioned in [2] and which method is used to avoid the singularity. In this paper, the DCM based attitude estimation compensating external acceleration is first derived. The estimator uses whole inertial sensors as input and adjusts the belief degree between gyrometers and accelerometers to generate the answer angles. The method is proved to have no singularity. The output could be used independently even though there was no other navigation solution available.

2   Nomenclatures

Mechanical relationships in this paper are written on four coordinate systems: inertial frame *"i"*, Earth Central Earth Fixed *"e"* , navigation frame *"n"* and body fixed frame *"b"*. And here are some notations:

DCM, $_b^n C$: Direction Cosine Matrix(from frame *n* to *b*)

$^n P$: Position in frame *n*

$^n V$: Velocity in frame *n*

$\delta ^n P, \delta ^n V$: Position/velocity error in frame *n*

$\delta ^n \psi$: Misalignment angle between the estimated attitude and the true one in frame *n*

$^b \omega_{ib}$: Angular rate of the frame *b* with respect to the frame *i*, considered in frame *b*

$\left[^b \omega_{ib} \times \right]$: Skew matrix made by the three component of $^b \omega_{ib}$

$^n f$: Acceleration consider in frame *n*

3   INS/GPS Integration

The INS/GPS integration algorithm use INS mechanization as the main navigated information source. One Kalman filter-based dynamics error model is using to estimate the error may occurs inside the INS mechanization compared with GPS information. Estimated error is used to fix the INS mechanization information. The navigation solution is solved in navigation frame.

3.1   INS Mechanization and Its Error Model

The INS mechanization takes inertial sensors data, propagates to get the position, velocity and also orientation. The classification of this module is low cost, strapdown INS - in contrast with gimbal INS. The main difference between gimbal and strapdown INS is at the orientation solution. In the strapdown INS orientation solution is computed from the angular rates measured by the three gyroscopes and the initial orientation to reach the desired orientation overtime. There are a number of algorithms available for orientation propagation namely, Euler, Direction Cosine and Quaternion methods. The Direction Cosine Method (DCM) is commonly used in all forms of inertial navigation. It is also offered the best representation for land navigation application.

In this research, INS mechanization is computed in navigation-frame. Inertial navigation equations written in navigation-frame which were mentioned in [6] could be addressed as Eqs. (1-3) below

$$^n\dot{P} = {}^nV \tag{1}$$

$$^n\dot{V} = {}^n_bC\,{}^bf - 2\,{}^n\omega_{in} \times {}^nV + \left({}^ng({}^nP) - {}^n\omega_{in} \times [{}^n\omega_{in} \times {}^nP]\right) \tag{2}$$

$$^n_b\dot{C} = {}^n_bC\,[{}^b\omega_{nb}\times] \tag{3}$$

Where, $[2\,{}^n\omega_{in} \times {}^nV]$ is Coriolis part. ${}^n\omega_{in} \times [{}^n\omega_{in} \times {}^nP]$ is Centrifugal part, when Earth rotation is considered. And for convenience, ${}^b\omega_{nb} = {}^b\omega_{ni} + {}^b\omega_{ib} \simeq {}^b\omega_{ib}$ because the relative rotation velocity of the navigation frame with respect to inertial frame, ${}^b\omega_{ni}$, is much smaller compared with ${}^b\omega_{ib}$ (gyrometer output).

Error model of given INS mechanization was built to evaluate the effect of error sources on sensors and on computed implementation to the inertial navigation system. This part presents the development of the error equations that is accomplished by perturbing or linearizing the nominal equation.

The orientation error ($\delta^n\psi$) is defined as the misalignment angle between the estimated orientation and the true orientation. The errors in the velocity (δ^nV) and position (δ^nP) are defined as the estimated quantity minus the true quantity. Follow [1], we have process error equations as following

$$\begin{bmatrix} \delta^n \dot{P} \\ \delta^n \dot{V} \\ \delta^n \dot{\psi} \end{bmatrix} = \begin{bmatrix} \delta^n V \\ [^n f \times] \delta^n \psi + {}^n_b C \delta^b f \\ -{}^n_b C \delta^b \omega_{ib} \end{bmatrix} \tag{4}$$

In which, variables error model is

$$\delta X = [\delta^n P, \ \delta^n V, \ \ \delta^n \psi]^T \tag{5}$$

From relationship in Eq.(4), process model is established as Eq.(6)

$$\delta \dot{X} = F_t \delta X + G_t w_t \tag{6}$$

$$F_t = \begin{bmatrix} 0 & I_{3\times3} & 0 \\ 0 & 0 & [\,^n f \times] \\ 0 & 0 & 0 \end{bmatrix}, \ G_t = \begin{bmatrix} 0 & 0 & 0 \\ 0 & {}^n_b C & 0 \\ 0 & 0 & -{}^n_b C \end{bmatrix} \tag{7}$$

Where, F_t, G_t are varied with time and must be linearized in each updated circle.

Process noises involve the error of accelerometer and gyrometer. It is assumed to be Gaussian white noise.

$$w_t = [0_{3\times3} \ \ \delta^b f \ \ \delta^b \omega_{ib}]^T \tag{8}$$

Error model in Eq.(6) becomes process model of a linearized Kalman filter. Error of INS mechanization is estimated by Kalman filter, the detailed algorithm and data flows are discussed in part 3.4 of this paper. The filter takes information from INS mechanization module, GPS, DCM and also raw gyrometers and accelerometers data to generate the estimated error as output.

3.2 Roll/Pitch Estimation

In this part, a new attitude determination algorithm is presented, in which external acceleration is considered and compensated. This module help generate two comparative inputs for error model in the previous part.

DCM is written in term of rotation matrix ($^n_b C$) that describes the orientation of coordinates frames *"b"* with respect to navigation frame *"n"*. Rotation order is about *zz'*, *yy'* and then *xx'* corresponding to Euler angles: yaw (ψ), pitch (θ), roll (ϕ). Rotation matrix $^n_b C$ can be expressed as Eq. (9) which was described in [5]:

$$^n_b C = \begin{bmatrix} \theta_c \psi_c & -\phi_c \psi_s + \phi_s \theta_s \psi_c & \phi_s \psi_s + \phi_c \theta_s \psi_c \\ \theta_c \psi_s & \phi_c \psi_c + \phi_s \theta_s \psi_s & -\phi_s \psi_c + \phi_c \theta_s \psi_s \\ -\theta_s & \phi_s \theta_c & \phi_c \theta_c \end{bmatrix} \tag{9}$$

The notation *"s"* refers to sine and *"c"* refers to cosine. The transformation matrix $^n_b C$ can be obtained with the following integration [5].

Normally, whole DCM is updated. However, we can update each line of DCM matrix separately as below:

$$
\begin{bmatrix} {}^n_b \dot{C}_{31} \\ {}^n_b \dot{C}_{32} \\ {}^n_b \dot{C}_{33} \end{bmatrix} = \begin{bmatrix} 0 & -{}^n_b C_{33} & {}^n_b C_{32} \\ {}^n_b C_{33} & 0 & -{}^n_b C_{31} \\ -{}^n_b C_{32} & {}^n_b C_{31} & 0 \end{bmatrix} \begin{bmatrix} {}^b \omega_x \\ {}^b \omega_y \\ {}^b \omega_z \end{bmatrix}
\tag{10}
$$

Where, ${}^n_b C_{3x}$ is the component of DCM at position (3,x) and ${}^b \omega_x, {}^b \omega_y, {}^b \omega_z$ are angular velocity state with respect to the body-fixed frame which measured by gyrometers. So, even though using the DCM method, only three states must be updated in each circle.

Measurements model:

$$
\begin{bmatrix} {}^b f_x \\ {}^b f_y \\ {}^b f_z \end{bmatrix} = {}^n_b \mathbf{C}^T \begin{bmatrix} 0 \\ 0 \\ g \end{bmatrix} = \begin{bmatrix} {}^n_b C_{31} \\ {}^n_b C_{32} \\ {}^n_b C_{33} \end{bmatrix} g
\tag{11}
$$

Although relationships in Eq.(10) means we integrate by time the angular velocity to have the true angle, however, the angles can be directly produced from accelerometer data. We use these accelerometer data as the measurements for the filter as Eq.(11). Output of the filter adapts from these two sources. To adjust the believable degree in which source, we change the measurement covariance values in Kalman Filter. The mechanical relationship is applied to the Kalman Filter as following form

$$
\dot{x}(t) = \Phi_t x(t) + w(t)
\tag{12}
$$

$$
z(t) = \Gamma x(t) + v(t)
\tag{13}
$$

Where Φ_t are varied with time and must be linearized in each updated circle as

$$
\Phi_t = \begin{bmatrix} 0_{3\times3} & C_3 \\ 0_{3\times3} & 0_{3\times3} \end{bmatrix}, C_3 = \begin{bmatrix} 0 & -{}^n_b C_{33} & {}^n_b C_{32} \\ {}^n_b C_{33} & 0 & -{}^n_b C_{31} \\ -{}^n_b C_{32} & {}^n_b C_{31} & 0 \end{bmatrix}
\tag{14}
$$

$$
\Gamma = \begin{bmatrix} 0_{6\times6} & I_6 \end{bmatrix}
$$

Notice that the state-variables include three components of DCM and three angular velocity components as Eq. (15). While, the measurement include totally six sensors output as Eq. (16)

$$
\mathbf{x}(t) = \begin{bmatrix} {}^n_b C_{31} & {}^n_b C_{32} & {}^n_b C_{33} & {}^b \omega_x & {}^b \omega_y & {}^b \omega_z \end{bmatrix}
\tag{15}
$$

$$\mathbf{z}(t) = \begin{bmatrix} {}^b f_x & {}^b f_y & {}^b f_z & {}^b \omega_x & {}^b \omega_y & {}^b \omega_z \end{bmatrix} \tag{16}$$

For simplicity reason, the notations in Eq. (16) imply the following original denotes ${}^b f_x \equiv {}^b f_{ibx}$ and ${}^b \omega_x \equiv {}^b \omega_{ibx}$ which are output signal from the accelerometers and gyrometers, respectively.

$$Q_{att} = E\{w(t)\, w^T(t)\} = \begin{bmatrix} 0_{9\times9} & 0_{9\times3} \\ 0_{3\times3} & q.I_{3\times3} \end{bmatrix} \tag{17}$$

$$R_{att} = E\{v(t)\, v^T(t)\} = \begin{bmatrix} r_a\, I_{3\times3} & 0_{3\times3} \\ 0_{3\times3} & r_g\, I_{3\times3} \end{bmatrix} \tag{18}$$

In Eq. (18) accelerometer measurement noise covariance values can be adjusted to compensate the effect of external acceleration. q-value is also chosen to maintain the smooth of output response

$$r_a = 5E - 4 \rightarrow 2 \,, r_g = 6.E - 3 \,, q = 2E - 4 \tag{19}$$

These adjustment based on the following criteria

$$f(a_x, a_y, a_z) = \left| a_x^2 + a_y^2 + a_z^2 - 1 \right| > \sigma \tag{20}$$

After DCM is updated, normalization is necessary for preventing the unit property. When the three components of DCM are estimated, attitude could be computed easily [7].

3.3 Heading Evaluation

Orientation consists of three Euler angles, two attitude angles which are estimated in previous section and the remaining one is heading angle. In order to evaluate heading, two methods are used. First, magnetic compass is used when the vehicle is in initial state or static state because of the slowly response of magnetic compass on angular changing and the magnetic disturbance of vehicle body. With the knowledge of attitude angles, the measured magnetic field vector can be projected on a horizontal plane parallel to the Earth's surface follows [10]

Second, GPS based heading is used when the vehicle moving upward with certain velocity:

$$\psi = atan\left(\frac{{}^n P_x \,|_k - {}^n P_x \,|_{k-1}}{{}^n P_y \,|_k - {}^n P_y \,|_{k-1}} \right) \tag{21}$$

3.4 Integration Method

For illustrating the algorithm, Fig. 1 shows the data flows from different sensors sources and the way to integrate them into a system.
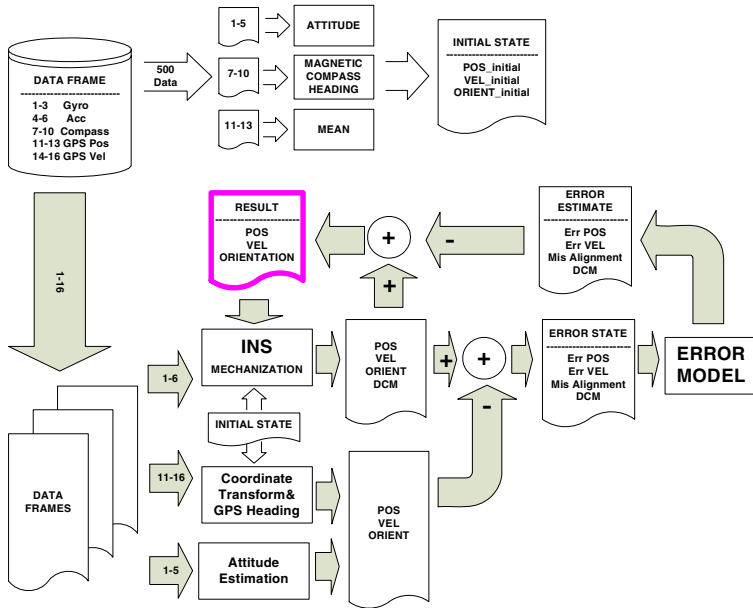
Fig. 1. Data flows for whole algorithm

Fig. 1 shows the direct feedback diagram in which result position, velocity and orientation. Characteristics of this direct feedback system could be addressed as below

1.   The frequency of INS is 100Hz, when the frequency of GPS is normally 1Hz so, in fact, the purpose is using INS mechanization to calculate the navigation information between the time of two GPS data. Error model is linear Kalman filter work on error states between inertial navigating states and GPS adding measurements during that time. Whenever a new GPS data is received, the error states are added to the propagation values from the INS mechanization to have the fixed states.
2.   INS is a high dynamic system, the inaccuracy in sensors data, such as bias, and intermediate information, such as translation velocity, causes large error in output result. So, the fixed states must be fed back to the INS mechanization to preserve the unbounded growth in error observations.

As input of integration filter, the measurements from the GPS receiver, attitude estimation and GPS heading are subtracted with INS mechanization output to generate error in position, velocity, attitude and heading. Whenever a measurement source become available, an error model measurement is obtained and the filter then updates the estimate of the error state in the inertial unit. Derivation of the filter process model is prescribed in section 3.1. Corresponding to process model in section 3.1, the measurement model is used as

$$z(k) = H(k)x(k) = I_{9\times9}\, x(k) \tag{22}$$

Using the measurements from GPS, attitude and heading which as follow

$$z(k) = \begin{bmatrix} P_{inertial}\,(k) - P_{GPS}\,(k) \\ V_{inertial}\,(k) - V_{GPS}\,(k) \\ \begin{bmatrix} \phi_{inertial}\,(k) \\ \theta_{inertial}\,(k) \end{bmatrix} - \begin{bmatrix} \phi_{Att\,Est}\,(k) \\ \theta_{Att\,Est}\,(k) \end{bmatrix} \\ \psi_{inertial}\,(k) - \psi_{GPS\,Cal.}\,(k) \end{bmatrix} \tag{23}$$

Filter is adjust based on the measurement noise covariance matrix as follow

$$R_{GPS}\,(k) \;= diag(\sigma^2_{PN}, \sigma^2_{PE}, \sigma^2_{PD}, \sigma^2_{VN}, \sigma^2_{VE}, \sigma^2_{VD}) \tag{24}$$

$$R_{ORNT}\,(k) = diag(\sigma^2_{ATT}, \sigma^2_{ATT}, \sigma^2_{HEAD}) \tag{25}$$

$$R(k) \;= \begin{bmatrix} R_{GPS}\,(k) & 0_{6\times3} \\ 0_{3\times6} & R_{ORNT}\,(k) \end{bmatrix} \tag{26}$$

Where, the individual variances of R_{GPS} are obtained from the GDOP values while σ^2_{ATT} and σ^2_{HEAD} are chosen experimentally up to the parameters in the attitude estimation and GPS position perturbation. And *diag()* is the diagonal matrix with given components in order.

4   Implementation

This work involves development of an INS/GPS integration system for the land vehicle application. Install the measurement system to a car ran around Mun-su Stadium, the contour covers about 3km un-flat road. The road map is showed in Fig. 5.

4.1   System Description

A low cost, strapdown IMU was made with some parameters in the Table 1. It is packaged as a compact, aluminium box and mounted in the car trunk (Fig. 2).
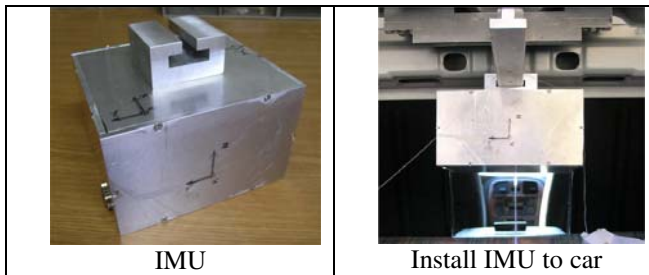

| IMU | Install IMU to car |

Fig. 2. Inertial Measurement Unit and its installation in car

Commercial accuracy GPS is used with antenna is put outside so that it is near the IMU. Data from IMU and GPS is collected parallel to process offline. See Table 1 for more details.

Table 1. Hardware Details

*Inertial Measurement Unit – (*Our product*)*	
Common	*100 Hz, 10 bit ADC, RS232 interface, baudrate 112500*
Accelerometer ADXL202B	*Range +/- 2g, dual axis* *Low cost (US30$)*
Gyrometer ADXRS150	*+/-150 degree/s Single Chip Yaw* *Low cost (US70$)*
Mag. Compass HMC2003	*3-axis magnetometer, 40 micro-gauss to ±2 gauss at 1V/gauss*
	GPS Module
GPS Novatel, Propak-G2	*Accuracy in single mode: 2m. Control mode: "bestxyz". Extracts data in format: WGS84 ECEF*

In the experiments, the vehicle must be kept stationary for 10 seconds for collecting the bias and initializing. That mean the first 500 data used to set the bias and then, next 500 data used to calculate the initial states. After that, the information is divided into two main sources, one to the INS mechanization to evaluate the main navigation information, the other to the measurement estimated part to produce the adding data. Subtract the main navigation with the adding navigation information; we have the inputs for the error model. Results are the main navigation states from INS mechanization subtracted with the errors estimated. And also, the results are fed back to INS mechanization to adjust its state.

4.2   Results

In this part, results of orientation estimation and INS/GPS integration system are shown. Comparisons focus on the improving in performance when number of measurement increases from six [1] up to nine components.

4.2.1   Orientation Estimation Parameters Turning

Fig. 3 shows the experiment result was made indoor to turn the attitude estimation Kalman filter. Without external acceleration adaptively compensating, the error between the estimated roll angle and the encoder reference angle is about 14 deg. It decreases to 6 deg when using compensation technique.

Fig. 4 shows the measurements from the attitude estimation and the GPS based heading in outdoor experiment. Property of the orientation solution without adding is that the bias errors and the calculated errors through time can not be fixed. So that, as time gone by, the accuracy of integration system decreases.

The property of the attitude estimation using accelerometers and gyrometers combination is that there still be some effects of external acceleration can not be compensated completely. However the dot lines show that, they contain no error
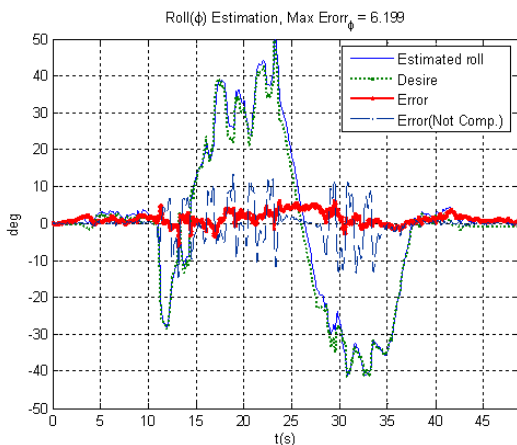
Fig. 3. Roll angle estimation with/without external acceleration adaptively compensating

accumulated. That mean as time gone by, the system could be prevented from orientation accumulated errors and then position accumulated errors.

The property of heading base on GPS position is that whenever the vehicle runs forward, this algorithm can show quite accuracy heading (the dot line reaches 360 deg when the vehicle covers one circle). Problem is the resolution of GPS is around meter, the same lever of vehicle movement in one seconds. Hence, there are some points that the measurement is wrong. Those are some drops in the third plot of Fig. 4. So, turning purpose is determining some criteria to decide which the accuracy measurements are and which wrong ones are. Name a few, limited max-min variation, limited gravity value and limited variant gravity value are applied.
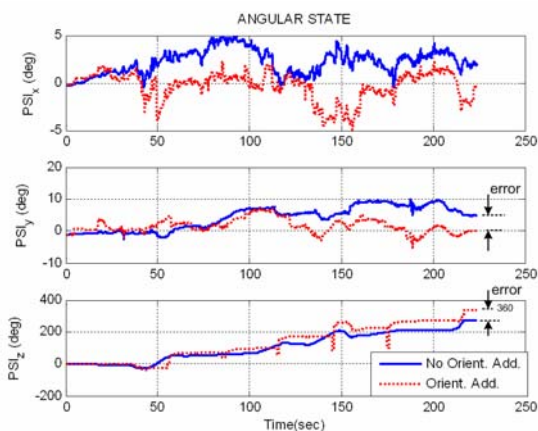

Fig. 4. Attitude and heading measurements

4.2.2   Error Model Noise Covariance Turning
In this test, GPS value is regularly enabled 20 seconds, disabled 10 seconds and so on. The purpose is testing the ability of the integration system in keeping trajectory without

GPS signal. In order to determine the parameters for error model, consider the error variance in two cases: first is error variance values in case every second GPS updated and second is error variance values in case 10 second GPS updated. From the analyzing of error variance values, we have some ideas that made the system work properly:

- First, measurement noise covariance values of error model can be chosen as the covariance values when GPS is updated every second
- Second, position and velocity process noise covariance values must be bigger than corresponding measurement noise covariance values several times to keep the system updates GPS position fast enough and the system can suffer the big error of 10s outage. The orientation noise covariance is smaller than corresponding measurement noise covariance as orientation information cause sensitive effect on output response, we should believe more in model to keep the system stable. The covariance values of error model are chosen as

Table 2. Error model noise covariance values

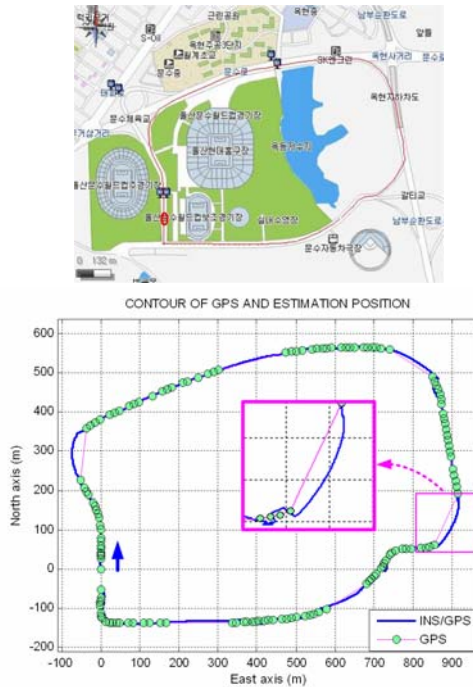| r/q	dPx	dPy	dPz	dVx	dVy	dVz	dPsix	dPsiy	dPsiz
r	13	24	3	0.7	0.5	0.2	1E-3	1.5E-3	0.01
q	100	100	15	3.5	2	0.1	5E-4	7E-4	0.005

Fig. 5. The contour in North-East plane.Full measurement error model. The car ran around Mun-su stadium with the velocity of 40-50 km/h. In comparison, the real map is put in upper corner.

Fig. 5 shows that even the outages happen at the corners in the contour, position plot on North-East plane is still smooth. In case of 10 second outage, the detailed plot of position error and orientation error are showed in Fig. 6 and Fig. 7.
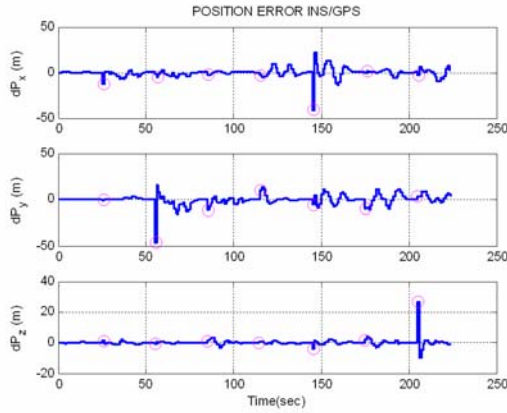

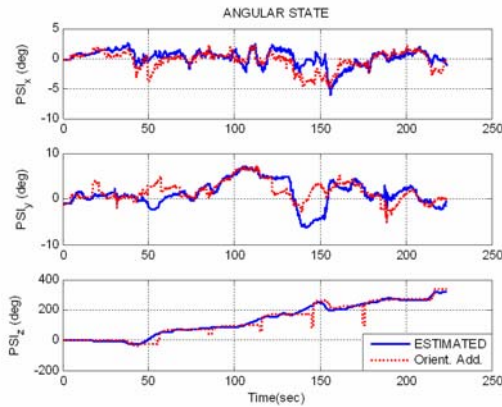**Fig. 6.** Position error in case of full measurement error model, 10 second GPS outage

Fig. 7. Estimated Attitude and Heading compared with the corresponding measurements

4.2.3 Full Measurements Error Model in Comparison with Six Measurements One

In this part, efficiency of applying orientation measurement to error model is evaluated. Compared model which has no orientation updated is used.

Analytically, let consider the errors of the two cases in Fig. 6 and Fig. 8. Errors are calculated as the subtraction between the INS position components with the correspondent GPS position components if GPS signal is enable. In case of GPS signal is not available, set the error to be zero until the next GPS signal.

Fig. 6 shows the error of full measurement (9 measurements). The maximum error is 46m after 10 second outage. The average error is much smaller, around 9m for 10 second outage.
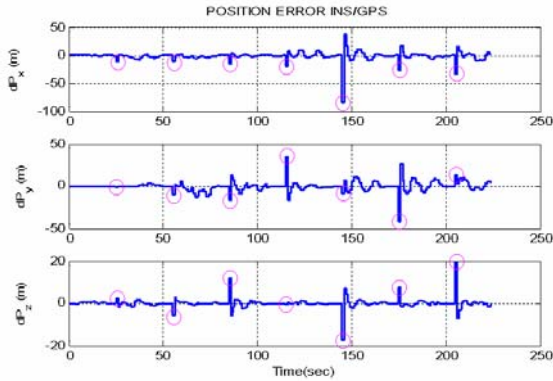

Fig. 8. Position error in case of six measurement error model, 10 second GPS outage

In order to compare the new method with the conventional one, Fig. 8 shows the error in the case without orientation measurements (6 measurement error model). The maximum error is 85m after 10 second outage. The average error is around 18m for 10 second outage.

So, with adding attitude and heading state for error model, the error in position is decrease. And as the result, both visible response and analytical evaluation show that the full measurements error model is better than the previous one with position and velocity measurements only.

The new orientation measurement scheme helps improving the accuracy of position and orientation determination two times up compared with previous scheme. But this rate becomes higher with worse calibrated gyrometers.

5   Conclusion

The developed system shows a stable and more accuracy solution for the INS/GPS integration. The reasons are explained as follows:

1. The integration method by means of error model is an efficient way in this case, when different frequency of data sources is used. Direct feedback error model which feedbacks the fixed output to the INS mechanization to keep error from growing in the integration loop.

2. Full measurements which are not only GPS position, velocity but also roll, pitch and heading are used for error model. Because every INS sensor errors are fixed, the system will maintain the accuracy without restarting.

The attitude estimation compensating external acceleration and compound method of heading estimation give quite good orientation solution for the integration system.

There still be some more works necessary for the accuracy improvement of this integration system. Those are: considering of vehicle model constraints, modelling sensor and adding bias states.

Acknowledgements

References

1. Salah, S., Eduardo, M.N., Hugh, F.D.-W.: A High Integrity IMU/GPS Navigation Loop for Autonomous Land Vehicle Applications. IEEE Transactions on Robotics and Automation 15(3), 572–578 (1999)
2. Suh, Y.S., Park, S.K., Kang, H.J., Ro, Y.S.: Attitude Estimation Adaptively Compensating External Acceleration. JSME International Journal 49(1), 172–179 (2006)
3. Kim, J.H., Salah, S.: 6DoF SLAM aided GNSS/INS Navigation in GNSS Denied and Unknown Environments. Journal of Global Positioning Systems 4(1-2), 120–128 (2005)
4. Hong, S., Lee, M.H., Chun, H.H., Kwon, S.H., Jason, L.S.: Observability of Error States in GPS/INS Integration. IEEE Transactions on Vehicular Technology 54(2), 731–743 (2005)
5. Nebot, E., Durrant-Whyte, H.: Initial Calibration and Alignment of Low cost Inertial Navigation Units for land vehicle applications. Journal of Robotics Systems 16(2), 81–92 (1999)
6. Jay, F., Matthew, B.: The Global Positioning System and Inertial Navigation. McGraw-Hill Companies, United States of America (1999)
7. Chatfield: Fundamentals of High Accuracy Inertial Navigation. American Institute of Aeronautics and Astronautics (1997)
8. David, M.B., Ryu, J., Gerdes, J.C.: Integrating INS Sensors With GPS Measurements for Continuous Estimation of Vehicle Side slip, Roll, and Tire Cornering Stiffness. IEEE Transations on intelligent transportation systems 7(4), 483–493 (2006)
9. Jianchen, G., Petovello, M.G., Cannon, M.E.: GPS/Low-Cost IMU/Onboard Vehicle Sensors Integrated Land Vehicle Positioning System. EURASIP Journal on Embedded Systems (2007)
10. Wang, J.H., Yang, G.: A new magnetic compass calibration algorithm using neural networks. Meas. Sci. Technol. 17, 153–160 (2006)
11. Gebre-Egziabher, D., et al.: Inertially-Aided GPS-based Attitude Heading Reference System (AHRS) for General Aviation Aircraft. In: Proceedings of the ION GPS 1997, Kansas City, MO (1997)

Synchronization Behavior Analysis for Coupled Lorenz Chaos Dynamic Systems via Complex Networks

Yuequan Yang[1], Xinghuo Yu[2], and Tianping Zhang[1]

[1] Department of Automation, College of Information Engineering
Yangzhou University, Yangzhou 225009, China
[2] School of Electrical and Computer Engineering
RMIT University, Melbourne VIC 3001, Australia
`yangyq@yzu.edu.cn`

Abstract. Of particular interest is intrinsic principle for synchronization behavior of complex networks. The synchronization behaviors of coupled Lorenz chaos systems via three kinds of networks are explored, respectively. The bounded property and synchronization criteria for complex networks under consideration are derived. The intrinsic principle for synchronization behavior depends on synchronization error matrix and coupling strength.

Keywords: complex networks, synchronization, Lorenz chaos systems.

1   Introduction

As known, Lorenz chaos system has quite complex and abundant property, such as homoclinic bifurcation, period doubling phenomena, pre-turbulence, intermittent chaos[1-3]. What will happen when multiple Lorenz chaos systems coupled via complex network? When and why do coupled Lorenz chaos systems become synchronization behavior via complex network? Motivated by research on complex network nowadays, this note attempts to make a further study on these problems. Especially, one of the interesting and significant phenomena in complex networks is the synchronization [8-16]. The famous master stability function has been established to decide whether or not any linear coupling arrangement produce stable synchronization dynamics in [8], while variations of desynchronization bifurcation mode have been revealed with change of coupling scheme and coupling strength. Based on master stability function rationale, synchronizability and synchronization enhancement of complex networks has been discussed using eigenratio λ_N / λ_2 of out-coupling matrix by rewiring links and assigning weights for links [9-16].

In this note, we attempt to seek intrinsic reason for variant synchronization behavior of complex network. More attention is paid to synchronization behavior of variant complex networks by employing an out-coupling parameter. The synchronization analysis here is established not on local linearization, but global bounded property of coupled Lorenz systems via complex network.

2 The Simplest Network with Two Nodes

For a complex network with N identical linearly and diffusively coupled Lorenz chaos nodes, where the ith node is a Lorenz chaos system

$$\dot{\zeta}_i(t) = f(\zeta_i(t)) + c\sum_{j=1}^{N} a_{ij}\Gamma\zeta_j(t) \tag{1}$$

where $\zeta_i(t) = (x_i, y_i, z_i)^T$ is system state of the ith node, c is called the coupling strength, $A = (a_{ij})_{N\times N}$ called out-coupling matrix, $\Gamma \in R^{3\times3}$ called inner-coupling matrix which is a constant 0-1 matrix linking between coupled system variables.

Assumption 1. Irreducible Out-coupling matrix A is diagonally dominant matrix

$$\mu + \sum_{j=1, j\neq i}^{N} a_{ij} = \mu + \sum_{j=1, j\neq i}^{N} a_{ji} = -a_{ii} \tag{2}$$

$i = 1, 2, \cdots, N$, $\mu \geq 0$, μ is called out-coupling parameter in this note.

Remark 1. It is worth pointing out that diagonally dominant out-coupling matrix guarantee its eigenvalues are negative if $\mu > 0$, which generates interesting synchronization exhibition.

Assumption 2. Inner-coupling matrix $\Gamma \in R^{3\times3}$ is defined as $\Gamma = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix}$.

According to Assumption 2, the simplest network with two autonomous nodes is

$$\zeta_i(t) = \begin{cases} \dot{x}_i = \sigma(y_i - x_i) \\ \dot{y}_i = rx_i - y_i - x_iz_i + ca_{i1}y_1 + ca_{i2}y_2 \\ \dot{z}_i = x_iy_i - bz_i \end{cases} \tag{3}$$

where $\sigma > 1$, $r > 1$, $i = 1, 2$, and the out-coupling matrix is $A = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}$.

To begin with, a following lemma is given as follows.

Lemma 1. For the coupled two Lorenz systems (3), the trajectories of system (3) are bounded and tend inside the following six dimensional hyper- sphere with the radius

$$R = 2r\max\left\{\sqrt{b/r}, 1, \sqrt{b}\right\} \tag{4}$$

Proof: Define the Lyapunov function candidate

$$V_1 = \sum_{i=1}^{2} rx_i^2 + \sigma y_i^2 + \sigma(z_i - 2r)^2 \tag{5}$$

Its time derivative satisfies

$$\dot{V}_1 = \sum_{i=1}^{2} [-2\sigma(rx_i^2 + y_i^2 + b(z_i - r)^2) + 2\sigma br^2] + 2c\sigma\varsigma^T A\varsigma \tag{6}$$

$$\leq \sum_{i=1}^{2} [-2\sigma(rx_i^2 + y_i^2 + b(z_i - r)^2) + 2\sigma br^2] \tag{7}$$

where $\varsigma = (y_1, y_2)^T$. It is easy to know (7) follows from the fact the Assumption 2. It is apparently seen that $\dot{V}_1 = 0$ determines a hyper-ellipsoid with six dimensions, beyond which the derivative of V_1 is negative. Then this hyper-ellipsoid lies inside the hyper-sphere (4), that is, the coupled systems (3) is bounded.

From now on, denote $\tilde{R}_y$ and $\tilde{R}_z$ as the bound of state y and z.

Theorem 1. Suppose $\tilde{R}_y^2 < 4rb < b\tilde{R}_z^2 + \tilde{R}_y^2$, if

$$c > (b\tilde{R}_z^2 + \tilde{R}_y^2 - 4br)/(8rb - 2\tilde{R}_y^2) \tag{8}$$

then the global synchronization of (3) is achieved.

Proof: Define the Lyapunov function

$$V_2 = 1/2\left[r(x_1 - x_2)^2 + \sigma(y_1 - y_2)^2 + \sigma(z_1 - z_2)^2 \right] \tag{9}$$

Its time derivative along with (3) satisfies

$$\dot{V}_2 = -r\sigma(x_1 - x_2)^2 + \sigma(y_1 - y_2)\{-(1+2c)(y_1 - y_2) - (x_1 - x_2)z_1\} \\ + \sigma(z_1 - z_2)\{(x_1 - x_2)y_1 - b(z_1 - z_2)\} \tag{10}$$

$$\leq -r\sigma(x_1 - x_2)^2 + \sigma|(y_1 - y_2)(x_1 - x_2)||z_1| - \sigma(1+2c)(y_1 - y_2)^2 \\ + \sigma|(z_1 - z_2)(x_1 - x_2)||y_1| - b\sigma(z_1 - z_2)^2 \tag{11}$$

$$= \zeta^T M \zeta = \zeta^T \begin{bmatrix} -r\sigma & \sigma\tilde{R}_z/2 & \sigma\tilde{R}_y/2 \\ \sigma\tilde{R}_z/2 & -\sigma(1+2c) & 0 \\ \sigma\tilde{R}_y/2 & 0 & -b\sigma \end{bmatrix} \zeta \tag{12}$$

where $\zeta = \left(\left| x_1 - x_2 \right|, \left| y_1 - y_2 \right|, \left| z_1 - z_2 \right| \right)^T$. It is necessary to note that (10) follows from the following facts $x_1 y_1 - x_2 y_2 = (x_1 - x_2) y_1 + x_2 (y_1 - y_2)$ and $x_1 z_1 - x_2 z_2 = (x_1 - x_2) z_1 + x_2 (z_1 - z_2)$, if

$$-r\sigma + \left[\sigma \tilde{R}_z / 2 \quad \sigma \tilde{R}_y / 2 \right] \begin{bmatrix} 1/(\sigma(1+2c)) & 0 \\ 0 & 1/(b\sigma) \end{bmatrix} \begin{bmatrix} \sigma \tilde{R}_z / 2 \\ \sigma \tilde{R}_y / 2 \end{bmatrix} < 0 \qquad (13)$$

For M is negative definite matrix which comes from that the coupling strength c satisfies (8), then there exists $\dot{V}_2 < 0$. It follows then that the following

$$\int_0^\infty \frac{1}{2} \left[r(x_1 - x_2)^2 + \sigma(y_1 - y_2)^2 + \sigma(z_1 - z_2)^2 \right] < V_2(0) \qquad (14)$$

By virtue of Barbalat lemma, it yields

$$\left| x_1 - x_2 \right| \to 0, \left| y_1 - y_2 \right| \to 0, \left| z_1 - z_2 \right| \to 0, \text{ as } t \to \infty \qquad (15)$$

Hence, global synchronization of the coupled system (3) is achieved.

3   Completely Coupled Complex Networks

Here, consider a completely coupled complex network system with the ith node

$$\dot{\zeta}_i(t) = \begin{cases} \dot{x}_i = \sigma(y_i - x_i) \\ \dot{y}_i = rx_i - y_i - x_i z_i + c \sum_{j=1}^{N} a_{ij} y_j(t) \\ \dot{z}_i = x_i y_i - b z_i \end{cases} \qquad (16)$$

where c is the coupling strength, for out-coupling matrix, $a_{ij} = a_{ji} = 1$, $i \neq j$.

Lemma 2. For complex network systems (16), the trajectories are bounded and tend inside following hyper-sphere with $3N$ dimensions with the radius

$$R = \sqrt{2N} r \max \left\{ \sqrt{b/r}, 1, \sqrt{b} \right\} \qquad (17)$$

Proof: Its proof is similar to the proof of Lemma 1, by employing the Lyapunov function candidate $V_3 = \sum_{i=1}^{N} rx_i^2 + \sigma y_i^2 + \sigma(z_i - 2r)^2$, and is omitted here.

Theorem 2. Suppose $\tilde{R}_y^2 < 4rb < b\tilde{R}_z^2 + \tilde{R}_y^2$, if

$$c > (b\tilde{R}_z^2 + \tilde{R}_y^2 - 4br)/((4rb - \tilde{R}_y^2)(N+\mu)) \tag{18}$$

then system (16) is global synchronization.

Proof: Define the Lyapunov function

$$V_4 = \frac{1}{2}\sum_{1 \le i < j \le N} r(x_i - x_j)^2 + \sigma(y_i - y_j)^2 + \sigma(z_i - z_j)^2 \tag{19}$$

Its time derivative along with (16)

$$\dot{V}_4 \le \sum_{1 \le i < j \le N} \{-r\sigma(x_i - x_j)^2 + \sigma|y_i - y_j||x_i - x_j||z_i| - \sigma(1+cN+c\mu)(y_i - y_j)^2$$

$$+ \sigma|z_i - z_j||x_i - x_j||y_i| - b\sigma(z_i - z_j)^2$$

$$\le \sum_{1 \le i < j \le N} \tilde{\zeta}_{ij}^T \Delta \tilde{\zeta}_{ij} \tag{20}$$

where $\tilde{\zeta}_{ij} = \left(|x_i - x_j|, |y_i - y_j|, |z_i - z_j|\right)^T \in R^3$, $\Delta = \begin{bmatrix} -r\sigma & \sigma\tilde{R}_z/2 & \sigma\tilde{R}_y/2 \\ \sigma\tilde{R}_z/2 & -\sigma(1+cN+c\mu) & 0 \\ \sigma\tilde{R}_y/2 & 0 & -b\sigma \end{bmatrix}$.

By virtue of Schur Complement, it is not difficult to know is Δ negative definite if

$$-r\sigma + \begin{bmatrix} \sigma\tilde{R}_z/2 & \sigma\tilde{R}_y/2 \end{bmatrix} \begin{bmatrix} 1/(\sigma(1+cN+c\mu)) & 0 \\ 0 & 1/(b\sigma) \end{bmatrix} \begin{bmatrix} \sigma\tilde{R}_z/2 \\ \sigma\tilde{R}_y/2 \end{bmatrix} < 0. \tag{21}$$

That is, if c satisfies (18), then $\dot{V}_4 < 0$. The coupled system (16) is global synchronization.

4   General Complex Networks

Consider a general complex networks with ith node

$$\zeta_i(t) = \begin{cases} \dot{x}_i = \sigma(y_i - x_i) \\ \dot{y}_i = rx_i - y_i - x_i z_i + c\sum_{j=1}^{N} a_{ij} y_j(t) \\ \dot{z}_i = x_i y_i - bz_i \end{cases} \tag{22}$$

where the out-coupling matrix $A = (a_{ij})_{N \times N}$ with general form, $a_{ij} = a_{ji} = 1$ if there is a link between node i and node j, and $a_{ij} = a_{ji} = 0$ otherwise. Define a matrix called synchronization error matrix as

$$\hat{M} = \begin{bmatrix}
-2a_{12} & -a_{13} & \cdots & -a_{1N} & a_{23} & a_{24} & \cdots & a_{2N} & 0 & \cdots & 0 & \cdots & 0 \\
-a_{12} & -2a_{13} & \cdots & -a_{1N} & -a_{23} & 0 & \cdots & 0 & a_{34} & \cdots & a_{3N} & \overset{\tilde{N}-3N+5}{0} & 0 \\
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\
-a_{12} & -a_{13} & \cdots & -2a_{1N} & 0 & 0 & \cdots & -a_{N2} & 0 & \cdots & -a_{N3} & \cdots & -a_{N-1,N} \\
a_{21} & -a_{31} & \overset{N-4}{0} & 0 & -2a_{23} & -a_{24} & \cdots & -a_{2N} & a_{34} & \cdots & a_{3N} & \overset{\tilde{N}-3N+5}{0} & 0 \\
a_{21} & 0 & -a_{41} & \overset{N-4}{0} & -a_{23} & -2a_{24} & \cdots & -a_{2N} & \overset{N-5}{-a_{43}} & 0 & 0 & \overbrace{a_{45} \cdots a_{4N}}^{N-4} & \overset{\tilde{N}-4N+9}{0} \\
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\
a_{21} & 0 & \overset{N-4}{0} & -a_{N1} & -a_{23} & -a_{24} & \cdots & -2a_{2N} & 0 & \overset{N-5}{0} & -a_{N3} & \cdots & -a_{N-1,N} \\
0 & a_{31} & a_{41} & \overset{N-4}{0} & a_{32} & -a_{42} & \overset{N-5}{0} & 0 & -2a_{34} & \cdots & a_{3N} & -a_{45} & \cdots \\
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\
0 & \overset{N-4}{0} & a_{N-1,1} & -a_{N1} & 0 & \overset{N-5}{0} & a_{N-1,2} & -a_{N2} & \overset{N-5}{0} & a_{N-1,3} & -a_{N3} & \cdots & -2a_{N-1,N}
\end{bmatrix}$$

Theorem 3. Suppose $\tilde{R}_y^2 < 4rb < b\tilde{R}_z^2 + \tilde{R}_y^2$, if

$$c > (b\tilde{R}_z^2 + \tilde{R}_y^2 - 4br)/((4rb - \tilde{R}_y^2)(\mu - \lambda_{\max}(\hat{M}))) \tag{23}$$

then system (22) is global synchronization.

Proof: Choose the Lyapunov function

$$V_5 = (1/2) \sum_{1 \le i < j \le N} r(x_i - x_j)^2 + \sigma(y_i - y_j)^2 + \sigma(z_i - z_j)^2 \tag{24}$$

Its time derivative along with (22) satisfies

$$\dot{V}_5 = \sum_{1 \le i < j \le N} r(x_i - x_j)(\dot{x}_i - \dot{x}_j) + \sigma(y_i - y_j)(\dot{y}_1 - \dot{y}_2) + \sigma(z_i - z_j)(\dot{z}_1 - \dot{z}_2)$$

$$= \sum_{1 \le i < j \le N} \{ r\sigma(x_i - x_j)\left[(y_i - y_j) - (x_i - x_j)\right]$$

$$+ \sigma(y_i - y_j)\left[r(x_i - x_j) - (y_i - y_j) - (x_i z_i - x_j z_j) - c\mu(y_i - y_j)\right] \tag{25}$$

$$+ \sigma(z_i - z_j)\left[(x_i z_i - x_j z_j) - b(z_i - z_j)\right]\} + \Phi$$

where $\Phi = c\sigma \sum_{i < j} (y_i - y_j)\left[\sum_{k=1}^{N} a_{ik} y_k - \sum_{k=1}^{N} a_{jk} y_k\right]$. Next, derive this term

$$\Phi = c\sigma \sum_{i<j} (y_i - y_j) \left[\sum_{k=1}^{N} (a_{ik} - a_{jk}) y_k \right] = \psi^T \bar{M} \xi \qquad (26)$$

where $\xi = (y_1, y_2, \cdots, y_N)^T$, $\psi = (y_1 - y_2, y_1 - y_3, \cdots, y_i - y_j, \cdots, y_{N-1} - y_N)^T$,

$$\bar{M} \in R^{\tilde{N} \times N}, \ \bar{M} = \begin{bmatrix} a_{11} - a_{21} & a_{12} - a_{22} & \cdots & a_{1N} - a_{2N} \\ a_{11} - a_{31} & a_{11} - a_{32} & \cdots & a_{1N} - a_{3N} \\ \vdots & \vdots & \vdots & \vdots \\ a_{i1} - a_{j1} & a_{i2} - a_{j2} & \cdots & a_{iN} - a_{jN} \\ \vdots & \vdots & \vdots & \vdots \\ a_{N-1,1} - a_{N1} & a_{N-1,2} - a_{N2} & \cdots & a_{N-1,N} - a_{NN} \end{bmatrix}$$. Further,

$$\Phi = \psi^T \hat{M} \psi \qquad (27)$$

where $\hat{M} \in R^{\tilde{N} \times \tilde{N}}$ is shown as above. Appling (25) and (27), it yields

$$\dot{V}_5 \le \sum_{i<j} \{ -r\sigma(x_i - x_j)^2 + \sigma |y_i - y_j| |x_i - x_j| |z_i|$$

$$+ \sigma |z_i - z_j| |x_i - x_j| |y_i| - b\sigma(z_i - z_j)^2 - \sigma(1 + c\mu - c\lambda_{max}(\hat{M})) \sum_{i<j} (y_i - y_j)^2$$

$$= \sum_{i<j} \begin{pmatrix} |x_i - x_j| \\ |y_i - y_j| \\ |z_i - z_j| \end{pmatrix}^T \begin{bmatrix} -r\sigma & \sigma \tilde{R}_z/2 & \sigma \tilde{R}_y/2 \\ \sigma \tilde{R}_z/2 & -\sigma - \sigma c\mu + \sigma c\lambda_{max}(\hat{M})) & 0 \\ \sigma \tilde{R}_y/2 & 0 & -b\sigma \end{bmatrix} \begin{pmatrix} |x_i - x_j| \\ |y_i - y_j| \\ |z_i - z_j| \end{pmatrix} < 0 \qquad (28)$$

Note that the last inequality follows from (23). Thus, similarly, by virtue of Barbalat lemma, there are $|x_i - x_j| \to 0, |y_i - y_j| \to 0, |z_i - z_j| \to 0$, as $t \to \infty$. Consequently, the goal of synchronization of system (22) is achieved.

Remark 2. From the synchronization error matrix $\hat{M}$, the insightful understanding of out-coupling matrix will help explore deep synchronization characteristic for general complex networks, which needs further study.

5 Simulation Tests

Consider two coupled Lorenz systems, completely coupled complex network, and general complex network cases. Let $\sigma = 10, b = 8/3$. To begin with, assume the

initial states are $\zeta_1 = \begin{bmatrix} -50 & -15.4 & 19 \end{bmatrix}$ and $\zeta_2 = \begin{bmatrix} -54 & 10 & -4 \end{bmatrix}$. Fig.1 shows the good synchronization behavior with $\mu = 0$, $c = 2$, and $r = 28$. For completely coupled complex network case, synchronization behaviors are shown in Fig.2 with $c = 0.8$ $r = 216.2$ and $\mu = 20$. For a general complex network with $N = 100$ Lorenz nodes is established. the synchronization behavior exhibits a periodic circle orbit with $c = 0.5$, $r = 216.2$ and $\mu = 10$, in Fig. 3.
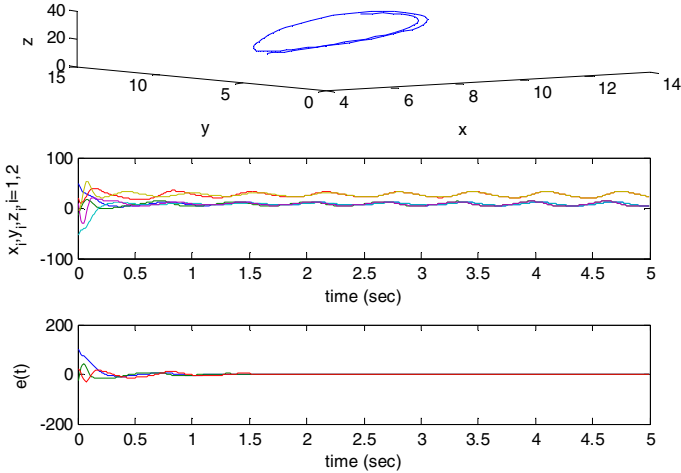

Fig. 1. Synchronization behavior of two coupled Lorenz chaos systems with $\mu = 0$
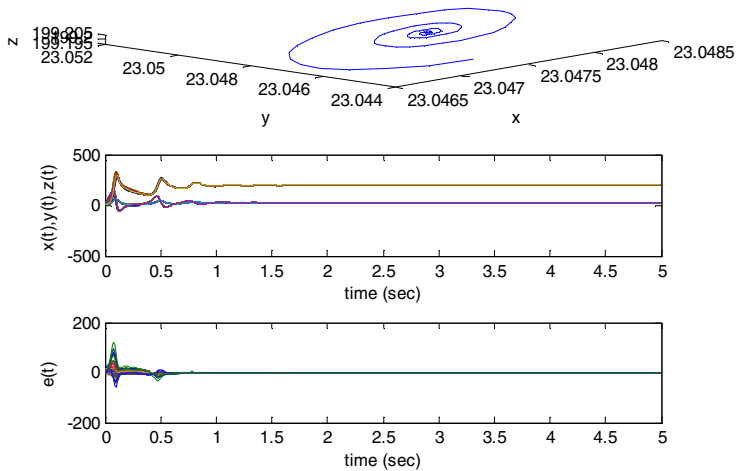

Fig. 2. Synchronization behavior of completely coupled network with $\mu = 20$
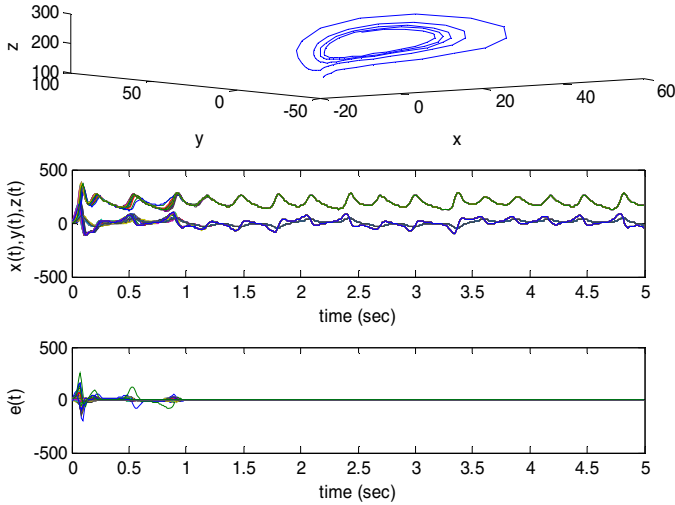
Fig. 3. Synchronization behavior over general complex network with $\mu = 10$

Remark 3. It is worth noting that much larger coupling strength may degenerate into desynchronization. One of reasons may be that out-coupling actions seriously deteriorate some individual chaos dynamics which is so sensitive to aggregated errors from the coupled nodes.

6   Conclusion

The dynamic behaviors of coupled Lorenz system via complex network were investigated. The bounded property and synchronization criterion for complex network under consideration were explored. Through observations, intrinsic principle for synchronization behavior of complex network lies in the synchronization error matrix and coupling strength.

Acknowledgment

References

1. Sparrow, C.: The Lorenz Equations: Bifurcations, Chaos, and Strange Attractors. Springer, New York (1982)
2. Fradkov, A.L., Pogromsky, A.Y.: Introduction to Control of Oscillations and Chaos. World Scientific Press, Singapore (1998)

3. Chacon, R.: Control of Homoclinic Chaos by Weak Periodic Perturbations. World Scientific Press, Singapore (1998)
4. Watts, D.J., Strogatz, S.H.: Collective Dynamics of Small-world Networks. Nature 393, 440–442 (1998)
5. Strogatz, S.H.: Exploring Complex Networks. Nature 410, 268–276 (2001)
6. Barabasi, A.L., Albert, R.: Emergence of Scaling in Random Networks. Science 286, 509–512 (1999)
7. Barabasi, A.L., Albert, R., Jeong, H.: Mean-field Theory for Scale-free Random Networks. Physcia A 272, 173–187 (1999)
8. Pecora, L.M., Carroll, T.L.: Master Stability Functions for Synchronized Coupled Systems. Physical Review Letters 80, 2109–2112 (1998)
9. Chavez, M., Hwang, D., Amann, A., et al.: Synchronizing is Enhanced in Weighted Complex Networks. Physical Review Letters 94, 218701 (2005)
10. Jalili, M., Ajdari, R.A., Hasler, M.: Enhancing Synchronizability of Dynamical Networks Using the Connection Graph Stability Method. Int. J. Ciurcuit Theory Appl. 35, 611–622 (2007)
11. Jalili, M., Ajdari, R.A., Hasler, M.: Enhancing Synchronizability of Weighted Dynamical Networks Using Betweenness Centrality. Physical Review E 78, 016105 (2008)
12. Motter, A.E., Zhou, C.S., Kurths, J.: Enhancing Complex-network Synchronization. Europhysics Letters 69, 334–340 (2005)
13. Donetti, L., Hurtado, P.I., Munnoz, M.A.: Entangled Networks, Synchronization, and optimal Networks. Physical Review Letter 95, 188701 (2005)
14. Wu, C.W.: Synchronization and Convergence of Linear Dynamics in Random Directed Networks. IEEE Trans. on Automatic Control 51(7), 1207–1210 (2005)
15. Wu, C.W.: Synchronization in Systems Coupled via Complex Networks. ISCAS IV, 724–727 (2004)
16. Wu, C.W.: Perturbation of Coupling Matrices and Its Effect on the Synchronizability in Arrays of Coupled Chaotic Circuits. Physics Letters A 319, 495–503 (2003)

A Fast Iteration Algorithm for Solving the Geocentric Earth Latitude Coordination Based on the Geocentric Cartesian Coordination⋆

Da Lu, Wen-Bo Zhao, and Ji-Yan Du

Artillery Academy of People Liberation Army, Heifei, AnHui230031, P.R. China
ludahyuk@yahoo.com.cn

Abstract. A fast iteration algorithm proposed is applied to solve the latitude of geocentric earth coordination from geocentric Cartesian coordination. Comparing with other well-known algorithms, the simulation results demonstrate that the algorithm is better in computation speed, computation precision and simpler in computation complexity; therefore, it is available for the portable.

Keywords: the geocentric Cartesian coordination; the geocentric earth coordination; the coordination transformation; the earth latitude.

1   Introduction

Currently the global positioning technology has been relatively mature .The Global Positioning System[1,2] has been widely used in many fields, such as military, transportation, aerospace, mining and so on. The format of Earth geographical information, which the Global Positioning System providing to the users, is named the geocentric earth coordination. The origin of the geocentric coordinates is the mass center of of the earth ellipsoid. The geocentric coordinates is the coordinates built up based on the relation between the earth rotation axes and the earth equator. The geocentric coordinates can be divided into the geocentric earth coordinates and the geocentric Cartesian (inert) coordinates. For the given point of the earth, its geocentric earth coordinate is expressed as the vector $[L, B, H]^T$, where the earth longitude L is the angle between the meridian plane located in this given point and Greenwich meridian plane, with eastwardly positive and westwardly negative. Where the earth latitude B is the angle between the normal line of the earth ellipsoid passing through the given point and the earth equator surface, with northwardly positive and southwardly negative. Where the altitude H is the distance from the sea level to the given point along the normal line of the earth ellipsoid. starting from the ellipsoid, with outwardly positive, inwardly negative. The definition rules of the origin and coordinate axes of the geocentric inert coordinates are described as followings. The origin O is the mass center (also called centroid) of the earth. Z axes

---

points to the north pole of the earth, and X axes is the intersection line between Greenwich meridian plane and the earth equator, and Y axes is perpendicular to XOZ plane, the given point of the earth can be expressed in the form of the geocentric inert coordinate as $[X_e, Y_e, Z_e]^T$. For some given point on earth, the two coordination forms have one to one corresponding relationships.

2 The Coordinates Transform Relations between the Geocentric Earth Coordinate System and the Geocentric Cartesian Coordinate System

For any point on earth, transform from the geocentric earth coordinates to the geocentric Cartesian coordinate conversion is relatively simple, detail expression is as follows:

$$\begin{cases} X_e = (N + H)\cos(B)\cos(L) \\ Y_e = (N + H)\cos(B)\sin(L) \\ Z_e = \left(N\left(1 - e^2\right) + H\right)\sin(B) \end{cases} \tag{1}$$

Where $[L, B, H]^T$ is the geocentric earth coordinates and $[X_e, Y_e, Z_e]^T$ is the geocentric Cartesian coordinates. N is the curvature radius of Earth prime vertical, its expression is $N = \frac{a}{\sqrt{1 - e^2 \sin^2(B)}}$. a is the radius of the long axes of Earth ellipsoid. e is the Earth first eccentricity, the terms on the ellipsoid of the earth, eccentricity and prime vertical can be consulted the reference[1,2].

According to (1),the coordinate transform form the Geocentric Cartesian coordinates to the Geocentric earth coordinates can be proved as the followings.

By the Geocentric rectangular coordinates Y_e and X_e of (1),the Geocentric earth longitude L can be obtained based on the following expression.

$$L = \arctan\left(\frac{Y_e}{X_e}\right) \tag{2}$$

Then, let's discuss to seek the expression of the Geocentric earth latitude B.

According to (1) , it can be obtained by the following expression.

$$\begin{cases} \cos(L) = \frac{X_e}{(N+H)\cos(B)} \\ \sin(L) = \frac{Y_e}{(N+H)\cos(B)} \\ \sin(B) = \frac{Z}{N(1-e^2)+H} \end{cases} \tag{3}$$

It is well known the (4) is correct for any angle L.

$$\sin^2(L) + \cos^2(L) = 1 \tag{4}$$

Inserting $sin(L)$ and $cos(L)$ of (3) into (4), it can be obtained as

$$\cos(B) = \frac{\sqrt{(X_e)^2 + (Y_e)^2}}{N + H} \tag{5}$$

For any angle B, the following trigonometric function transform is correct.

$$\tan(B) = \frac{\sin(B)}{\cos(B)} \tag{6}$$

Inserting $cos(B)$ of (5) and $sin(B)$ of (3) into (6), it can be expressed as

$$\tan(B) = \frac{Z_e}{\sqrt{(X_e)^2 + (Y_e)^2}} \frac{N+H}{N(1-e^2)+H} \tag{7}$$

Simplifying (7) , following can be is

$$\tan(B) = \frac{Z_e}{\sqrt{(X_e)^2 + (Y_e)^2}} \left(1 + \frac{N_e^2}{N(1-e^2)+H}\right) \tag{8}$$

Inserting $N = \frac{a}{\sqrt{1-e^2 \sin^2(B)}}$ and $\frac{Z_e}{sin(B)} = N(1-e^2)+H$ into (7),

$$\tan(B) = \frac{Z_e}{\sqrt{(X_e)^2 + (Y_e)^2}} \left(1 + \frac{ae^2}{Z_e} \cdot \frac{\sin(B)}{\sqrt{1-e^2\sin(B)}}\right) \tag{9}$$

By them , the Geocentric earth latitude B can be obtained according to the Geocentric Cartesian coordinates as the following.

$$B = \arctan\left(\frac{Z_e}{\sqrt{(X_e)^2 + (Y_e)^2}} \left(1 + \frac{ae^2}{Z_e} \cdot \frac{\sin(B)}{\sqrt{1-e^2\sin(B)}}\right)\right) \tag{10}$$

Then the Altitude H will be discussed as the following section. In (1), summing the square of X_e and square of Y_e can be obtained as the following expression

$$H = \frac{\sqrt{(X_e)^2 + (Y_e)^2}}{\cos(B)} - N \tag{11}$$

Currently, the Geocentric earth coordinates can be obtained based on form of the Geocentric Cartesian coordinates as following as.

$$\begin{cases} B = \arctan\left(\frac{Z_e}{\sqrt{(X_e)^2+(Y_e)^2}}\left(1 + \frac{ae^2}{Z_e} \cdot \frac{\sin(B)}{W}\right)\right) \\ L = \arctan\left(\frac{Y_e}{X_e}\right) \\ H = \frac{\sqrt{(X_e)^2+(Y_e)^2}}{\cos(B)} - N \\ W = \sqrt{1 - e^2 \sin^2(B)} \end{cases} \tag{12}$$

3   The Fast Iteration Algorithm to Solve Geocentric Earth Latitude

3.1   Algorithm Principle

From (12) it can be seen that the expression of the Geocentric earth latitude B still contains itself expression $sin(B)$, it is necessary to find some other methods to solve. Taylor expansion method proposed in Reference[2] was used to approximate the analytical solution of Latitude B. However, the drawback for this method is that the solution of the physical process is difficult to understand, and the analytical computing processing had demand of a greater computational complexity, it is not suit for the application in the Pocket PC, portable devices such as mobile phone. This section a fast iterative algorithm was proposed to solve this problem. The proposed basic idea of the algorithm is that the given initial latitude B_0, the algorithm along the given direction to approximate the real latitude B. In the following analysis the realization of the principle of iteration will be given in detail. Supposed that (13) is

$$f(B) = B - \arctan\left(\frac{Z_e}{\sqrt{(X_e)^2 + (Y_e)^2}}\left(1 + \frac{ae^2\sin(B)}{Z_e\sqrt{1 - e^2\sin^2(B)}}\right)\right) \qquad (13)$$

the objective function, solution of the true value of latitude B^* , should be satisfied to the following equation as.

$$\min_B f(B) = f(B^*) \qquad (14)$$

In this way, the calculation process to solve the real value of latitude can be transformed to a algorithm that make (14) achieve the minimum value by variable B. There were many methods to Solve these optimization problems, such as dichotomy algorithm introduced in Reference[3], the dogleg trust-domain algorithm introduced in Reference[4], Levenberg-Marquardt(lm) algorithm introduced in Reference[5,6,7], and Gauss-Newton(gn) algorithm introduced in Reference[8]. Whether dogleg , lm, or gn algorithm, its computational complexity is relatively high, not suitable for fast calculation demands in portable devices. In the practical of engineering applications, the general requirement for quick computation is to have not only a more fast speed of computing, but also have a more high calculation accuracy. Based on the above requirement, considering the completing transform from Geocentric Cartesian coordinates to Geocentric earth coordinates, a kind of fast iterative algorithm was proposed in this paper.

Assumed that in the kth iterations approaching value of latitude is B_k, the $k+1th$ express of iterations value along gradient direction as the following as.

$$B_{k+1} = B_k - F^{-1}(B_k) f(B_k) \qquad (15)$$

Where $F^{-1}(B)$ is the first-order derivative about B, that is, $F(B) = \frac{df(B)}{d(B)}$.

In order to avoid solving the first-order derivative, difference quotient about B is introduced to replace the derivative, that is,

$$\frac{df(B_k)}{dB_k} = \frac{f(B_k + S_p) - f(B_k)}{S_p} \tag{16}$$

Where S_p is the iterative step, when the step is little enough, the above given expression can be approximated to Derivative $F(B_k)$ of Objective function $f(B)$ in B_k.

Further supposed that

$$\delta_k = F^{-1}(B_k)f(B_k) \tag{17}$$

Then the iterative formulation of (15) in the gradient direction can be written as

$$B_{k+1} = B_k - \delta_k \tag{18}$$

For the given point, it can be seen that the key, for the iterative method of latitude, is to solve δ_k, namely, how to find the optimal of step size of the latitude iterative algorithm, in the limited number of iterative steps of the iterative process the final accuracy requirements are satisfied. Based on this issue, the following paper will be discussed for the selection of the iteration step size. In according with (17), the following equation can be given by

$$\delta_k F(B_k) = f(B_k) \tag{19}$$

Inserting (16) into (19), it can be obtained by the following equation.

$$\frac{\delta_k}{S_p + \delta_k} f(B_k + S_p) = f(B_k) \tag{20}$$

Supposed that

$$Z_k = \frac{\delta_k}{S_p + \delta_k} \tag{21}$$

Equation (20) can be transformed to Equation (22)

$$Z_k f(B_k + S_p) = f(B_k) \tag{22}$$

The expression of Z_k can be obtained as follows:

$$Z_k = \frac{f(B_k)}{f(B_k + S_p)} \tag{23}$$

Inserting Z_k of (23) into (20), the optimal iterative step size can be given by.

$$\delta_k = \frac{S_p Z_k}{1 - Z_k} \tag{24}$$

So the iterative expression of latitude variables based on (15) is as following as:

$$B_{k+1} = B_k - \frac{S_p Z_k}{1 - Z_k} \tag{25}$$

3.2   The Routine of the Algorithm

According to the above analysis, the steps of the algorithm are as follows

(1)Setting of the initial value setting The initial values to be set are: initial value of latitude iteration B_0, initial value of the derivatives calculation step S_p. The selection of the latitude initial value B_0, will be directly impact on the number of the iteration calculation steps , in general, can be used the method by offording the estimated value in advance, which accuracy can be about 5 degrees.

From the equation (16),it can been known that, with S_p smaller, the numerical derivatives will be closer to the theory of derivatives of the objective function, so, in general, S_p will be selected as the smaller positive number, such as such as 0.01,0.1. When the iterative value closing the real value, with a large step size, the algorithm will appear oscillation , in order to ensure the sufficient accuracy, the last step size will be used in this paper as following as Equation (26).

$$S_P = c_f \times S_p \tag{26}$$

Where c_f is the control coefficient of step-size.
(2)According to Equation (13) to calculate $f(B_k)$ from of Step k
(3)According to Equation (23) to calculate Z_k
(4)According to Equation (24) to calculate δ_k
(5)According to Equation (25) to calculate B_{k+1}
(6)According to Equation (26) to calculate S_p
(7)Estimate the iterative error $|B_{k+1} - B_k|$ in order to determine whether a given standard error satisfied according to the follow equation.

$$|B_{k+1} - B_k| \le \varepsilon \tag{27}$$

If satisfied, then stop the calculation, otherwise, $k = k + 1$, turn to Step (2), continue to implement the algorithm step-by-step.
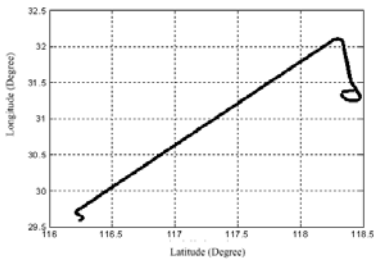

Fig. 1. 760 test points geographical location (geocentric earth coordination format)
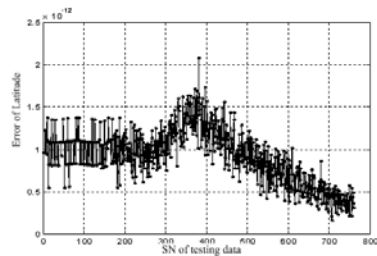

Fig. 2. The curve of absolute error of the proposed algorithm for 760 testing data
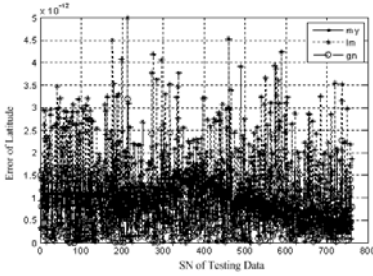
Fig. 3. Performance Comparative plot among the proposed algorithm (my), Levenberg Marquardt algorithm (lm) and Gauss-Newton algorithm (gn)
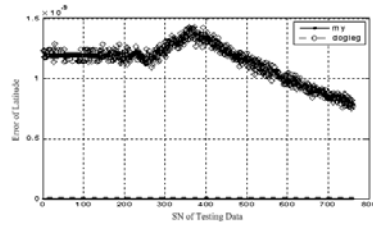
Fig. 4. Comparative performance plot between the proposed algorithm(my) and dogleg algorithm

Table 1. The result of the time-consuming of caculation were from 20 rutines of simulation testing

Algorithm name	The proposed algorithm	Dogleg algorithm	L-M algorithm	Gauss-Newton algorithm
time-consuming (sec.)	0.5573	9.2518	7.8318	12.9960

4   Algorithm Simulation and Result Analysis

In this paper, the position data recorded by the differential type GPS device were used to verify the validity of the proposed algorithm. The formation of the GPS data recorded was in Geogetic earth coordinates. The starting points of recorded data is Point $[118.3988589, 31.39042554, 5.25]^T$, and the end point is $[116.2413184, 29.59455702, 8094.88]^T$, totally 760 data were intercepted for testing, the locations of the testing data points were shown as Figure 1.

The detail testing methods were that, transform 760 data in Geogetic earth coordinates to Geocentric Cartesian coordinates, and then use the new algorithm to compute the Geogetic latitude from the Geocentric Cartesian coordinates for these 760 data.

The relative parameters of the simulation testing were given as follows, $S_p = 0.1$, $c_f = 0.1$, $\varepsilon = 10^{-6}$, $a = 6378137.0m$, $e^2 = 0.00669437999013$, $B_0 = 20°$.

The testing conditions for the simulation was as follows: the hardware platform is Intel Pentium 4 CPU, with frequency of 1.60GHz, with memory of 512MB. Software computing environments be that, the operating system is Windows XP, application software for calculation be Matlab 7.1.0.246 Service Pack3.

The testing results obtained by the proposed algorithm was shown as Figure 2. The algorithm error be very small. In order to comparative analysis the performance of the algorithm, for the same testing sample set, the dogleg, Levenberg-Marquardt and Gauss-Newton algorithm were used to test, includeding calculation time-consuming and computation errors . Comparative testing results were seen as Figure 3 and Figure 4. From Figure 3, the calculation error mean of the proposed algorithm in this paper is the same as Levenberg-Marquardt algorithm and the

Gauss-Newton algorithm ,but precision of the proposed algorithm is better than the above method, that is, error dispersion degree of the proposed algorithm is less than ones of Levenberg-Marquardt algorithm and Gauss-Newton algorithm.From Figure 4 it can be seen that the calculation error and the accuracy of the algorithm is much better than the dogleg algorithm. Four kinds of algorithms on the same sample set of calculated time-consuming was shown as Table 1 (In Table 1 the result of the time-consuming of calculation were from 20 routines of simulation testing). It can be seen that under the same conditions the time-consuming of calculating by the proposed the algorithm of this paper is the least. Therefore, the algorithm proposed in this paper,with faster calculation speed and higher accuracy, is a useful solution method with fast iterative process from the Geocentric Cartesian coordinates to the Geocentric earth coordinates.

References

1. Tong, X.R.: GPS/INS combination navigation & location and application, Change Sha: blishing House of University of Science and Technology of National Defense (1998)
2. Zhang, S.X.: The Theory and Aplication of GPS Satellite Position, Change Sha: blishing House of University of Science and Technology of National Defense (1996)
3. The materials preparation of froup, Operations Research, 154–158. Tsinghua University Press (1990)
4. Powell, M.J.D.: A Fortran Subroutine for Solving Systems of Nonlinear Algebraic Equations. In: Rabinowitz, P. (ed.) Numerical Methods for Nonlinear Algebraic Equations, ch. 7 (1970)
5. Levenberg, K.: A Method for the Solution of Certain Problems in Least-Squares. Quarterly Applied Mathematics 2, 164–168 (1944)
6. Marquardt, D.: An Algorithm for Least-squares Estimation of Nonlinear Parameters. SIAM Journal Applied Mathematics 11, 431–441 (1963)
7. Mor, J.J.: The Levenberg-Marquardt Algorithm: Implementation and Theory. In: Watson, G.A. (ed.) Numerical Analysis. Lecture Notes in Mathematics, vol. 630, pp. 105–116. Springer, Heidelberg (1977)
8. Dennis Jr., J.E.: Nonlinear Least-Squares. In: Jacobs, D. (ed.) State of the Art in Numerical Analysis, pp. 269–312. Academic Press, London

A Novel Time-Domain Structural Parametric Identification Methodology Based on the Equivalency of Neural Networks and ARMA Model

Bin Xu[1,2,*], Ansu Gong[1], Jia He[1], and Sami F. Masri[1,3]

[1] College of Civil Engineering, Hunan University, Changsha, 410082, Hunan, China
Tel./Fax: (86) 731 882 1856
[2] Key Laboratory of Building Safety and Energy Efficiency (Hunan University),
Ministry of ducation, Changsha, Hunan, 410082, P.R. China
[3] Department of Civil Engineering, University of Southern California, Los Angeles,
CA 90089, USA
`binxu@hnu.cn`

Abstract. On one hand, it has been demonstrated theoretically and verified numerically that neural networks can act as a time domain nonparametric modeling approach of engineering dynamic systems by forecasting their dynamic responses according to them in the past consequent time steps. On the other hand, as a time-domain auto-regressive method, the auto-regressive and moving average (ARMA) model has been widely employed to describe the mapping between structural dynamics response at a current time instant and them in the past previous time instants. The equivalency of the physical meaning of the neural network nonparametric model and the ARMA parametric model for dynamic systems is testified firstly in this paper. Then, a novel structural parametric identification methodology based on the nonparametric neural network model is proposed by the use of excitation and dynamic response measurement time series. The accuracy and efficacy of the proposed strategy for a multi-storey frame structure model are validated using the excitation and acceleration measurement time series of impact test.

Keywords: BP neural network; ARMA; multi-storey shear structure; dynamic response; time series; health monitoring; parameter identification.

1 Introduction

Many civil infrastructures are now deteriorating due to aging, misuse, lacking of proper maintenance, and, in some cases, overstressing as a result of increasing load demands and changing environments. Failure of these infrastructures often leads to a high social consequence. It is therefore critical to evaluate their current reliability, performance, and condition for the prevention of potential catastrophic events.

In the last few decades, many mathematical-model-based structural parameters identification algorithms for structural damage detection using eigenvalues or mode

---

* Corresponding author.

shapes and/or their derivatives have been proposed and widely employed to identify parameters of engineering structures. Comprehensive review on the development of the vibration-based identification methodologies can be found in references[1,2]. Besides the eigenvalues and/or mode shape based identification algorithms, several time-domain sequential algorithms such as the recursive least squares method and the sequential prediction error method have been developed.

On one hand, as one of the typical time domain auto-regressive method in the form of describing the mapping between the dynamics response at a current time instant and those in the previous time instants, the auto-regressive and moving average (ARMA) model has been widely employed to identify structural physical parameters. The engineering significance, characteristics, and the modeling approach of ARMA models for system parameters estimation have been introduced by Yang et al. and Lee et al. [3,4]. The ARMA model is a typical parametric modeling strategy.

On the other hand, Modeling and identification of civil engineering structures in parametric way present a challenging problem because of the complexity and nonlinearity which widely exist. Therefore, neural networks have been drawn considerable attention in civil engineering community as a non-parametric modeling strategy for linear and nonlinear structural members or components and structural systems, due mainly to their ability to approximate any arbitrary continuous function and mapping, and due to its parallel computation performance. For example, a magnetorheological (MR) damper is an important structural member for vibration control of engineering structures and it is difficult to model in a parametric format because there are so many parameters to be included and identified in the model. Chang and Roschke presented an alternative representation of an MR damper in terms of a multi-layer perceptron neural network [5]. Zapateiro et al. presented an approach of modeling a shear-mode MR damper based on the Bouc-Wen and Neural Network models [6].

Neural networks also show their ability in damage detection of engineering structures in both qualitative and quantitative way. Nakamura et al. employed a neural network based approach for the detection of changes in the characteristics of structure-unknown systems [7]. For nonlinear structure-unknown systems, Masri et al. proposed a nonparametric structural damage detection methodology based on neural network approaches for the health monitoring [8]. Even it has been widely recognized that neural network is a typical nonparametric identification and modeling methodology with strong robustness, neural networks can also be employed for parametric identification. A study by Masri et al. demonstrated that neural networks are powerful tools for the identification of systems typically encountered in the structural dynamics field [9]. Yun and Bahng presented a method for estimating the stiffness parameters of a complex structural system by using neural network [10]. Unlike any conventional system parameter identification technique that involves inverse analysis, those strategies provided by Xu et al., with the direct use of displacement, velocity, and/or acceleration or strain responses, can give the structural parameter identification results in a substantially faster way with two neural networks [11-12]. For large scale infrastructures, Xu and Wu proposed a decentralized and localized identification method [13-14].

In this paper, the equivalency of the above mentioned two time domain methodologies of ARMA model and neural network model is verified theoretically by the discrete solution of structural dynamic equation. Since the matrices in the ARMA model are fully determined by structural mass, stiffness and damping matrices, and the ARMA model and neural network based model describe the same physical mapping, it

is possible to determine structural matrices by the weights and thresholds of the trained neural network model of the corresponding structure, even though the trained neural network is a nonparametric model. From this point of view a novel structural parametric identification methodology based on the equivalency of neural network model and ARMA model is proposed by the use of excitation and acceleration response measurement time series. The accuracy and efficacy of the proposed strategy are validated experimentally with a lab scale 4-DOF frame model structure under hammer impact excitation.

2   Equivalency of Neural Network and ARMA Model and Parameters Identification Methodology

2.1   Equation of Motion

Consider a structure that can be modeled as an n-DOF linear system, the dynamics of the system can be described by the following equation of motion

$$M_0\ddot{x}(t) + C_0\dot{x}(t) + K_0 x(t) = L_0 u(t) . \tag{1}$$

where $\ddot{x}(t)$, $\dot{x}(t)$, and $x(t)$ are n-dimensional response vectors, $u(t)$ is a m-dimensional input vector, M_0, C_0 and K_0 are mass, damping, and stiffness matrices, respectively, L_0 is the n×m input coefficient matrix,   and dot denotes differential with respect to time.

The identification problem is to estimate the unknown system parameters based on observations of the input and the response of the system. To make the problem unique, the equation of motion can be rewritten as

$$\ddot{x}(t) + J\dot{x}(t) + Kx(t) = Lu(t) . \tag{2}$$

where $J = M_0^{-1}C_0$, $K = M_0^{-1}K_0$, and $L = M_0^{-1}L_0$.

Since the observations are usually made at sampling instants of time, it is analytically convenient to convert Equation (2) into a corresponding discrete time equation. Furthermore, in actual experiments, the response of the system can be measured in terms of displacement, velocity or acceleration, depending on the sensor types.

2.2   Continuous and Discrete Time State Space Equation

For the effectiveness in the discretization of continuous time variables, the state space equation is generally employed. Hence, defining the state vector $x_0(t)$ as

$$x_0(t) = \begin{Bmatrix} x(t) \\ \dot{x}(t) \end{Bmatrix} . \tag{3}$$

Equations (1) and (2) can be rewritten in the following state space equation as

$$\dot{x}_0(t) = Ax_0(t) + Bu(t) . \tag{4}$$

$$y(t) = Cx_0(t) + Du(t) . \tag{5}$$

where $y(t)$ is defined as acceleration response, and $D = L$.

$$A = \begin{bmatrix} -J & I \\ -K & 0 \end{bmatrix}, \quad B = \begin{bmatrix} B_1 \\ B_2 \end{bmatrix} = \begin{bmatrix} -JL \\ -KL \end{bmatrix}, \quad C = \begin{bmatrix} I & 0 \end{bmatrix}. \tag{6a-6c}$$

In this study, it is assumed that the input $u(k\Delta t)$ is constant over the time interval $k\Delta t \leq t < (k+1)\Delta t$ in which Δt is the sampling period. Then, the discrete time equations corresponding to Equation (4) and (5) can be obtained as

$$x_0(k+1) = Px_0(k) + Qu(k). \tag{7}$$

$$y(k) = Cx_0(k) + Du(k). \tag{8}$$

where $u(k)$, $y(k)$, and $x_0(k)$ denote the sampled data at $t = k\Delta t$, and

$$P = \exp(A\Delta t) = \sum_{i=0}^{\infty} \frac{(A\Delta t)^i}{i!}. \tag{9a}$$

$$Q = \int_0^{\Delta t} \exp(A \cdot \tau) d\tau \cdot B = A^{-1}(P - I)B. \tag{9b}$$

Even though the state space equation in continuous time is of the 'observable canonical form' (OCF), its counterpart in discrete time is not generally so. For the present purpose, it is convenient to transform Equations (7) and (8) into the OCF. Since the matrix C is of rank n, the observability index of Equations (7) and (8) is equal to two. Hence, the inverse of the observability matrix can be written as

$$\begin{bmatrix} C \\ CP \end{bmatrix}^{-1} = \begin{bmatrix} S_1 & S_2 \end{bmatrix}. \tag{10}$$

where S_1, and S_2 are $2n \times n$ submatrices. Then, the transformation matrix R into the OCF can be obtained as

$$R = \begin{bmatrix} PS_2 & S_2 \end{bmatrix}. \tag{11}$$

and Equations (7) and (8) can be transformed into the OCF as

$$x_c(k+1) = P_c x_c(k) + Q_c u(k). \tag{12}$$

$$y(k) = C_c x_c(k) + Du(k). \tag{13}$$

where

$$x_c(k) = Rx_0(k). \tag{14a}$$

$$P_c = R^{-1}PR = \begin{bmatrix} P_1 & I \\ P_2 & 0 \end{bmatrix}, \quad Q = R^{-1}Q = \begin{bmatrix} Q_1 \\ Q_2 \end{bmatrix}, \quad C_c = CR = \begin{bmatrix} I & 0 \end{bmatrix}. \tag{14b}$$

and P_1 and P_2 are $n \times n$ submatrices of P_c.

2.3 Equivalency of ARMA Model and Neural Networks Model and Paremeters Identification Strategy

By taking the Z-transform of Equation (12) and (13), the transfer function matrix in discrete time, $G(z)$, can be easily obtained as

$$y(k) = \overline{G}(z)u(k) = [(z^2 I - z P_1 - P_2)^{-1}(z Q_1 + Q_2) + D]u(k). \tag{15}$$

Equation (14) can be, equivalently, written as

$$y(k) = P_1 y(k-1) + P_2 y(k-2) + Du(k) + E_1 u(k-1) + E_2 u(k-2). \tag{16}$$

where $E_1 = Q_1 - P_1 D$ and $E_2 = Q_2 - P_2 D$.

The above equation is referred to as the ARMA model which is commonly used for sampled data systems. It should be noted that the auto-regressive (AR) parameters [P_1, P_2] are identical for each case of the measured response components. However, the moving average (MA) parameters [D, E_1, E_2] vary depending on the measured response components. If $y(k)$ is acceleration, however, it includes the terms related to $u(k)$ as well as $u(k-1)$ and $u(k-2)$.

It has been widely recognized that the ability of neural network to approximate arbitrary continuous function and its parallel computation character provide an efficient soft computing strategy for inverse analysis. Neural network can act as a nonparametric model of a structure by forecasting structural dynamic responses based on the past time steps. Therefore, the Equation (16) can be approximated and mapped by a back propagation (BP) neural net work which is trained to represent the mapping between the input $y(k-1)$, $y(k-2)$, $u(k)$, $u(k-1)$, $u(k-2)$ and the output $y(k)$. Since the trained neural networks describes the same mapping as the ARMA model shows, the coefficients of P_1, P_2, D, E_1, and E_2 of the ARMA model can be determined by the trained neural network.

2.4 Paremeters Identification Strategy Based on the Equivalency of ARMA Model and Neural Networks Model

From Equation (16), it can be seen that if $y(k-1)=[1; 0; 0;0]$, $y(k-2)=[0; 0; 0;0]$, $u(k)=[0; 0; 0;0]$, $u(k-1)=[0; 0; 0;0]$ and $u(k-2)=[0; 0; 0;0]$, the ratio of the output of the neural network to 1 is the first tier of P_1. Similarly, if $y(k-1)=[0; 1; 0;0]$, $y(k-2)=[0; 0; 0;0]$, $u(k)=[0; 0; 0;0]$, $u(k-1)=[0; 0; 0;0]$ and $u(k-2)=[0; 0; 0;0]$, the ratio of the output of the neural network to 1 is the second tier of P_1. The other coefficient matrix tiers of the ARMA model can be determined by the same way.

It should be noted that the input and the output vectors must have the same preprocessing and same post-processing as the training data for the BP neural network. In addition, in order to improve the of the identification accuracy, the value of l can be taken of a number of random real numbers between -1 and 1. Finally, the average of the identified matrices are treated as the final estimation of the parameters matrices P_1, P_2, D, E_1 and E_2 in Equation (16). In this paper, a group of one hundred group random real numbers are employed as input to the trained neural network.

After the parameters of P_1, P_2, D, E_1, and E_2 of the ARMA model are estimated, the system parameters $[J, K, L]$ of the differential equation can be recovered by the above proposed approach.

3   Experimental Verification

3.1   Experimental Setup

The experimental structure is the 4-story, 1-bay by 1-bay shear scale-model structure shown in Figure 1. The structure is 0.3m×0.4m in plan and is 1.2m tall. The overall mass of the structure is 12.81kg on the first, second and third floor, 12.53kg on the fourth floor. The cross-section of the columns is 30mm×5mm, and the thickness of the floor plates is 10mm.
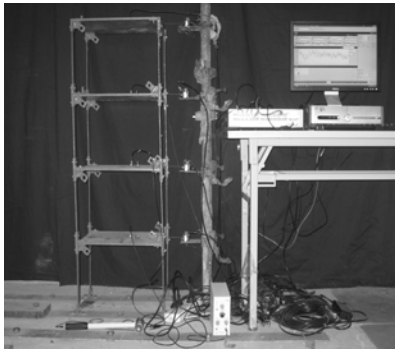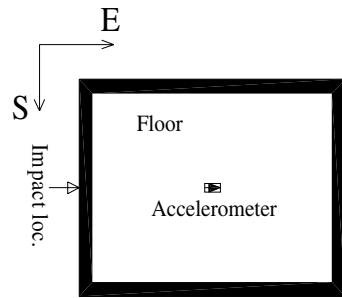

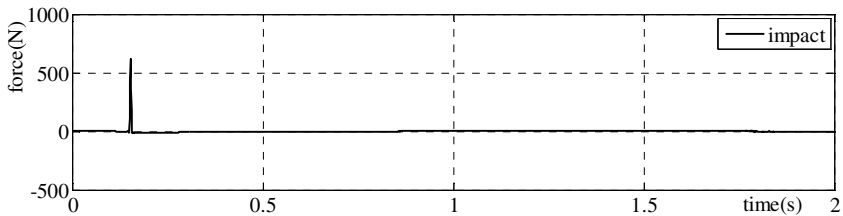**Fig. 1.** Steel scale-model structure   **Fig. 2.** Impact location and sensor placement

Fig. 3. Excitation force on the fourth floor of the structure

The forced vibration of the model structure is carried out by impact hammer in the test. The impact hammer has a force transducer, and this measurement was recorded during the hammer tests. The impact location was selected on the west face of the fourth floor of the frame structure and in the east direction. An accelerometer is install at the center of each floor of the model structure and therefore the acceleration responses of the four floors are measured. The excitation location and placement of accelerometers are shown in Figure 2.

3.2 Data Acquisition

The excitation and the acceleration responses are collected simultaneously using an LMS data acquisition system. A hammer external excitation force shown in Figure 3 is applied on the fourth floor of the structure. The sampling frequency is taken as 1024Hz. For the purpose of eliminating the effect of noise in the acceleration measurement, the acceleration responses obtained from the test are band-pass filtered with 3.0-50.0Hz.
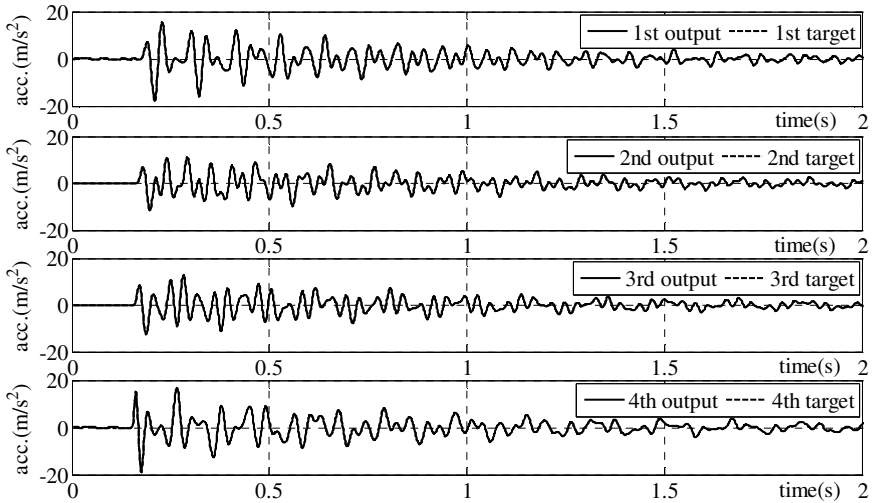

Fig. 4. Compare between the outputs and the targets of BP neural network

3.3 Training of Neural Networks

As described in section 2.3, a three-layer neural network is constructed and trained to model the 4-DOF structure under impact on the fourth floor. The input of the neural network model includes the acceleration responses at time steps $k-1$ and $k-2$ and the impact force at time steps k, $k-1$ and $k-2$. The output of the neural network is the acceleration response at time step k. For the 4-DOF structure, the neurons in the input, hidden and output layer of the neural network are 20, 35, and 4, respectively. Using the impact force and acceleration responses in the first 2 seconds, a total of 2048 sets of training data are constructed and employed to train the neural network.

Through the traditional error back-propagation algorithm, the neural network is trained to act as a nonparametric model of the structure in the form of forecasting the acceleration response. The comparison of the output vector and the target vector of the BP neural network are shown in Figure 4. The relative errors and root mean square (RMS) errors between the experimental measurements and the output of the trained neural networks are shown in Table 1. It is clear that the neural network is an ideal nonparametric model of the 4-DOF structure.

Table 1. The relative errors of the two curves in Fig. 4

Floor	1st (%)	2nd (%)	3rd (%)	4th (%)
Max relative error	0.45	0.24	0.39	0.82
Min relative error	-0.49	-0.28	-0.50	-0.46
RMS error	7.36e-2	6.00e-2	7.45e-2	6.03e-2

3.4 Structural Parameters Identification Base the Trained BP Neural Network and ARMA Model

When the parameters P_1, P_2, D, E_1 and E_2 of the ARMA model are estimated, the structural parameters other than mass matrix, based on acceleration response records, can be determined with the procedure described in section 2.2 and the matrices identification results are as follow,

$$\overline{C} = \begin{bmatrix} 41.41 & -2.90 & -3.20 & -3.62 \\ -2.59 & 26.36 & -3.85 & 0.42 \\ -13.53 & 9.77 & 53.44 & -42.63 \\ -15.84 & -8.43 & -16.30 & 17.40 \end{bmatrix} N.s/m, \tag{17}$$

$$\overline{K} = \begin{bmatrix} 3.074 & -1.871 & 0.283 & 0.021 \\ -1.865 & 3.564 & -1.905 & 0.100 \\ 0.042 & -1.918 & 3.209 & -1.534 \\ 0.003 & 0.074 & -1.584 & 1.564 \end{bmatrix} \times 10^5 N/m. \tag{18}$$

In order to evaluate the parameter identification results, a free vibration test on the model structure is carried out to extract the global system frequencies from the acceleration measurements. From the Fourier Transform of the acceleration responses of free vibration of the model structure, the natural frequencies are 5.79Hz, 17.12Hz, 26.96Hz, and 36.37Hz as shown in Figure 5. Based on the identified stiffness matrix and the mass matrix from the proposed methodology, the natural frequency of the structure is 5.75Hz, 17.06Hz, 26.89Hz and 35.33Hz, respectively, which is close to them from the free vibration test. It can be seen that the proposed method can identify structural matrices with acceptable accuracy.
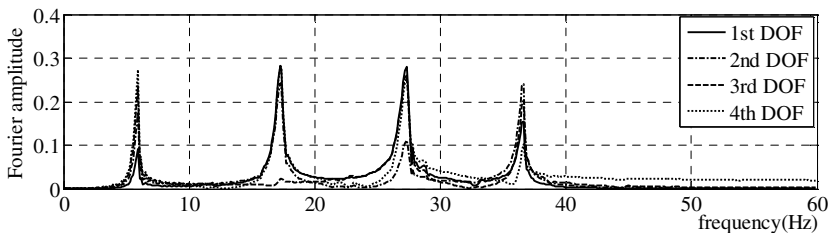

Fig. 5. Fourier Transform of acceleration responses

4 Concluding Remarks

In this paper, a novel structural parameters identification method for linear MDOF structural dynamic system was presented and the efficacy of the proposed strategy was validated experimentally for a multi-storey shear model structure under impact loading. The results show that the proposed methodology by matching ARMA and neural network model is a new efficient way for structural parameter identification with acceptable accuracy, only using structural response and excitation measurement time series.

Acknowledgement

References

1. Wu, Z.S., Xu, B., Harada, T.: Review on Structural Health Monitoring for Infrastructure. Journal of Applied Mechanics, JSCE 6, 1043–1054 (2003)
2. Doebling, S.W., Farrar, C.R., Prime, M.B.: A Summary Review of Vibration-based Damage Identification Methods. Shock and Vibration Digest. 30(2), 91–105 (1998)
3. Yang, S.Z., Wu, Y., Xuan, J.P.: Time Series Analysis in Engineering Application (in Chinese), pp. 175–275. Huazhong University of Science and Technology Press, Wuhan (2007)
4. Lee, C.G., Yun, C.B.: Parameter Identification of Linear Structural Dynamic Systems. Computers and Structures 40(6), 1475–1487 (1990)
5. Chang, C.C., Roschke, P.: Neural Network Modeling of a Magnetorheological Damper. Journal of Intelligent Material Systems and structures 9, 755–764 (1998)
6. Zapateiro, M., Luo, N., Taylor, E., Dyke, S.J.: Modeling and Identification of a Class of MR Fluid Foam Dampers. In: III Eccomas Thematic Conference on Smart Structures and Materials. Gdansk, Poland (2007)
7. Nakamura, M., Masri, S.F., Chassiakos, A.G., Caughey, T.K.: A Method for Non-parametric Damage Detection through the Use of Neural Networks. Earthquake Engineering and Structural Dynamics 27, 997–1010 (1998)
8. Masri, S.F., Smyth, A.W., Chassiakos, A.G., Caughey, T.K., Hunter, N.F.: Application of Neural Networks for Detection of Changes in Nonlinear Systems. Journal of Engineering Mechanics, 666–676 (2000)
9. Masri, S.F., Chassiakos, A.G., Caughey, T.K.: Identification of Nonlinear Dynamic Systems Using Neural Networks. J. Appl. Mech. Trans. ASME 60, 123–133 (1993)
10. Yun, C.B., Bahng, E.Y.: Substructural Identification Using Neural Networks. Computers and Structures 77(2000), 41–52 (2000)

11. Xu, B., Wu, Z.S., Chen, G., Yokoyama, K.: Direct Identification of Structural Parameters from Dynamic Responses with Neural Networks. Engineering Applications of Artificial Intelligence 17(8), 931–943 (2004)
12. Xu, B., Wu, Z.S., Yokoyama, K., Harada, T., Chen, G.: A Soft Post-earthquake Damage Identification Methodology Using Vibration Time Series. Smart materials and Structures 14, S116–S124 (2005)
13. Xu, B., Chen, G., Wu, Z.S.: Parametric Identification for a Truss Structure Using Axial Strain. Computer-Aided Civil and Infrastructure Engineering 22(3), 210–222 (2007)
14. Xu, B., Wu, Z.S.: Decentralized and Localized Identifications of Large-scale Structures. Computer Assisted Mechanics and Engineering Sciences 14(2), 361–378 (2007)

Solar Radiation Forecasting Using Ad-Hoc Time Series Preprocessing and Neural Networks[*]

Christophe Paoli[1], Cyril Voyant[1,2], Marc Muselli[1], and Marie-Laure Nivet[1]

[1] University of Corsica, CNRS UMR SPE 6134, (Vignola, Rte des Sanguinaires, 20000 Ajaccio, Campus Grimaldi, 20250 Corte), France
[2] Hospital of Castelluccio, Radiotherapy Unit, BP 85, 20177 Ajaccio, France
{christophe.paoli,cyril.voyant,marc.muselli,
marie-laure.nivet}@univ-corse.fr

Abstract. In this paper, we present an application of neural networks in the renewable energy domain. We have developed a methodology for the daily prediction of global solar radiation on a horizontal surface. We use an ad-hoc time series preprocessing and a Multi-Layer Perceptron (MLP) in order to predict solar radiation at daily horizon. First results are promising with nRMSE < 21% and RMSE < 998 Wh/m². Our optimized MLP presents prediction similar to or even better than conventional methods. Moreover we found that our data preprocessing approach can reduce significantly forecasting errors.

Keywords: Time Series, Preprocessing, Seasonality, Multi-Layer Perceptron.

1 Introduction

Artificial intelligence techniques are becoming more and more popular in the renewable energy domain [1], [2] and particularly for the prediction of meteorological data such as solar radiation [3], [4] [5] [6]. Thereby many research works have shown the ability of Artificial Neural Networks (ANNs) to predict times series of meteorological data. In this study and according to electricity suppliers, we focus on the prediction of global solar irradiation on a horizontal plane for daily horizon. In this way, we have investigated time series forecasting which is a challenge in many fields. Because it has made tremendous progress in the past twenty years in terms of theory, algorithms and applications, we have chosen to study ANNs. Moreover, if we compare to conventional algorithms based on linear models, ANNs offer an attractive alternative by providing nonlinear parametric models. Through the proposed study, we will particularly look at the Multi-Layer Perceptron (MLP) network which has been the most used of ANN architectures in the renewable energy domain [1], [2]. The originality of our study is to add an ad-hoc time series preprocessing step before using neural networks. Indeed, as seen in [7] a data preprocessing including deseasonalization and detrending can improve ANN forecasting performance. As global solar radiation has a deterministic part, we want to take into account this specificity.

The paper is organized as follow: the section 2 describes the physical phenomena we want to predict and introduces our ad-hoc time series preprocessing. Section 3

---

presents the neural network architecture we designed. Results are presented and discussed in section 4 where several conventional methods for estimation and modeling of the meteorological data are compared with our methodology. Section 5 concludes and suggests perspectives.

2   Data Analysis and Preprocessing

There are two approaches that allow quantifying solar radiation: the "physical modeling" based on physical processes occurring in the atmosphere and influencing solar radiation [8], and the "statistic solar climatology" mainly based on time series analysis [8]. As already said we have chosen to combine these two methods in a gray box approach to improve the quality of prediction. In this work, we have used the physical phenomena in an attempt to overcome the seasonality of the resource.

2.1   Meteorological Data

Our study proposes to analyze the radiation time series (Wh.m-2) measured at the meteorological station of Ajaccio (METEO FRANCE, Corsica, France, 41°55'N, 8°44'E). The data representing the global solar radiation were measured on a daily basis from January 1971 to December 1989. Thus we have a X_t time series to forecast for time t+1; that is at horizon 1. To achieve this, we choose to use a gray box (or semi-physical) model where time series prediction and modeling are mixed. For time series prediction only past values are used to forecast the future values at a given horizon. In the case of modeling, the different physical processes involved are taken into account in order to represent the variable however the horizon is. In the remainder of this paper, we choose the following naming convention: X_t designate the time series, and $X_{d,y}$ is the modeling of the variables, where d is the day of the year y. In the next section, an explanation of the physical phenomenon is proposed, and then we describe our time series preprocessing.

2.2   Physical Phenomenon

We can observe in Figure 1 that the global radiation consists of three types of radiation: direct radiation, diffuse radiation and ground-reflected radiation [10]. The ground-reflected radiation does not concern us because we try to predict the radiation on a horizontal surface where the ground reflected radiation does not make sense. For clear sky, ie without cloud cover, global radiation is relatively easy to model because it is primarily due to the distance from the sun sensor [9], [10], [11], [12], [13]. It is not the same, when there are clouds near the detectors. Indeed, these are mostly stochastic phenomena, which depend on the weather site.

The spectral analysis of the global radiation series highlights the high periodicity of the phenomenon (almost 365 days period). As proposed in [7], it appeared wise to make the time series stationary as much as possible. Deseasonalize and detrend a series allows to eliminate seasonal and trend components without changing the other information. In the present study, we have not considered the solar irradiation like a trend process but only like a seasonal process. The choice of the methodology used
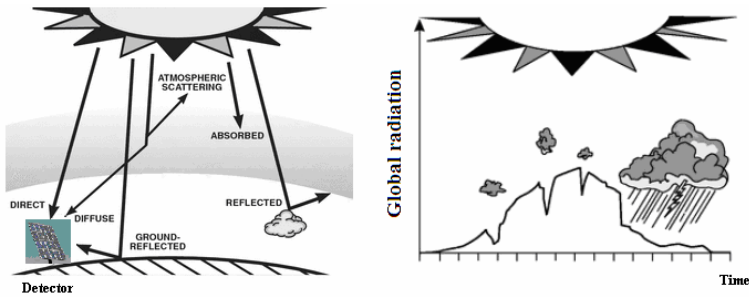
Fig. 1. Origin of the three types of radiation: direct radiation, diffuse radiation and ground-reflected radiation on a detector at ground level (left of the figure). Modification of the global irradiation profile accordingly to clouds cover (right of the figure).

depends on the nature of the seasonality. As in our case the seasonality is very pronounced and repetitive, so very deterministic and not stochastic. Moreover, it is possible to physically quantify the components of our irradiation series. Like we have seen in figure 1, the global solar irradiance on horizontal plane depends on direct and diffuse radiation. This specificity allows us to apprehend the periodicity of the phenomenon, and to reduce the non-stationarity of the series.

In the next section, we propose to stationarize radiation data to overcome the deterministic component which is easily quantifiable. Thus we devote to the prediction of cloud cover included in the global radiation of the site.

2.3   Ad-Hoc Time Series Preprocessing

In fact, by dividing the series by daily extraterrestrial radiation [14], we can quantify the annual periodicity. It is the first step of our stationarization process. In first approximation, it is possible to quantify the deterministic component of global radiation by the extraterrestrial solar radiation alone (H_0). Thus we apply on the original series $X_{d,y}$ (where "d" is the day and "y" is the year) the ratio to trend method. This leads to a new series ($S_{d,y}$), known as series of index clarity:

$$S_t = S_{d,y} = x_{d,y} \quad / \quad H_0^d .$$
(1)

After this step a new rigid seasonality is updated, we can lift it with the use of periodic seasonal factors [14]. This treatment aims to create a new distribution without periodicity. Although this pre treatment tends to stationarize the time series, a test of Fisher shows that seasonality was not optimal. According Bourbonnais [14], after using a ratio to trend method (H_o in this case) to correct rigid seasonalities, we can use a ratio to moving average. This second ratio can be applied when there is no analytical expression of the trend. In our case, we find that H_o led a new seasonality which is difficult to model. That's why we made a moving average ratio to overcome the seasonality. In the case of flexible seasonality, ie random in amplitude or period, the filtering techniques by successive moving averages are recommended.

$$y_{d,y} = S_{d,y} \quad / \quad \frac{1}{2.m+1} . \sum_{i=-m}^{m} S_{d+i,y} .$$
(2)

In our case as 2m+1 = 365 days, we obtain m = 182. To complete the process, then we use the 365 seasonal factors (y$_d$). These are in fact coefficients which allow to overcome rigid seasonality by a moving average ratio described above. In order to not distort the series, we have considered that the total sum of the components of the series is the same before and after the report (final seasonal factors y$_d$* of the equation 5). The transition coefficients (N = 18, the number of years of history) and the average coefficients of the regular 365 days are given by the equations 3 and 4. A new series seasonally adjusted that represents only the stochastic component of global radiation is given by the equation 6.

$$y_d = \frac{1}{N}\left(\sum_{y=1}^{N} y_{d,y}\right).$$ (3)

$$\overline{y}_d = \frac{1}{365}\left(\sum_{y=1}^{365} y_d\right).$$ (4)

$$y_d^* = y_d \ / \ \overline{y_d}.$$ (5)

$$S_{d,y}^{corr} = S_{d,y} \ / \ y_d^*.$$ (6)

In the next section, we present the RNA architecture used.

3   Neural Network Architecture and Design

The search for the ideal network structure is a complex and crucial task. We have adopted a feed forward Multi-Layer Perceptron (MLP), which is the most commonly used in the renewable energy domain [1], [2]. In order to determine the best network configuration, we have tried to study all the parameters available in this network architecture. To perform this optimization, we have considered the practice hypothesis that parameters are orthogonal. We have optimized parameters by considering each other constants.

We used the Matlab® software and its neural network toolbox to implement our network. The Matlab® training and testing data sets were set respectively to 80% and 20%. As a result of this iterative process, the selected network (see figure 2) has three neuron layers: input, hidden and output layer. There was no significant difference in the use of 1, 2 and 3 hidden layer architectures. One hidden layer was used in order to minimize the complexity of the proposed ANN model. We tried several input layer configurations: 1, 2, 4, 8 and 15 variables. Best results were obtained with 8 inputs which received as input the endogenous entries S$_{t-1}$,.. , S$_{t-8}$ normalized on {0,1}. The same process was used to fix the number of hidden layer neurons. We found that 3 neurons were sufficient. Finally, we have one neuron on the output layer $\hat{s}_t$. Concerning the transfer functions the best results were obtained with the Gaussian (hidden layer) and linear (output layer) function. Regarding the training algorithm, many experiences have enabled us to choose the Levenberg–Marquardt optimization (second-order algorithm) with 5000 epochs and a decrease factor (mu) of 0.5. Default values have been use for all other parameters. The early stopping technique was set to the maximum validation failure (the parameter max_fail = 5).
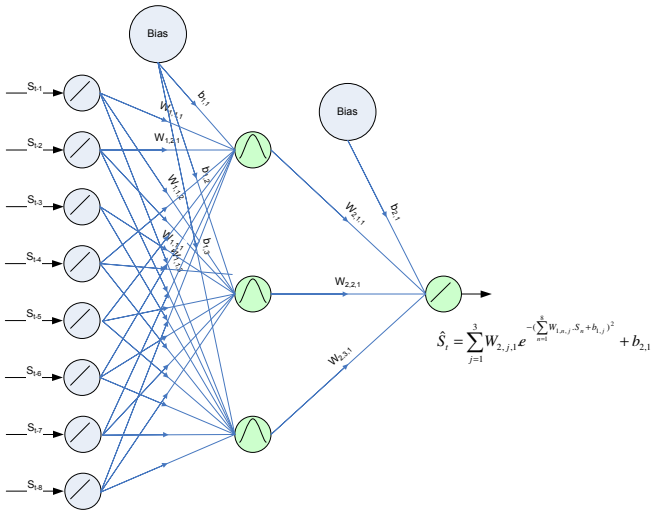
Fig. 2. Architecture of the optimized MLP

The learning has concerned the years 1971 to 1987 and the performance function was mean square error – MSE.

4   Results and Discussions

Figure 3 summarizes the protocol that has allowed us to conduct our experiences and validate our approach. A first treatment (step 1 of the figure 3) allows to clean the
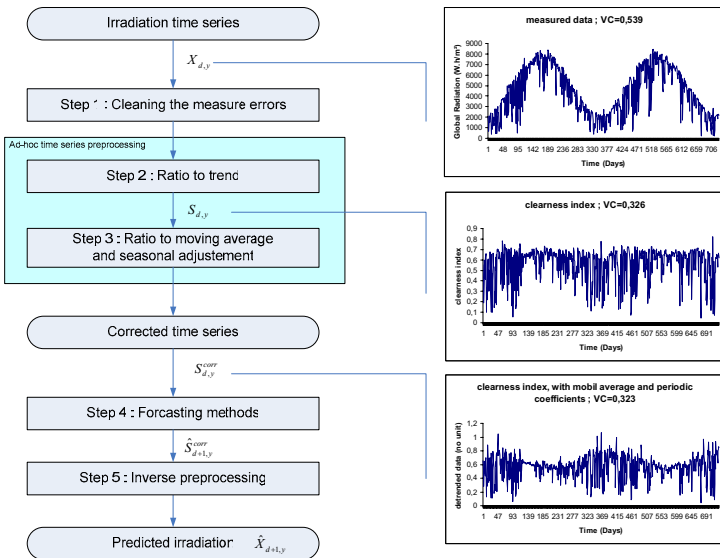

Fig. 3. Summarize of the protocol followed to obtain the predicted irradiation

series of atypical points. We have replaced them by the average over the 19 years for the hours and the days corresponding to the problems.

Steps 2 and 3 have been described in the previous section and lead to a series corrected. In the step 4 we compare classical forecasting methods outlined in the next section with our optimized MLP. Finally step 5 allows to reverse the preprocessing treatment and to obtain the prediction of global irradiation.

4.1 The Classical Forecasting Methods

In order to measure the effectiveness of our approach, we have decided to compare it with the following classical forecasting methods.

The ARIMA techniques [15], [16] are reference estimators in the prediction of global radiation field. It is a stochastic process coupling autoregressive component (AR) to a moving average component (MA). After several experiments, we have obtained and decided to use an ARMA (2,2). Bayesian inference [17], [18] is another classical technique. In this method evidences or observations are used to update or to newly infer the probability that a hypothesis may be true. We have identified that the prediction was better if we had 50 classes and an order of 3. Some authors have tried to use so-called Markov process [19], specifically the Markov chains, which is a stochastic process. The mean idea of this technique is that the description of the present state fully captures all the information that could influence the future evolution of the process. Future states will be reached through a probabilistic process instead of a deterministic one. In our case, we obtained 50 for the dimension of the transition matrix (number of class) and an order of 3 for the chain (determination of the prediction lag). The k-nearest neighbors algorithm (k-NN) [20] is a method for classifying objects based on closest training examples in the feature space. k-NN is a type of instance-based learning, or lazy learning where the function is only approximated locally and all computation is deferred until classification. It can also be used for regression. Unlike previous ones this tool does not use a learning base. The method consists in looking into the history of the series for the case the most resembling to the present case. In our study we choose a k equal to 10. The following section presents the results obtained.

4.2 Results

To determine whether our network was really interesting in terms of daily prediction of solar radiation, we compared its performances with the forecasting results obtained with a naive predictor (average over 18 years of the day considered), order 3 Markov chains, order 3 Bayesian inferences, an order 10 k-NN, an order 8 AR without preprocessing, an ARMA(2,2) with preprocessing. The predicted results for each combination were compared statistically using three parameters: the Root Mean Square Error (RMSE), the normalized RMSE (nRMSE), and the Mean Bias Error (MBE):

$$RMSE = \sqrt{\frac{1}{N}\sum_{i+1}^{N}(C_i - M_i)^2} \cdot$$

(6)

$$nRMSE = \sqrt{\frac{1}{N}\sum_{i+1}^{N}(C_i - M_i)^2} \quad / \quad \sqrt{\frac{1}{N}\sum_{i=1}^{N}(M_i)^2}. \tag{7}$$

$$MBE = \frac{1}{N}\sum_{i+1}^{N}(C_i - M_i). \tag{8}$$

Table 1 presents results we have obtained in the case of an annual error for daily prediction of global solar radiation.

Table 1. Annual error for all prediction methods, years 1988 and 1989, 8 simulations

Prediction methods	nRMSE	Confidence interval
Naïve predictor	26 %	0%
Markov Chain (order 3)	25,11 %	0%
Bayes (order 3)	25,16 %	0%
k-NN (order 10)	25,20 %	0%
AR(8) without preprocessing	21,18 %	0,2%
ARMA(2,2) with preprocessing	20,31 %	<0,1%
ANN[8,3,1] without preprocessing	20,97 %	0,15%
ANN[8,3,1] with preprocessing	20,17 %	0,1%

We highlight that the predictors other than ARMA and ANN give the same results, slightly higher than those obtained with a naive predictor. Even without preprocessing, ARMA and ANN are the best predictors. The preprocessing improves the quality of prediction and allows access to the 20% error. C_i and M_i are respectively the i_{th} calculated and measured values and N is the total number of observation. Table 2 details in the MLP case, the annual prediction errors obtained for the years 1988 and 1989.

Table 2. Annual prediction error for the years 1988 and 1989 with our MLP

	Arithmetic Mean	95% confidence interval
nRMSE	20,2 %	0,1 %
RMSE	997,971	6,333
MBE	-104,239	28,794
R^2	0,801	0,002
Monthly average error	3,9 %	0,4 %

The confidence interval is calculated after 10 simulations. Given the small size of the confidence intervals, we can say that there are very few local minimums. The monthly average error represents the error for the value of the irradiation. We obtain that the combination of the prediction of global radiation received after 1 month is different from an average of 4% of the aggregate measured. The negative MBE means that we underestimate the solar potential on averaged over the year. Since we have an atypical day of low irradiation then there is a tendency to overestimate. The determination coefficient R^2 is

greater than 0,8. Figure 4 shows the errors of prediction and distinguishes the seasons for the years 1988 and 1989. As we can see best results; i.e. less important error, in term of forecast are obtained in summer. These results can be used for example by energy managers who need to avoid using hydraulic power plants in dry season.
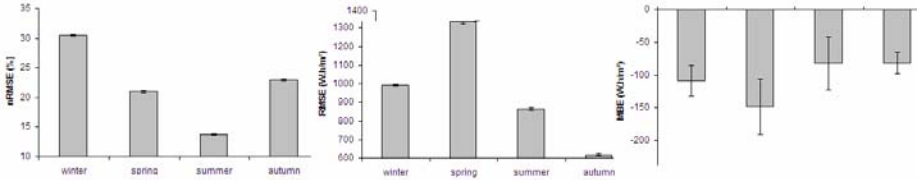

Fig. 4. Seasonal errors for the daily prediction of the years 1988 and 1989 (mean with 95% confidence interval)

There should be a compromise between RMSE and nRMSE. The nRMSE are useful for comparison and optimization. But for the interpretation of energy, we must look at the RMSE. The spring season is the most difficult to predict. The absolute error is consistent. However, we find that in summer the error does not exceed 900 Wh / m² , while the irradiation is important. MBE are found negative, which indicates an underestimation. The MBE is not significantly different from one season to another. Thus we will always have the same prediction error, whatever the season. Finally, Figure 5 compares the real data of solar radiation with the results obtained with our MLP with preprocessing. The error of prediction is also drawn. The increase in errors at the beginning of the cycle corresponds to the spring when the cloud disturbances are very important. We can see it is very difficult to predict the radiation cause of very noisy variable. Figure 6 shows the correlation between experimental and simulated global irradiation.

We systematically overestimate the days when the irradiation is minimal (winter). The points that lie at the very top of the line y = x shows that it is very difficult to predict the days when the irradiation had to be theoretically important. We would undoubtedly have improved the results optimizing an ANN by season, but it would complicate the procedure, and tend to decrease the robustness of the procedure. The next section concludes this paper and suggests prospects.
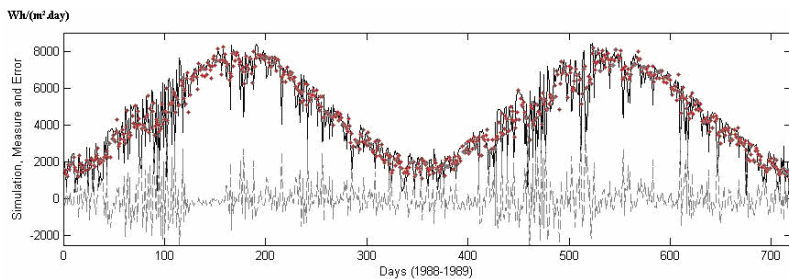

Fig. 5. Real data of solar radiation and results obtained with our MLP with preprocessing and error of prediction. Dashed line is the error; solid line is the real data of solar radiation and red points are the prediction.
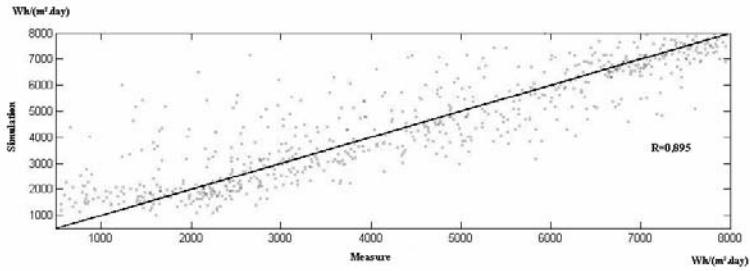
Fig. 6. Correlation between experimental and simulated global irradiation

5 Conclusion and Perspectives

This paper has developed an ANN predictor approach to determine global irradiation at daily horizon in order to help electrical managers. We have used an ad-hoc time series preprocessing and a time series prediction designed MLP. Although the location was very specific, with the proximity to the sea and the mountain that can greatly affect nebulosity, we have obtained relevant results. Seasonal RMSE are less than 998 Wh/m-2 (nRMSE < 21%). ANN processes presents a great interest compared to classical stochastic predictor like ARIMA. Moreover we found that our data preprocessing approach reduce significantly forecasting errors.

The next step of our work will be to validate our predictor on real photovoltaic system. It was recently installed in our laboratory and we are awaiting data. In the future, it seems important to study shorter time horizons. As matter of fact, electrical managers are also interested to horizons that can range from ½ hour to several hours: from 3 hours to 24 hours. Thus others ANN architecture types have to be studied: recurrent ANNs, adaptative ANNs, etc. An ongoing study will be based on the implementation of exogenous variables on the input neurons like METARs data (pressure gradient, temperature, etc.). Determining the relevant variables could be done by the random probe method [21].

References

1. Mellit, A., Kalogirou, S.A., Hontoria, L., Shaari, S.: Artificial intelligence techniques for sizing photovoltaic systems: A review. Renewable and Sustainable Energy Reviews 13(2), 406–419 (2009)
2. Mellit, A., Kalogirou, S.A., Hontoria, L., Shaari, S.: Artificial intelligence techniques for photovoltaic applications: A review. Progress in Energy and Combustion Science (2008)
3. Mubiru, J.: Predicting total solar irradiation values using artificial neural networks. Renewable Energy 33(10), 2329–2332 (2008)
4. Mubiru, J., Banda, E.J.K.B.: Estimation of monthly average daily global solar irradiation using artificial neural networks. Solar Energy 82(2), 181–187 (2008)
5. Kaligirou, S.A.: Artificial neural networks in renewable energy systems applications: a review. Renewable and sustainable energy review 5, 373–401 (2001)
6. Hocaoglu, F.O., Gerek, O.N., Kurban, M.: Hourly solar forecasting using optimal coefficient 2-D linear filter and feed-forward neural networks. Solar energy (2008)

7. Zhang, G.P., Qi, M.: Neural network forecasting for seasonal and trend time series. European journal of operational research 160, 501–514 (2005)
8. Badescu, V.: Modelling Solar radiation at the earth surface (2008)
9. Reindl, D.T., Beckman, W.A., Duffie, J.A.: Evaluation of hourly tilted surface radiation models. Solar Energy 45, 9–17 (1990)
10. Liu, B.Y.H., Jordan, R.C.: Daily insolation on surfaces tilted towards the equator. Trans. SHRAE 67, 526–541 (1962)
11. Hay, J.E., Davies, J.A.: Calculation of the solar radiation incident on an inclined surface. In: Proc. first Canadian solar radiation workshop, pp. 59–72 (1980)
12. Perez, R., Ineichen, P., Seals, R.: Modelling daylight availability and irradiance components rom direct and global irradiance. Solar Energy 44, 271–279 (1990)
13. Ineichen, P., Guisan, O., Perez, R.: Ground-reflected radiation and albedo. Solar Energy 44, 207–214 (1990)
14. Bourbonnais, R., Terraza, M.: Analyse des séries temporelles (2004)
15. Hamilton, J.D.: Times series analysis (1994) ISBN 0-691-04289-6
16. Poggi, P., Muselli, M., Notton, G., Cristofari, C., Louche, A.: Forecasting and simulating wind speed in Corsica by using an autoregressive model. Energy Conversion and Management 44(20), 3177–3196 (2003)
17. Diday, E., Lemaire, L., Pouget, J., Testu, F.: Éléments d'analyse de données, Dunod (1982)
18. Celeux, G., Nakache, J.P.: Analyse discriminante sur variables qualitatives (1994)
19. Muselli, M., Poggi, P., Notton, G., Louche, A.: First Order Markov Chain Model for Generating Synthetic 'Typical Days' Series of Global Irradiation in Order to Design PV Stand Alone Systems. Energy Conversion and Management (2001)
20. Mohammed, S., Donald, H.B.: Simulating climate change scenarios using an improved K-nearest neighbor model (2006)
21. Dreyfus, G.: Random probes for variable selection, Multiple Simultaneous Hypothesis Testing, Paris (2007)

Using Non-extensive Entropy for Text Classification

Lin Fu and Yuexian Hou

School of Computer Science and Technology, Tianjin University, China
anlan187@gmail.com, yxhou@tju.edu.cn

Abstract. This paper proposes the use of non-extensive entropy for text classification. Non-extensive entropy technique is used for text classification by estimating the conditional distribution of the class variable given the document. The underlying principle of non-extensive entropy is that without external knowledge, one should prefer distributions that are uniform. This paper proposes two models for text classification based on maximum entropy principle. The first model extends Shannon entropy into non-extensive entropy to simplify the form of classifier, the other one introduces high-level constraints into non-extensive model to impose constraints on the pairs of entities. Model with high-level constraints constructs relations between word pairs which builds semantic constraints, for the sake of advancing accuracy of text classification. Experiments on the 20 newsgroup set demonstrate the advantage of non-extensive model and non-extensive model with high-level constraints.

Keywords: Non-extensive entropy, High-level constraints, Text classification.

1 Introduction

With the rapid development of online resources, there are more and more information on the website. Consequentially, auto classification techniques are required to discriminate useful information against redundant urgently. For text classification, there are a variety of mature techniques which have been demonstrated reasonable performances, such as: Naïve Bayes [Lewis, 1998; McCallum and Nigam, 1998; Sahami, 1996], k-nearest neighbor [Yang, 1999], support vector machine [Joachims, 1998; Dumais *et al.*, 1998], boosting [Schapire and Singer, 1999], rule learning algorithms [Cohen and Singer, 1996; Slattery and Craven, 1998] and maximum entropy [Berger *et al.*, 1996; Pietra *et al.*, 1997].

Maximum entropy is a probability distribution estimation technique for estimating probability distribution from data. The motivating idea is that one should perform the most uniform models that also satisfy any given constraints. Recently, there have been many studies of maximum entropy technique for natural language processing, such as language modeling [Chen and Rosenfeld, 1999; Rosenfeld, 1994], text segmentation [Beeferman *et al.*, 1999], part-of-speech tagging [Ratnaparkhi, 1996] and prepositional phrase attachment [Ratnaparkhi *et*

al,. 1994]. Results show that maximum entropy is a promising technique for natural language processing which warrants further investigation.

This paper proposes two models for text classification which are built on the principle of maximum entropy. One is non-extensive entropy model and the other is its expanded form, non-extensive entropy model with high-level constraints. In text classification, non-extensive entropy is used to estimate the conditional distribution of the class variable given the document. Labeled training data and unlabeled test data are required, training data is used to derive a set of constraints for the model that characterize the class-specific exceptions for the distribution, test data is used to evaluate the accuracy of the obtained classifier. Constraints in text classification are represented as expected value of feature functions. When learning the non-extensive entropy model, one important issue that has to obtain is conditional distribution which is consistent with the given constraints from training data. During this process, Lagrange multiplier method is used to estimate the parameters and then a classifier is generated. Further analysis illuminates that the features in non-extensive model only construct word-class combination which can't express constraints completely. Motivated by this, we investigate in this paper how to inject high-level constraints which stand for cooccurrence relations between word pairs into the first proposed model. By increasing high-level constraints, the model establishes semantic relation constraints additionally. Relations between word pairs are of substantial assistance to enhance the accuracy of text classification and achieve more success.

This paper is organized as follows. Section 2 briefly introduces related work. Non-extensive entropy and its application in text classification are discussed in section 3. Section 4 details an improved model based on non-extensive entropy in text classification by inducting high-level constraints. Experiments and evaluations are presented in section 5 and section 6 concludes the paper.

2   Related Work

The maximum entropy concept has a long history, its constrained form has been widely applied in various contexts. It is used to construct a stochastic model that accurately represents the behavior of a random process and performs the most uniform distribution while satisfy any given constraints [1].

In text classification, maximum entropy is used by estimating the conditional distribution of the class variable given the document. Training data is used to set constraints on the conditional distribution. Each constraint expresses a characteristic of the training data that should also be present in the learned distribution. We let any real-valued function of the document and the class is a feature, $f_i(c, d)$. Maximum entropy allows us to restrict the model distribution to have the same expected value for this feature as seen in the training data D. Thus, we stipulate that the learned conditional distribution $p(c|d)$ must have the property:

$$\sum_{c,d} \widetilde{p}(c, d) f_i(c, d) = \sum_{c,d} \widetilde{p}(d) p(c|d) f_i(c, d) \,. \tag{1}$$

Thus, when using maximum entropy, the first step is to identify a set of feature functions that will be useful for classification. Then, for each feature, measure its expected value over the training data and take this to be a constraint for the model distribution.

When constraints are estimated in this fashion, it is guaranteed that a unique exponential distribution exists that has maximum entropy[1]:

$$p(c|d) = \frac{1}{Z(d)} \exp\left(\sum_i \lambda_i f_i(c,d)\right), \tag{2}$$

where each λ_i is a parameter to be estimated and $Z(d)$ is simply the normalizing factor to ensure a proper probability:

$$Z(d) = \sum_c \exp\left(\sum_i \lambda_i f_i(c,d)\right). \tag{3}$$

After obtaining the exponential form of the classifier, parameters λ are to be calculated. Improved Iterative Scaling (IIS), which performs hill climbing parameter log likelihood space, is chosen (Details in [2] and [3]).

3   Non-extensive Entropy Model

In order to simplify the form of classifier, we substitute non-extensive entropy[1] for Shannon entropy in maximum entropy model. Non-extensive entropy S_q is the generalization of Boltzmann-Gibbs entropy $S_{BG}(S_1 = S_{BG})$, it is based on the notion of q-exponential [8]. For the sake of analytical and computing simplicity, we only consider the non-extensive entropy with $q = 2$ in this paper. Under this assumption, we can rewrite the expression of non-extensive entropy as $S_2 = -\sum_i p_i^2$.

3.1   Constraints and Features

As the notion of features and constraints described above, each constraint restricts the model distribution to have the same expected value as seen in training data. It is expressed as:

$$\sum_{c,d} \widetilde{p}(c,d) f_i(c,d) = \sum_{c,d} \widetilde{p}(d) p(c|d) f_i(c,d), \tag{4}$$

where each $f_i(c,d)$ is a feature. All the feature functions f_i which determine statistics in modeling the process are are gathered as set F.

In text classification, we choose word counts as our features. More specifically, in this paper for each word-class combination, we instantiate a feature as:

$$f_{w,c'}(c,d) = \begin{cases} \frac{N(d,w)}{N(d)} & \text{if } c' = c \\ smooth & \text{Otherwise}, \end{cases} \tag{5}$$

---

[1] The non-extensive entropic form is $S_q = k \frac{\sum_i p_i^q - 1}{1-q}$ $(\sum_i p_i = 1)$.

where $N(d, w)$ is the number of times word w occurs in document d, $N(d)$ is the number of words in d, $smooth$ is the parameter of absolute discounting smooth method [7]. In this paper, we assume $smooth = 0.0001$.

With this representation, if a word occurs often in one class, we would expect the weight for that word-class pair to be higher than for the word paired with other classes. Note that we utilize the scaled counts as features in order to avoid hurts for long documents where repeated words are given too strong a weight.

3.2 Model

The mathematical measure of the uniformity of a conditional distribution $p(c|d)$ is provided by the non-extensive conditional entropy:

$$S_2(C|D) = \sum_d \widetilde{p}(d) S_2(C|d) = \sum_d \widetilde{p}(d)(1 - \sum_c p^2(c|d)) = 1 - \sum_{c,d} \widetilde{p}(d) p^2(c|d) \,. \tag{6}$$

Thus, non-extensive entropy model for text classification is established as: Model 1:

$$\arg \min_{p(c|d)} \sum_{c,d} \widetilde{p}(d) p^2(c|d)$$

$$s.t. \quad \sum_{c,d} \widetilde{p}(c,d) f_i(c,d) = \sum_{c,d} \widetilde{p}(d) p(c|d) f_i(c,d), \forall f \in F \tag{7}$$

$$\sum_c p(c|d) = 1, \forall d \in D$$

$$p(c|d) \geq 0 \,.$$

3.3 Parametric Form

Through Lagrange multiplier method, we obtain the form of conditional distribution:

$$p(c|d) = \frac{1}{Z(d)} \left(\sum_i \lambda_i f_i(c,d) \right), \tag{8}$$

where each $f_i(c,d)$ is a feature, λ_i is a parameter to be estimated and $Z(d)$ is a normalizing constant determined by the requirement that $\sum_c p(c|d) = 1$:

$$Z(d) = \sum_c \sum_i \lambda_i f_i(c,d) \,. \tag{9}$$

3.4 Lagrange Multiplier Method

Because of nonexponential form of conditional distribution in non-extensive entropy model, the matured IIS method is not appropriate for calculating parameters any longer. Motivated by this, we investigate in this paper how to obtain

model distribution in non-extensive entropy model. Lagrange multiplier method is selected instead of IIS. Basic procedures of parameters estimating are briefly outlined below:

Input: A collection D of labeled data and a set of feature functions $f_i(c, d)$.

Lagrange function $F(S_2, \lambda)$ is constructed as[2] :

$$F(S_2, \lambda) = \sum_{c,d} \widetilde{p}(d) p^2(c|d)$$
$$- \sum_i \lambda_i (\sum_{c,d} \widetilde{p}(d) p(c|d) f_i(c, d) - \sum_{c,d} \widetilde{p}(c, d) f_i(c, d)) \,. \tag{10}$$

By differentiating with respect to variable λ_i and $p(c|d)$ respectively, the Lagrange function is transformed into two sets of formulae. Let both of them be zero, two series of equations are received.

$$\frac{\partial F}{\partial p(c|d)} = 2\widetilde{p}(d) p(c|d) - \sum_i \widetilde{p}(d) \lambda_i f_i(c, d) = 0 \,. \tag{11}$$

$$\frac{\partial F}{\partial \lambda_i} = \sum_{c,d} \widetilde{p}(d) p(c|d) f_i(c, d) - \sum_{c,d} \widetilde{p}(c, d) f_i(c, d) = 0 \,. \tag{12}$$

From Eq. 11 we can get the expression $p(c|d) = \sum_i \lambda_i f_i(c, d)/2$. Incorporating it into Eq. 12, the set of equations only contain variables λ, so parameters λ can be derived by solving these equations. When we gain the parameters λ, a text classifier is established completely.

Output: A text classifier that takes an unlabeled document and predicts a class label.

4   Non-extensive Entropy Model with High-Level Constraints

In this section we present an improved form of non-extensive entropy model which is named non-extensive model with high-level constraints, the model is expected to improve the accuracy of text classification by increasing constraints. In this paper, we assume high-level constraints are only second-order dependence between word pairs which express semantic relations in a document. Intuitively, accuracy should be promoted to some extent through establishing high-level constraints based on the previous model.

Non-extensive entropy model constructs constraints with word-class combination, it simply considers contribution created by single words. A closer analysis of text classification indicates that relations between word pairs affect document label in a big way. For example, a document containing the word "apple"

---

[2] Ignore the constraints $\sum_c p(c|d) = 1$ in Lagrange function by using normalization factor in form of model distribution to ensure a proper probability.

can be labeled either "Computer" or "Life" or other related category by setting constraints with such a word in non-extensive entropy model. But if the information between word pairs cooccurrence are picked up additionally, the document will be labeled more accurately. Supposing word "banana" or "notebook" is existed together with word "apple", the document will be categorized into "Life" or "Computer" respectively. This example makes clear that by introducing cooccurrence relation between useful word pairs into the previous linear constraints model, accuracy can be enhanced considerably because semantic relations are taken into consideration. Furthermore, we hypothesize the cooccurrence between word pairs is only second-order dependence.

4.1   Constraints and Features

Keeping all existing constraints constant, we introduce word pairs cooccurrence constraints into the non-extensive entropy model for text classification. The form of word pairs cooccurrence constraints is established as:

$$\sum_{c,d} \widetilde{p}(c,d)R_j(c,d) = \sum_{c,d} \widetilde{p}(d)p(c|d)R_j(c,d)\,, \tag{13}$$

where $R_j(c,d)$ is a feature for word pairs-class combination. We notate the set of all feature functions as R.

In this formula, $R_j(c,d)$ is expressed by:

$$R_{w,c'}(c,d) = \begin{cases} \frac{N(d,w)}{N(d)} & \text{if } c' = c \\ smooth & \text{Otherwise}\,, \end{cases} \tag{14}$$

where w is the combination of two words w_i and w_j in document d, $N(d,w)$ is the two words' cooccurrence times in document d, $N(d)$ is the number of words in d and $smooth$ is the smooth parameter. We assume $smooth = 0.0001$.

4.2   Model

Appending cooccurrence constraints between word pairs described above, the model is defined by combining the linear constraints and high-level constraints together as:

Model 2:

$$\underset{p(c|d)}{\arg\min} \sum_{c,d} \widetilde{p}(d)p^2(c|d)$$

$$s.t. \quad \sum_{c,d} \widetilde{p}(c,d)f_i(c,d) = \sum_{c,d} \widetilde{p}(d)p(c|d)f_i(c,d), \forall f_i \in F$$

$$\sum_{c,d} \widetilde{p}(c,d)R_j(c,d) = \sum_{c,d} \widetilde{p}(d)p(c|d)R_j(c,d), \forall R_j \in R \tag{15}$$

$$\sum_{c} p(c|d) = 1, \forall d \in D$$

$$p(c|d) \geq 0\,.$$

4.3 Parametric Form

With Lagrange multiplier method, the distribution form is obtained as:

$$p(c|d) = \frac{1}{Z(d)}(\sum_i \lambda_i f_i(c,d) + \sum_j \delta_j R_j(c,d)), \qquad (16)$$

where each $f_i(c,d)$ is a feature for word-class combination and $R_j(c,d)$ is a feature for word pairs-class combination, λ and δ are parameters to be estimated and $Z(d)$ is simply the normalizing factor to ensure a proper probability:

$$Z(d) = \sum_c (\sum_i \lambda_i f_i(c,d) + \sum_j \delta_j R_j(c,d)). \qquad (17)$$

4.4 Lagrange Multiplier Method

As described above, non-extensive model with high-level constraints employs Lagrange multiplier method as parameter calculating algorithm. Different with the previous model, not only λ for word-class combination but also δ for word pairs-class combination are parameters to be estimated. Solution procedure is elided, consult details in section 3.4.

5 Results

This section provides some preliminary empirical evidence that non-extensive entropy model and its improved form are competitive and valuable algorithms in text classification. Results show that it is significant to take cooccurrence relations between word pairs into account.

5.1 Data Set

The experiments are set up on the 20_newsgroups data set[3]. It includes about 19,997 articles evenly divided among 20 UseNet discussion groups. Nearly one thousand messages from each of the twenty newsgroups were chosen at random and partitioned by newsgroup name.

5.2 Text Preprocessing

Remove Stop Words. Stop words, sometimes known as noise words, make little sense to text classification and should be ignored in preprocessing.

Feature Selection. Vocabulary selection is performed by selecting top words according to TF-IDF[4] values, top word pairs are chosen by cooccurrence frequency between two words which own higher TF-IDF value. Numbers of top words and word pairs can be selected by threshold values respectively.

---

[3] Available on the Internet. See http://www.cs.cmu.edu/~TextLearning
[4] TF-IDF : Term Frequency - Inverse Document Frequency.

5.3   Experiments

In order to validate the advantage of our proposed models, we compare both the models to maximum entropy model, so accuracy is adopted to evaluate the performance of text classification. Our experiments are performed with ten trials of randomly selected train-test splits. In each trial, 90% of the documents are held-out for training and the remaining are used for testing.

Two sets of experiments are set up to evaluate the performance of two models proposed in this paper. In the first set of experiments, non-extensive entropy model is executed across different vocabulary sizes. Additionally, compare the results to non-extensive model without absolute discounting smooth technique[5]. In the second set of experiments, we intend to further study the effect of model introduced in chapter 4, an appended set of high-level constraints expressed by word pairs' cooccurrence are appended. Empirical results are compared to the model without absolute discounting smooth technique.

Table 1. Classification accuracy of non-extensive entropy model, compared to the model without absolute discounting smooth technique

Vocabulary Size	20000	25000	30000	35000
Model1	85.47%	89.24%	90.85%	88.63%
Model1 without smooth	84.64%	88.11%	89.61%	87.12%

Table 1 presents classification accuracy of non-extensive entropy model under different vocabulary sizes. Results are compared to non-extensive model without absolute discounting smooth technique. From the results, we observe that:

1) Accuracy of non-extensive entropy classifier reduces slightly after a certain amount of vocabulary sizes. This shows that feature selection is an important factor for non-extensive entropy.

2) Between our non-extensive entropy models, the model with smooth technique achieves higher accuracy. With deeper analysis, it clearly states the importance of avoiding sparse feature in text classification.

Table 2. Classification accuracy of non-extensive entropy model with high-level constraints, compared to the model without absolute discounting smooth technique

Word Pairs Size	4000	4500	5000	5500
Model2	86.74%	89.91%	91.72%	89.15%
Model2 without smooth	85.31%	88.73%	90.23%	87.54%

With the optimal vocabulary size (30000) chosen in the first set of experiments, table 2 reports the performance of model 2 and its simply form without smooth technique across different word pairs sizes. Conclusions are listed as follows:

---

[5] A feature without absolute discounting smooth method is instantiated as :
$$f_{w,c'}(c, d) = \begin{cases} \frac{N(d,w)}{N(d)} & \text{if c'=c} \\ 0 & \text{Otherwise}. \end{cases}$$

1) Identical to the first set of experiments, the error of non-extensive entropy model with high-level constraints increases slightly after a certain amount of word pairs sizes. The results indicate that excessive features not only take a toll on efficiency in text classification, but also cut down the accuracy. The essentiality of feature selection is emphasized once again.

2) Model with smooth technique outperforms the one without. It further proves that we should attach great attention to sparse feature.

3) In comparison with the first set of experiments, we can apparently discover the model with high-level constraints behaves better than the model which only contains linear constraints. As intuition, accuracy is improved by taking word pairs cooccurrence relations into account.

Table 3. The proposed models' accuracy under respective optimal sizes, compared to maximum entropy model and maximum entropy model with a prior.

	Basic MaxiEnt	MaxEnt w/Prior	Model1	Model2
Vocabulary Size	57040	57040	30000	30000
Word Pairs Size				5000
Accuracy	84.23%	84.86%	90.85%	91.72%

Compare the two proposed models with basic maximum entropy model and maximum entropy model with prior at their respective optimal vocabulary sizes, table 3 shows classification accuracy for the four algorithms on our data set. From table 3 we can observe :

1) As intuition, maximum entropy with prior is more accurate than basic maximum entropy method[2].

2) The performance of our proposed methods are remarkably better than maximum entropy models, it is because that IIS is adopted to calculate parameters, as an iterative algorithm, IIS suffers some problems such as numerical stability and slow convergence rate with respect to certain initial configurations. On the other hand, the computation procedure of non-extentsive entropy models is not more than solving a linear system, which could be more efficient and governable in most cases.

3) Non-extensive entropy model with high-level constraints achieves the highest accuracy with a fairly number of active constraints. In intuition, performance will be improved by introducing high-level constraints which stands for semantic relation into the previous model, the result that the latter model performs better supports the idea.

Experimental results demonstrate the advantage of the non-extensive entropy model and non-extensive entropy with high-level constraints.

6   Conclusion

In summary, this paper proposes two models for text classification on the basis of maximum entropy. Replace Shannon entropy with non-extensive entropy to

simplify the form of classifier, furthermore, high-level constraints which stand for semantic relation between word pairs are injected to advance the accuracy of text classification. Experiments on 20_newsgroup data set show that both the proposed models are effective.

Acknowledgements

References

1. Berger, A.L., Della Pietra, S.A., Della Pietra, V.J.: A Maximum Entropy Approach to Natural Language Processing. Computational Linguistics 22(1), 39–71 (1996)
2. Nigam, K., Lafferty, J., McCallum, A.: Using Maximum Entropy for Text Classification. In: IJCAI 1999 Workshop on Machine Learning for Information Filtering, Stockholm, Sweden, pp. 61–67 (1999)
3. Berger, A.: The Improved Iterative Scaling Algorithm: A Gentle Introduction (unpublished manuscript) (1997)
4. Pietra, S., Pietra, V., Lafferty, J.: Inducing Features of Random Fields. IEEE Transactions on Pattern Analysis and Machine Intelligence 19(4), 380–393 (1997)
5. Kazama, J., Tsujii, J.: Evaluation and Extension of Maximum Entropy Models with Inequality Constraints. In: Proceedings of EMNLP 2003, Sapporo, Japan, pp. 137–144 (2003)
6. Ratnaparkhi, A.: A Maximum Entropy Model for Part-of-Speech Tagging. In: Proceedings of EMNLP 1996, Philadelphia, pp. 133–142 (1996)
7. Chen, S.F., Goodman, J.: An Empirical Study of Smoothing Techniques for Language Modeling. Technical Report TR-10-98. Harvard University (1998)
8. Tsallisa, C., Baldovina, F., Cerbinob, R., Pierobon, P.: Introduction to Nonextensive Statistical Mechanics and Thermodynamics. Physica A: Statistical Mechanics and its Applications 305(1), 129–136 (2004)
9. Chen, S.F., Rosenfeld, R.: A Survey of Smoothing Techniques for ME Models. IEEE Transactions on Speech and Audio Processing 8(1), 37–50 (2000)
10. Tsallis, C.: Possible Generalization of Boltzmann-Gibbs Statistics. Journal of Statistical Physics 52(1-2), 479–487 (1998)
11. Abe, S., Rajagopal, A.K.: Nonadditive Condition Entropy and its Significance for Local Realism. Physical A: Statistical Mechanics and its Applications 289(1-2), 157–164 (2001)
12. Tsallis, C.: Entropic Nonextensivity: A Possible Measure of Complexity. Chaos, Solitons & Fractals 13(3), 371–391 (2002)
13. Sven, M., Hermann, N., Jrg, Z.: Smoothing Methods in Maximum Entropy Language Modeling. In: Acoustics, Speech and Signal Processing, IEEE International Conference, Phoenix, Arizona (1999)

E-Learning Systems
with Artificial Intelligence in Engineering

Wojciech Kacalak and Maciej Majewski

Koszalin University of Technology, Faculty of Mechanical Engineering
Raclawicka 15-17, 75-620 Koszalin, Poland
`wojciech.kacalak@tu.koszalin.pl, maciej.majewski@tu.koszalin.pl`

Abstract. This paper presents a new concept of intelligent e-learning systems with intelligent two-way speech communication between an e-learning system and the user. The system uses intelligent methods for analysis, evaluation and assessment of user knowledge and skills as well as e-learning process control, supervision and optimization. Developed as a prototype for mobile technologies, the communication system by speech and a natural language between the intelligent e-learning system and external users consists of intelligent mechanisms for biometric user identification, speech recognition, word and sentence recognition, sentence meaning analysis, and user reaction assessment. Also discussed are selected problems of the new concept of intelligent e-learning systems using intelligent speech communication. The discussion focuses on recognition and evaluation of spoken natural language sentences with use of hybrid neural networks.

Keywords: E-learning system; User-computer interaction; Speech interface; Artificial intelligence; Mobile technology; Cybernetics.

1 Introduction

For the purposes of this research, mobile technologies can be understood as a classification of all tasks in which receiving and passing of data, observation, supervision and control of processes, decision taking, passing of rights, access to means, initiation of undertakings and tasks, as well as prediction of results, can take place without limitations involving movement and location of tasks and computer devices. Mobile technologies aided by artificial intelligence methods produce higher quality, especially when it is necessary to use the expert's assistance regarding distant effects and decision connections, heuristic circumstances of decisions, sudden changes of conditions, while the usage of this knowledge can take place via remote communication.

Building and applying intelligent personal computer and remote mobile systems of realization, control, supervision and optimization of distance learning processes which communicate users with an e-learning system by speech and a natural language, is an important goal. These systems enable a remote supervision of the e-learning process quality using spoken commands in a natural

language and allow to achieve a high quality e-learning process. It is very significant for development of new effective and flexible e-learning systems. It can also contribute for increase of efficiency and decrease of costs of e-learning processes. The system for remote realization, control, supervision and optimization is an innovative solution making it possible to exploit better the e-learning methods used nowadays. It also allows for more complete advantages of applied e-learning processes. The presented solution can be included to the attempts of creating the standard of intelligent personal computer and mobile systems for realization, control, supervision and optimization of e-learning processes using two-way communication by speech and a natural language between the e-learning system and the user for distance learning systems. This subject area will soon become one of the most important for the technological development of e-learning systems.

2   The State of the Art

Speech communication between the e-learning system and the user involves spoken language understanding which is composed of speech recognition and natural language understanding. Many sites in USA, Canada, Europe, China and Japan have been researching spoken language understanding systems. The recent advances in development of prototypes of such systems are described in articles in [1,5,7]. Vocabularies for these systems are usually about tens of thousands of words, and the speech and language are spontaneous. The speech recognition utterance error rates in the 2009 benchmarks was about 5% to 15%. The utterance understanding error rates range from 6% to 41%. These error rates may be compatible with many potential applications. The integration of speech recognition and natural language in applications is faced with many of the same challenges that each of the components face: accuracy, robustness, portability, speed, and size. The integration also gives rise to some new challenges which include: integration strategies, coordination of understanding components with system outputs, the effective use in natural language of a new source of information from speech recognition, and the handling of spontaneous speech effects. With few exceptions, current research in spoken language systems has focused on the understanding of spoken input. However, many if not most potential applications involve a collaboration between the human and the computer. In many cases, spoken language output is an appropriate means of communication that may or may not be taken advantage of, because of lack of coordination of understanding components with system outputs. This paper proposes an approach to deal with the above mentioned problem.

As the Internet gains wide popularity around the world, e-learning is taken by the learners as an important study aid. E-learning systems offer new possibilities in learning, because a user can get immediate feedback on solutions to problems, learning paths can be individualized, etc. On-line learning is a growing business. The number of organizations working on online learning and the number of courses available on the Internet is growing rapidly. E-learning creates a tremendous opportunity for specific institutions and training organizations to

provide on-demand education and training via virtual learning environments. At present, a lot of e-learning tools with varying functionality and purposes exist [2,3,4]. E-learning is an alternative concept to the traditional tutoring system. E-learning forms the revolutionary and new way to empower a workforce with the necessary skills and knowledge [6]. Towards this goal, various e-learning systems have been developed during the last years, however, most of them form old-fashioned applications, missing functionalities like sophisticated capabilities which could be delivered with use of artificial intelligence methods.

3   A New Concept of Intelligent E-Learning Systems

The new concept involves e-learning systems which use artificial intelligence methods and are equipped with intelligent two-way speech communication between the e-learning system and the user (fig. 1). According to the new concept, the e-learning system contains intelligent methods for analysis, evaluation and assessment of the user's knowledge and skills as well as e-learning process control, supervision and optimization. The complete intelligent e-learning system through hybrid artificial neural networks is equipped with an intelligent two-way voice communication system which can optionally use mobile technology (fig. 2). The communication system by speech and a natural language between the intelligent e-learning system and its users consists of intelligent mechanisms of biometric user identification, speech recognition, word and sentence recognition, sentence meaning analysis, and user reaction assessment. The intelligent
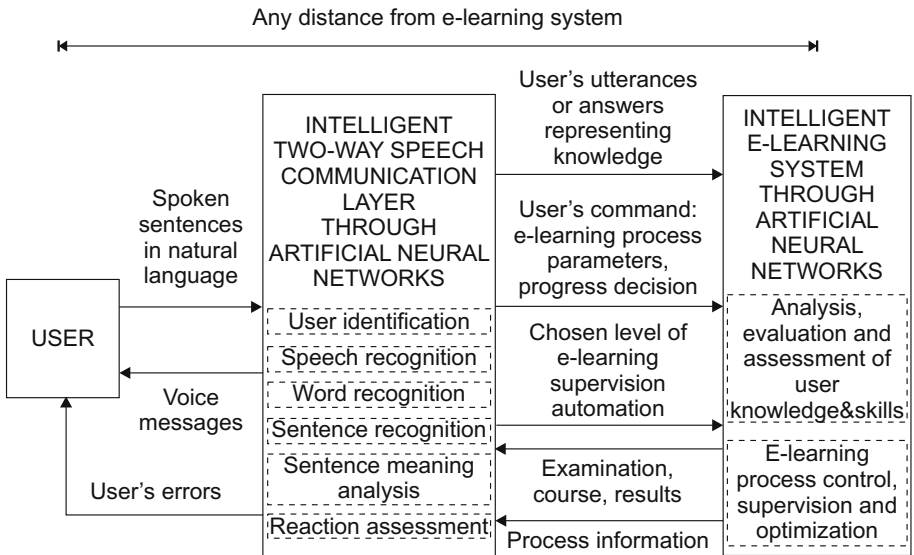

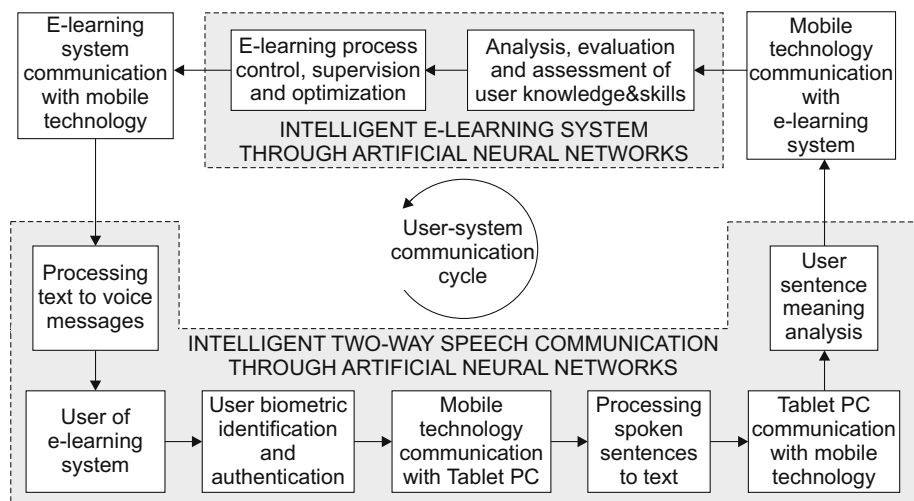**Fig. 1.** Concept of an intelligent e-learning system

Fig. 2. E-learning system for mobile technology

e-learning system (fig. 3A) can be developed for personal computers as well as various mobile technology devices (fig. 3B).

The advantages of intelligent e-learning systems using intelligent two-way communication by speech and a natural language between the e-learning system and the user include the following:

– More robustness against user's errors and more efficient realization, control, supervision and optimization of the e-learning process with the chosen level of supervision automation.
– Improvement of the co-operation between a user and an e-learning system in respect to the richness of communication.
– Achievement of a higher level of organization of a distance learning process that is equipped with an intelligent two-way speech communication system, which is relevant for its efficiency, flexibility, ergonomics and economy.
– E-learning decision and optimization systems can be remote elements with regard to an e-learning system.

The complete intelligent e-learning system is shown in abbreviated form on fig. 4. Upon biometric identification and authentication of the user by the mobile system, the automatic recognition process of the user's utterance or answer in a natural language is performed. The sentences produced in continuous speech are processed to text and numeric values with the module for processing spoken sentences to text on a Tablet PC. The speech recognition engine is a continuous density mixture Gaussian Hidden Markov Model system. The speech signal is transmitted from the PDA to the Tablet PC and after a successful utterance recognition, a text sentence in a natural language is sent to the PDA. Individual words treated here as isolated components of the text are processed with
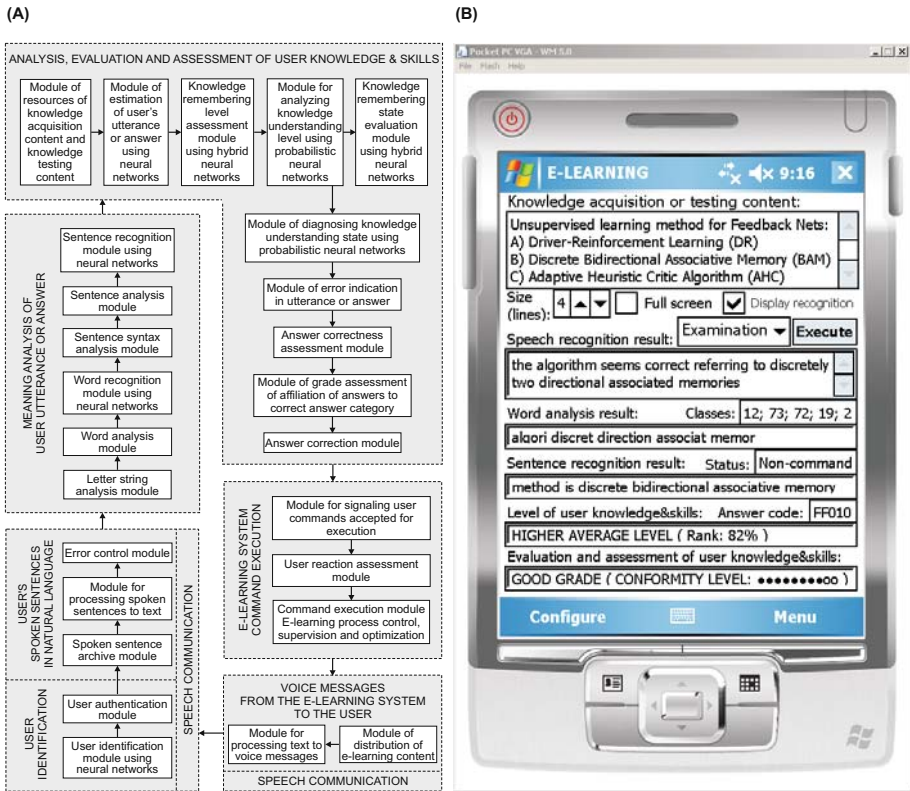
(A)

(B)

Fig. 3. (A) Complete intelligent e-learning system; (B) Prototype of the intelligent e-learning system for mobile technology: an example question and answer by speech and a natural language and the user knowledge evaluation and assessment

the letter string analysis module. The letters grouped in segments are then processed by the word analysis module. The analyzed word segments are inputs of the evolvable fuzzy neural network for recognizing words. The network uses a training file containing also words and is trained to recognize words as sentence components, with words represented by output neurons. In the next stage, the recognized words are transferred to the sentence syntax analysis module which uses sentence segment patterns. It analyses and divides sentences into segments with regards to meaning, and codes sentences as vectors. They are sent to the sentence segment analysis module using Hamming neural networks equipped with sentence segment patterns. The sentence become inputs of the sentence recognition module. The module uses a 3-layer evolvable fuzzy Hamming neural network either to recognize the sentence and find its meaning or else it fails to recognize it (fig. 5). The neural network of this module uses a training file containing patterns of possible meaningful sentences. The word and sentence recognition modules contain Hamming neural networks which feature evolvable architectures and learning (fig. 6).
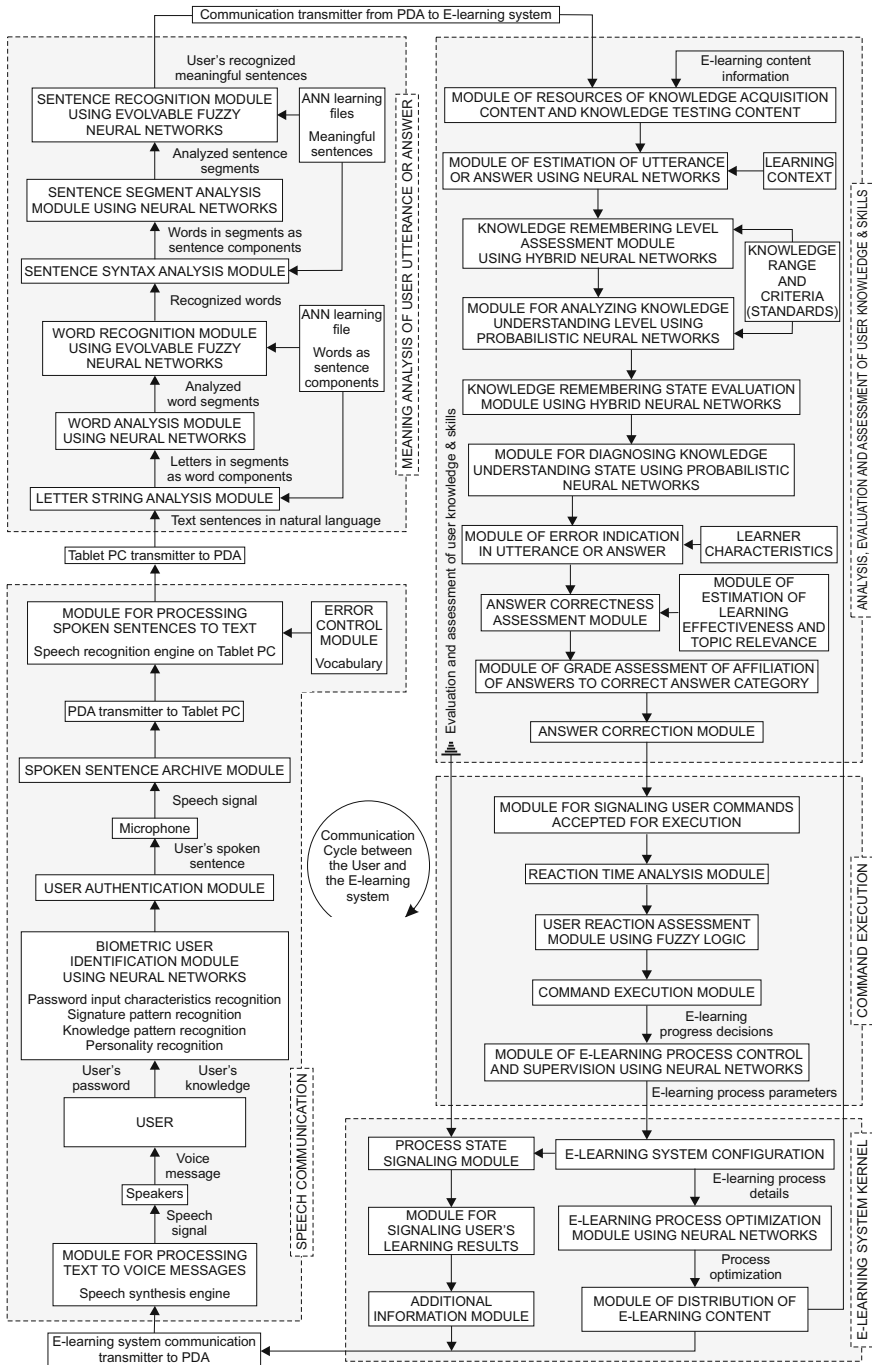
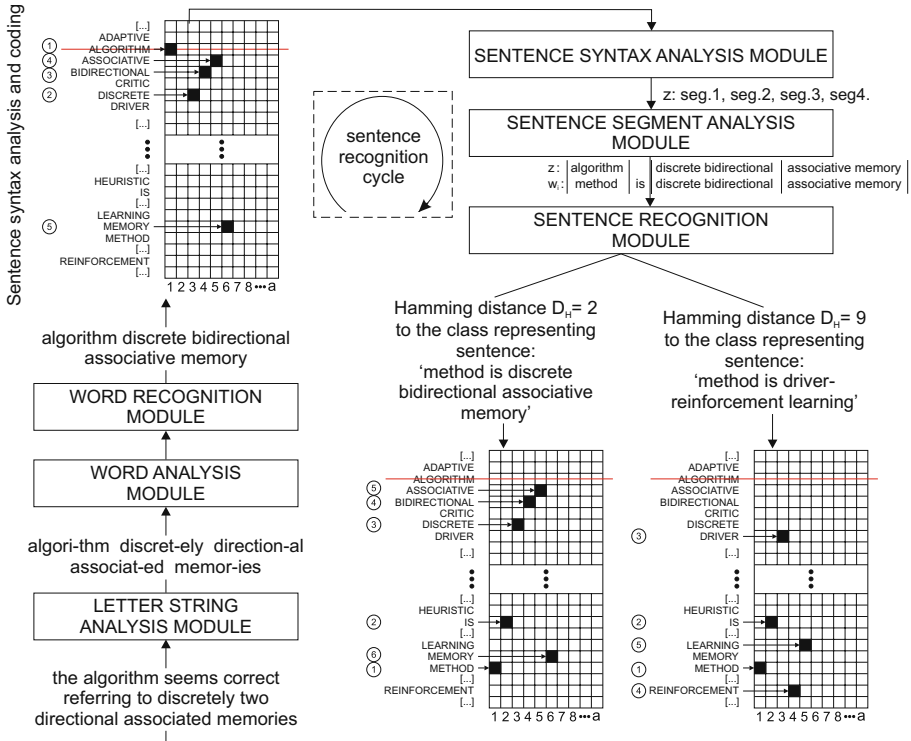**Fig. 4.** Architecture of the complete intelligent e-learning system

Fig. 5. Illustration of a cycle of meaning analysis of the user's utterance or answer

After the meaning analysis process of the user's utterance or answer, their recognized meaningful sentences are processed for analysis, evaluation and assessment of the user's knowledge and skills. The e-learning system estimates the user's utterances or answers, assesses the knowledge remembering level, analysis the knowledge understanding level, evaluates the knowledge remembering state, and consecutively diagnoses the knowledge understanding state. The intelligent methods are applied for these tasks using hybrid and probabilistic neural networks. The system also indicates errors in the utterances or answers, assesses the answer correctness, estimates learning effectiveness and topic relevance, assigns the answers to the correct categories, and makes corrections. The command execution tasks include assessment of the reaction time and reactions of the user, and control of the e-learning process with regards to the definition of progress decisions as well as parameters for the e-learning process. The process supervision is performed using neural networks. The e-learning system kernel is composed of an e-learning system configuration module, modules for intelligent optimization of the e-learning process, as well as signaling the process state and user's learning results. The speech communication to the user involves processing text to voice messages.
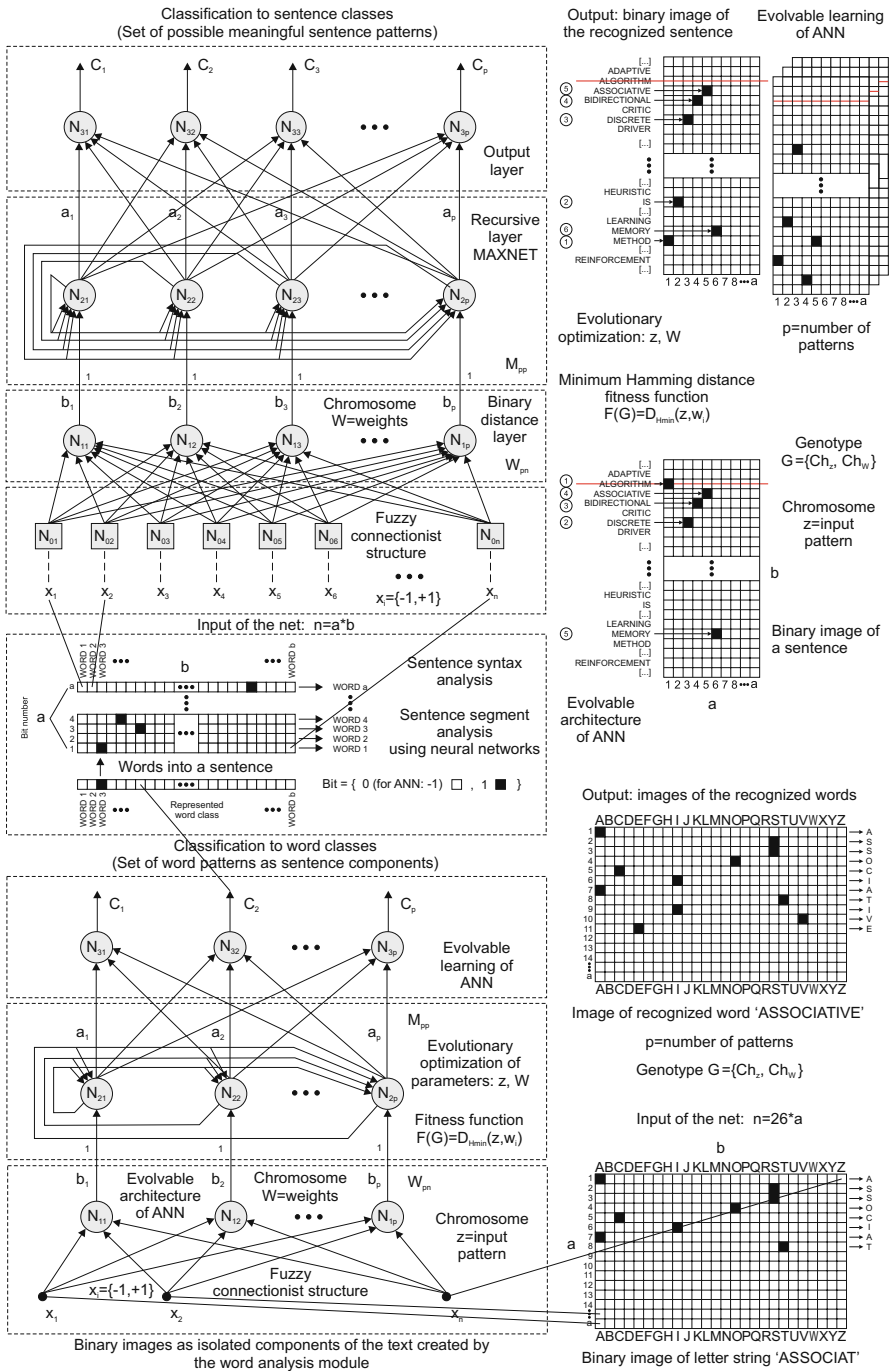
Fig. 6. Evolvable fuzzy neural networks for word and sentence recognition

4   Experimental Results

The prototype of the intelligent communication system was developed for Windows Mobile for a PDA (Pocket PC) and a Tablet PC for Windows.

As shown in Fig. 7A, the ability of the implemented evolvable fuzzy neural network to recognize a word depends on the number of letters of that word. For best performance, the neural network requires a minimum number of letters of each word being recognized as its input.

The ability of the evolvable fuzzy neural network to recognize the sentence depends on the sentence length as shown in Fig. 7B. Similarly, for best sentence recognition, the neural network requires a certain minimum wordcount of the given sentence.
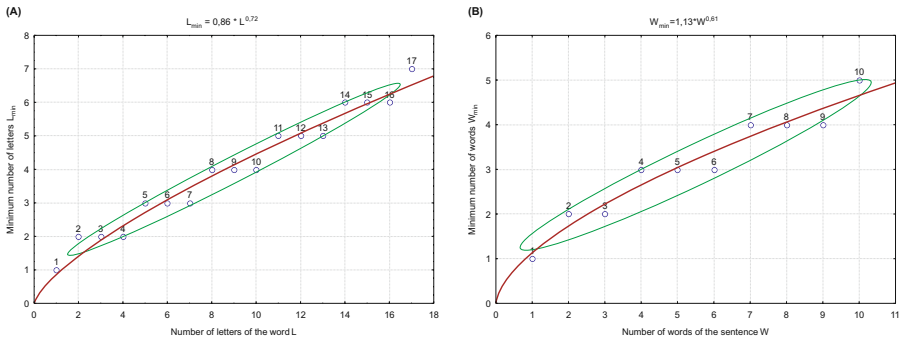

Fig. 7. Experimental results: (A) Sensitivity of spoken word recognition: minimum number of letters of the word being recognized vs. number of word letters; (B) Sensitivity of spoken sentence meaning recognition: minimum number of words of the sentence being recognized vs. number of sentence component words

The experimental implementation of the intelligent e-learning system prototype allowed for an attempt of its evaluation. The testing has enabled to determine for the system the following attribute values [6]:

- Suitability, expressing the degree of appropriateness of the system to the tasks which have to be accomplished.
- Learnability, which conveys how easy it is for the user to learn the system and how rapidly the user can begin to work with it.
- Error rate, which reflects the error ratio while working with the system.

5   Conclusions and Perspectives

A speech interface between users and e-learning systems using the natural language, is ideal because it is the most natural, flexible, efficient, and economical

form of human communication. Application of hybrid neural networks allows recognizing sentences of similar meanings but of different lexical and grammatical patterns, which will undoubtedly be the most important way of communication between humans and computer systems. This presented approach can be included to the methods of natural language processing.

The condition of the effectiveness of the presented system is to equip it with mechanisms of meaning analysis of the user's utterance or answer, as well as analysis, evaluation and assessment of the user's knowledge and skills. In the e-learning systems, the condition of high quality communication between users and e-learning systems is the analysis of the e-learning software and process state before an utterance is performed by the user, and use of artificial intelligence for the analysis, evaluation and assessment of the answer or command.

The developed flexible intelligent e-learning system can be extended for various applications. The experimental results of the proposed system show its promising performance for further development and experiments. The system described in this paper is a conceptually new approach to the problem of lack of coordination of understanding components with system outputs in a spoken language system, although it is an appropriate means of communication that may or may not be taken advantage of.

References

1. Ayres, T., Nolan, B.: Voice Activated Command and Control with Speech Recognition over WiFi. Science of Computer Programming 59(1-2), 109–126 (2006)
2. Govindasamy, T.: Successful Implementation of E-Learning. The Internet and Higher Education 4(3-4), 287–299 (2002)
3. Ismail, J.: The Design of an E-Learning System. The Internet and Higher Education 4(3-4), 329–336 (2002)
4. Larkin, M.: Handheld Use Increasing for E-Learning and Clinical Decision Making. The Lancet 361(9351), 93 (2003)
5. Paek, T., Chickering, D.M.: Improving Command and Control Speech Recognition on Mobile Devices: Using Predictive User Models for Language Modeling. User Modeling and User-Adapted Interaction 17(1-2), 93–117 (2007)
6. Rosenberg, M.J.: E-Learning: Strategies for Delivering Knowledge in the Digital Age. McGraw-Hill, New York (2001)
7. Wald, M., Bain, K.: Universal Access to Communication and Learning: the Role of Automatic Speech Recognition. Universal Access in the Information Society 6(4), 435–447 (2008)

Developing the KMKE Knowledge Management System Based on Design Patterns and Parallel Processing

Lien-Fu Lai[1], Chao-Chin Wu[1], Liang-Tsung Huang[2,*] and Ya-Chin Chang[1]

[1] Department of Computer Science and Information Engineering,
National Changhua University of Education, Taiwan
{lflai,ccwu}@cc.ncue.edu.tw, yachang09@gmail.com
[2] Department of Information Communication, MingDao University, Taiwan
larry@mdu.edu.tw

Abstract. KMKE provides a knowledge engineering approach to integrating knowledge management activities (such as knowledge modeling, knowledge verification, knowledge storage and knowledge querying) into a systematic framework. In this paper, we develop the KMKE knowledge management system based on design patterns and parallel processing. First, several design patterns are applied to develop the KMKE system for enhancing its flexibility and extensibility. Making the KMKE system flexible and extensible is useful to deal with continuous changes originated in knowledge. Second, JAVA programs and CLIPS programs are bound to offer the capability of knowledge inference for the KMKE system. Knowledge verification and knowledge querying can then be performed through the execution of CLIPS rules. Finally, we propose the Parallel CLIPS to shorten the execution time of the KMKE system. Since a large amount of knowledge may increase the execution time substantially, parallelizing the execution of CLIPS rules in cluster system could effectively reduce the search space of the CLIPS inference engine.

Keywords: Knowledge Engineering, Knowledge Management, Expert Systems, Design Patterns, Parallel Processing.

1   Introduction

In the coming age of knowledge, using information technology to assist in acquiring, storing, and applying knowledge becomes a major topic in knowledge management (KM) [9,10]. Many researchers have pointed out several major problems that KM activities may encounter [1,6,13].

- The lack of a systematic approach makes it difficult to pursue and integrate KM activities.
- Both declarative knowledge and procedural knowledge must be captured, stored, and analyzed to solve problems.
- KM activities must support the continuous change of knowledge.
- A common understanding of knowledge is required among knowledge workers for sharing and interoperating knowledge.
- In times of increasing information overload, finding information relevant to the task at hand is becoming increasingly critical.

---

* Corresponding author.

We have proposed a KMKE approach (Knowledge Management through Knowledge Engineering) [7] to alleviating these problems: (1) It offers a systematic approach to integrating the knowledge process of modeling, verification, storage, querying, and updating. (2) It expresses both declarative knowledge and procedural knowledge in a unified CGs formalism, which is helpful in organizing and integrating knowledge. (3) It supports the continuous change of knowledge in the knowledge process. The verification mechanism can be applied to guide the knowledge update process and ensure the consistency of knowledge models. (4) It supports knowledge sharing and knowledge interoperations with its hierarchical ontology system. Higher-level ontologies serve as the common understanding of knowledge to enhance knowledge sharing, while lower-level ontologies facilitate the finding of specific knowledge relevant to a given domain. (5) It alleviates the problem of information overload with its knowledge querying system and the hierarchical ontology system.

KMKE [7] is a knowledge engineering approach to performing knowledge management. An overview of the systematic framework of KMKE is shown in Figure 1. A KMKE system consists of four components: (1) *Knowledge Modeling*. Knowledge is analyzed and captured to form a set of knowledge models. A unified CGs formalism [15] is used to organize and express different types of knowledge. The formal knowledge representation helps in capturing the semantics of knowledge and in offering the capability of reasoning with the knowledge. (2) *Knowledge Verification*. A verification mechanism is proposed to verify knowledge models based on the formal semantics of the knowledge representation. When knowledge models are created or changed, the verification mechanism is applied to ensure their consistency. (3) *Knowledge Storage*. Knowledge models can be classified and stored in a hierarchical ontology system. Ontologies serve as the common understanding of knowledge and facilitate the finding of specific knowledge relevant to a given domain. (4) *Knowledge Querying*. A knowledge query language is designed to enhance the dissemination of knowledge. Knowledge can be reused by way of matching, retrieving, and inferring from knowledge models.
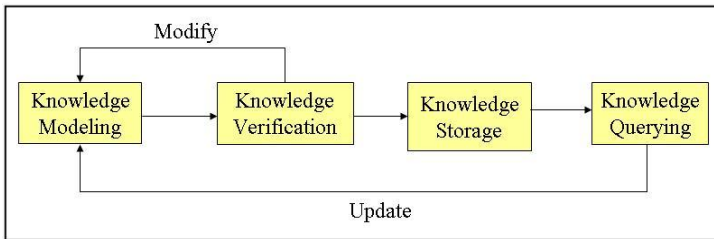

Fig. 1. An overview of the KMKE approach

In this paper, we develop the KMKE system based on design patterns and parallel processing. First, several design patterns are applied to develop the KMKE system for enhancing its flexibility and extensibility. Making the KMKE system flexible and extensible is useful to deal with continuous changes originated in knowledge. Second, JAVA programs and CLIPS programs are bound to offer the capability of knowledge inference for the KMKE system. Knowledge verification and knowledge querying can then be performed through the execution of CLIPS rules. Finally, we propose the

Parallel CLIPS to shorten the execution time of the KMKE system. Since a large amount of knowledge may increase the execution time substantially, parallelizing the execution of CLIPS rules in cluster system could effectively reduce the search space of the CLIPS inference engine.

2   Design Issues in Developing the KMKE System

KM systems are usually modified to reflect changes in the real world and in users' requirements. The implementation of the KMKE system raises several design issues:

- The software architecture should be flexible and extensible. Since the requirements of KM are frequently changed, the KMKE system must be easy to modify and easy to extend.
- The formulation of knowledge should be capable of inference. Since useful knowledge often derives from known facts and rules via knowledge inference, the reasoning capability is required for the KMKE system.
- The execution of the KMKE system should be efficient. Since knowledge verification and knowledge querying are achieved by matching CGs and inferring CGs, a large amount of knowledge may increase the execution time substantially. Shortening the execution time is required for KMKE to enhance its usability.

2.1   Design Issue 1: How to Make the KMKE System Flexible and Extensible?

To enhance the flexibility and the extensibility, the Model-View-Controller (MVC) pattern [17] (see Figure 2) is applied to construct the software architecture of the KMKE system. The MVC architecture divides the system into three subsystems. The model classes are responsible for maintaining the data, the view classes are responsible for displaying the data, and the controller classes are responsible for controlling the system flow. In the MVC architecture, the lower layers are defined independently of any upper layers. For example, the model is independent of the view. It would therefore be able to update the view classes without affecting the model classes.
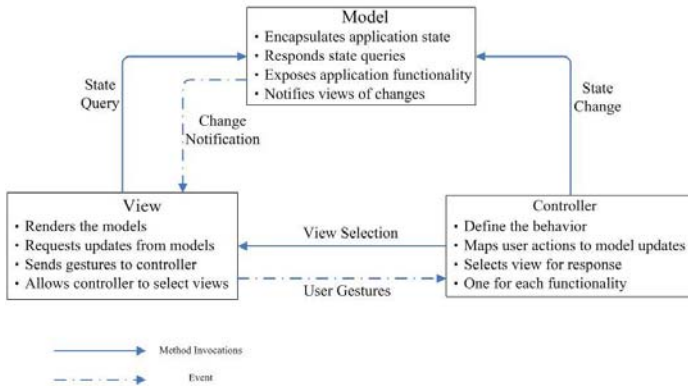

Fig. 2. MVC architecture pattern

Whenever knowledge nodes in the model classes are changed, conceptual graphs in the view classes should be updated to reflect the change. That is, the model classes need to inform the view classes for triggering a view update. It would cause a violation of the MVC architecture in which the model classes should be independent of the view classes. To solve the dependency-violating problem, we apply the Observer design pattern [16] (see Figure 3) to preserve the dependencies in the MVC architecture.
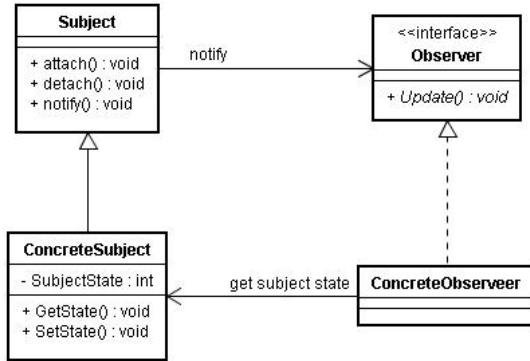
Fig. 3. Observer design pattern

The model classes (i.e. Subject) maintain references to a set of observers. When the state of a model object changes, a Update() message would be sent to the observer. The observers are implemented by the ConcreteObserver classes defined in the view layer, but the model objects only access them through the observer interface defined in the model layer. A change to a model object would notify an observer interface defined in the model layer, and a ConcreteObserver object in the view layer would implement the Update() operation to reflect the change.

Fig. 4. Transfer Object design pattern

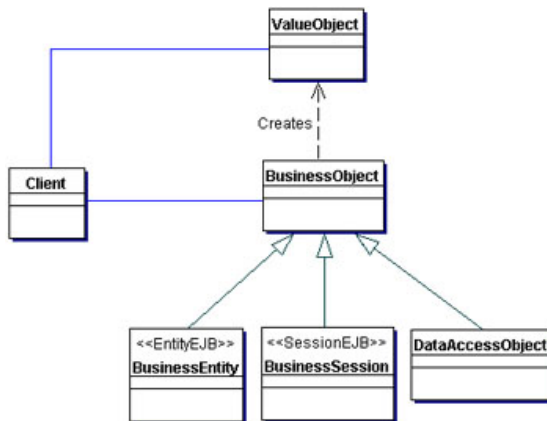Entities often include a large group of attributes within objects. The client needs to invoke a object's get methods multiple times until it obtains all the attribute values. If the application accesses only a few attributes from databases each time, using multiple calls to get methods that return a few attribute values is inefficient for obtaining data values from an enterprise bean. We apply the Transfer Object design pattern [18] (see Figure 4) to solve the database access problem. A Transfer Object is used to encapsulate the data, and a single method call is used to send and retrieve the Transfer Object. When the client requests the enterprise bean for the data, the enterprise bean can construct the Transfer Object, populate it with its attribute values, and pass it by value to the client. The related attributes of objects are combined into value objects (VO) to avoid wasting time and system resources.
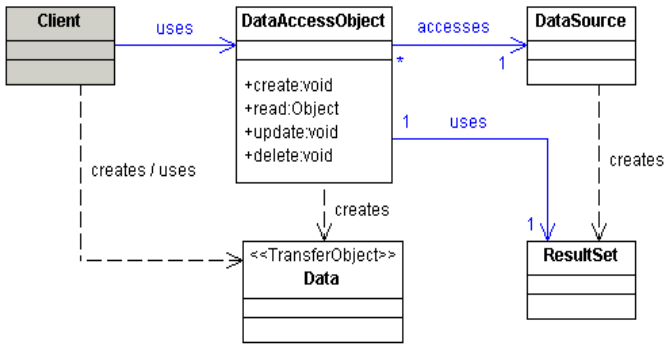

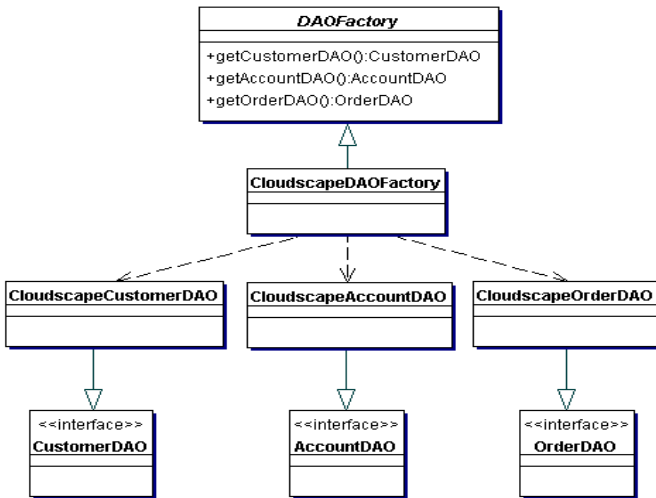**Fig. 5.** Data Access Object (DAO) design pattern

Fig. 6. Apply Factory Method to construct DAO

Code that depends on specific features of data sources would tie together business logic with data access logic. This makes it difficult to replace or modify an application's data sources. We apply Data Access Object (DAO) design pattern [19] (see Figure 5) to abstract and encapsulate all access to the data source. The DAO manages the connection with the data source to obtain and store data. Therefore, the DAO can serve as the adapter between the business logic and the data source to allow data access mechanisms to change independently of the code that uses the data.

To access different data sources, we apply Factory Method [3] (see Figure 6) to let a class differ instantiation to subclasses. We could define an DAOFactory interface for creating an object, and let subclasses decide which class to instantiate. It enables the subclasses to provide an extended version of an object, because creating an object inside a class is more flexible than creating an object directly in the client.

2.2   Design Issue 2: How to Offer the Capability of Knowledge Inference for the KMKE System?

To verify and deduce knowledge models automatically, CLIPS are used as the target language to implement the KMKE system. Whenever a knowledge model has been constructed or updated, conceptual graphs in the knowledge model will be translated into their counterparts in CLIPS automatically. In knowledge verification, three types of verification (i.e., syntactic, structural, and semantic) are performed on each domain to ensure consistency of the knowledge models. They are implemented as verification rules in CLIPS. Through the deductions of the CLIPS inference engine, inconsistency in knowledge models can be found automatically. In knowledge querying, proper knowledge in CGs can be retrieved via graph matching. The graph-matching mechanism is implemented by CLIPS rules based on semantic matching and analogical matching. The answer to a knowledge query can be deduced from knowledge models based on the CG's rules of inference, which are also implemented by CLIPS rules.

Hence, it is required to integrate JAVA programs with CLIPS programs. We use the exchange of files to communicate the procedural language (i.e. JAVA) with the non-procedural language (i.e. CLIPS). For example, the JAVA program generates facts and writes them to a common file. The CLIPS program reads facts from the common file and makes inference on the facts. The results of inference are written back to the common file, and the JAVA program reads them for updating display. To integrate JAVA programs with CLIPS programs, we use JCLIPS package [11] to bind JAVA and CLIPS. JCLIPS embeds CLIPS cores and the inference engine in the JAVA package. It allows JAVA programs to connect with CLIPS's inference engine and to control the execution of CLIPS programs including loading files, running the inference engine, and executing commands.

2.3   Design Issue 3: How to Shorten the Execution Time of the KMKE System?

In KMKE, knowledge verification and knowledge querying are performed by matching CGs and inferring CGs. Both graph matching rules and knowledge inference rules are implemented as CLIPS rules. Therefore, the number of CGs nodes in a domain is the main factor that affects the response time of executing knowledge queries and verifying knowledge models. A large amount of knowledge may increase the execution time substantially.

We use two mechanisms to raise the execution efficiency of the KMKE system.

● The hierarchical ontology system is used to classify and store knowledge models. By retrieving the appropriate domain in the hierarchy, the search space can be reduced.

● The cluster system and the MPI library are applied to parallelize CLIPS programs. Both knowledge verification rules and knowledge querying rules can execute in parallel to shorten the execution time of the KMKE system.

A cluster system [2,4,5] is a type of parallel and distributed processing system, which consists of a collection of interconnected stand-alone computer working together as a single integrated computing resource. We adopt MPI de-facto standard [20,21] to extend CLIPS to be a parallel programming language.
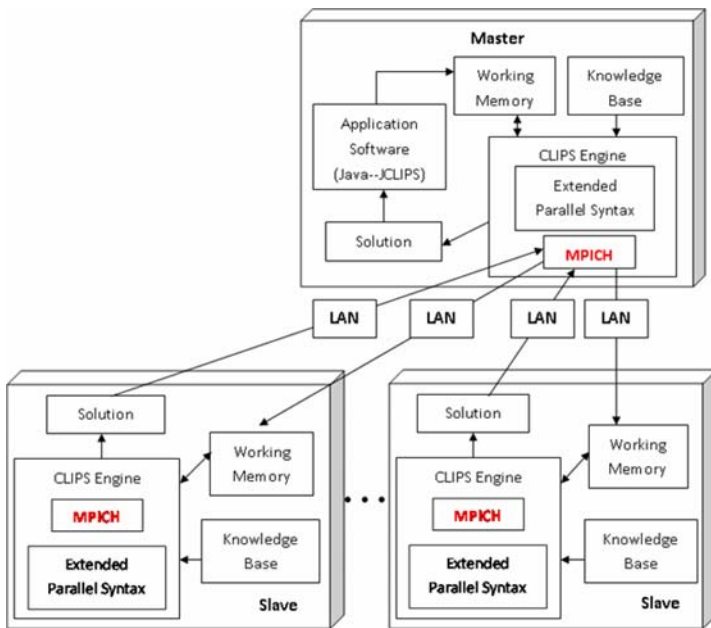
Fig. 7. Execution of a parallel CLIPS program in the cluster system

To extend CLIPS to be a parallel programming language, we need to add some new syntax. The CLIPS inference engine provides two main methods for integrating with user-defined code. The first one is to integrate the CLIPS engine with external functions. A developer can add external functions to the CLIPS inference engine, pass arguments to them, and return values from them. The second one is to embed the CLIPS inference engine within other programs. Calls to the CLIPS inference engine are made like any other subroutine in the main program. The CLIPS inference engine has implemented many functions to provide users with the ability to access internal information.

The first method is appropriate to define new syntax for use in writing CLIPS parallel program. New syntax should follow the same style of CLIPS language and will

be parsed and executed by a corresponding external function implemented inside the CLIPS inference engine. Furthermore, a new and simple syntax can perform a complicated operation on internal information. In this way, programming a parallel CLIPS application can be simplified.

To execute a CLIPS application in parallel, the CLIPS inference engine is parallelized with MPICH as shown in Figure 7. A target application is required to be coded with extended syntax in a CLIPS program, which is an input file for the CLIPS inference engine during execution.

4   Implementing the KMKE System

We apply the object-oriented software engineering approach [14] to developing the KMKE system. The layered architecture pattern (i.e. MVC architecture) is applied to organize the KMKE system. The software architecture is divided into five layers: a Presentation layer, an Application layer, a Controller layer, a DAO layer, and a VO layer. Classes concerned with the user interface are placed within the Presentation layer. Classes that maintain the system state and implement the business logic of the application are placed in the Application layer. The Controller layer is responsible for system flow control and serves as a bridge to communicate the Presentation layer and the Application layer. The DAO layer encapsulates all database access within DAO objects and constructs DAOFactory objects as the interface to access DAO objects.
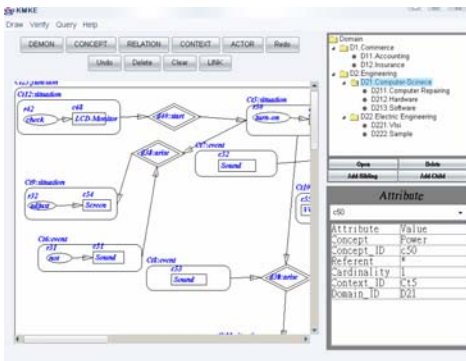
Fig. 8. Editing knowledge models in KMKE   **Fig. 9.** Querying knowledge in KMKE

The VO layer contains the value objects of CGs nodes to encapsulate data. Each layer is represented by a package to group related classes or nested packages. The Application layer is defined independently of the Presentation layer and the Controller layer. Therefore, we can modify the user interfaces defined in the Presentation layer or the control flow defined in the Controller layer without making any changes to the core classes of the application. Similarly, both the DAO layer and the VO layer are defined independently of the Application layer, and the modification of the Application layer has no influence on the two layers. Furthermore, four design patterns are applied to refine the class diagram, including Observer, Transfer Object, Data Access Object and Factory.

Java is used to implement the application software, and CLIPS is used to implement the knowledge inference in knowledge verification and knowledge querying. In the proposed Parallel CLIPS, both rules and data could be divided to execute in parallel. On the one hand, the module structure in CLIPS is adopted to implement the separation of rules for parallel processing. Each module contains a group of rules which can execute independently. MPICH assigns modules to different CPUs for parallel execution and calls the functions of CLIPS for transmitting data. On the other hand, the master could distribute data to slaves for parallel processing. Each slave executes with the assigned partial data and returns results to the master via the external functions of CLIPS. Figure 8 shows the editing of knowledge models and Figure 9 shows the result of knowledge querying.

5   Conclusion

In this paper, we develop the KMKE system based on design patterns and parallel processing. First, several design patterns are applied to develop the KMKE system for enhancing its flexibility and extensibility. Second, JAVA programs and CLIPS programs are bound to offer the capability of knowledge inference for the KMKE system. Finally, we propose the Parallel CLIPS to shorten the execution time of the KMKE system.

Considering the cohesion of modules, rules for knowledge querying are divided into seven modules. Each module contains a subset of rules related to one question type and is allocated to a processor. Table 1 shows the response times of executing knowledge queries in the non-parallel and parallel environments, respectively. A knowledge query that contains seven question types is applied to a domain consisted of 1000 knowledge nodes. Subsequently, the domain is repeatedly enlarged with duplicated 1000 knowledge nodes to 10000 knowledge nodes. The best performance improvement ratio is 4.945 while seven processors are used to parallel the queries of 6000 knowledge nodes. The average performance improvement ratio is 3.428 in 1000-10000 knowledge nodes.

Rules for knowledge verification are also divided into seven modules. Each module contains a subset of rules related to one type of verification and is allocated to a processor. Table 2 shows the response times of executing knowledge verification in the non-parallel and parallel environments, respectively. Seven types of verification are applied to 1000-10000 knowledge nodes. The average performance improvement ratio is 3.626 while seven processors are used to parallel the verification of 1000-10000 knowledge nodes. The performance improvement ratio is good in the cases of 1000-5000 knowledge nodes.

Table 1. The execution time of knowledge querying

KMKE_query (# of knowledge nodes)	1000	2000	3000	4000	5000	6000	7000	8000	9000	10000
non-parallel(sec)	1.146	2.439	4.94	9.267	17.035	32.29	45.331	51.25	67.451	101.781
parallel(sec)	0.348462	0.69222	1.640259	2.937267	4.865162	6.529676	11.172121	17.076257	24.467055	33.47546

Table 2. The execution time of knowledge verification

KMKE_verification (# of knowledge nodes)	1000	2000	3000	4000	5000	6000	7000	8000	9000	10000
non-parallel(sec)	1.605	3.931	8.469	21.704	36.664	40.726	62.398	74.977	96.6	118.624
parallel(sec)	0.252749	0.874932	1.930685	3.721334	7.536087	15.145918	26.769478	37.25112	53.634791	78.95628

Therefore, each domain in the hierarchical ontologies had better contain no more than 5000 knowledge nodes. If there are more than 5000 knowledge nodes in a domain, the KMKE system will recommend that users should split the domain into two.

The advantages of the proposed system are as follows.

- It provides a systematic and automatic framework to integrate several KM activities. JAVA programs and CLIPS programs are bound to offer the capability of knowledge inference for the KMKE system.
- Applying design patterns makes the KMKE system more flexible and extensible. It is useful to deal with continuous changes originated in knowledge.
- Parallelizing the execution of CLIPS rules in cluster system could effectively reduce the search space of the CLIPS inference engine. It is useful to shorten the execution time of the KMKE system.

References

1. Fischer, G., Ostwald, J.: Knowledge Management: Problems, Promises, Realities, and Challenges. IEEE Intelligent Systems, 60–72 (January 2001)
2. Franck, C., Olivier, R., Daniel, E.: Understanding Performance of SMP Clusters Running MPI Programs. Future Generation Computer Systems, 711–720 (April 2001)
3. Gamma, E., Helm, R., Johnson, R., Vlissides, J.: Design Patterns. Addison-Wesley, Reading (1995)
4. Goscinski, A.M., Wong, A.K.L.: A Study of the Concurrent Execution of Parallel and Sequential Applications on a Non-dedicated Cluster. Parallel Computing, 69–91 (2008)
5. Guo, J., Bhuyan, L.N.: Load Balancing in a Cluster-Based Web Server for Multimedia Applications. IEEE Transactions on Parallel and Distributed Systems, 1321–1334 (2006)
6. Kemp, L.L., Nidiffer, K.E., Rose, L.C., Small, R., Stankosky, M.: Knowledge Management: Insights from the Trenches. IEEE Software, 66–68 (November 2001)
7. Lai, L.F.: A Knowledge Engineering Approach to Knowledge Management. Information Sciences 177, 4072–4094 (2007)
8. Lee, J., Lai, L.F.: A High-Level Petri Nets Based Approach to Verifying Task Structures. IEEE Transactions on Knowledge and Data Engineering 14(2), 316–335 (2002)
9. Liao, S.H.: Knowledge Management Technologies and Applications – Literature Review from 1995 to 2002. Expert System with Applications 25, 155–164 (2003)
10. McBriar, I., Smith, C., Bain, G., Unsworth, P., Magraw, S., Gordon, J.L.: Risk, Gap and Strength: Key Concepts in Knowledge Management. Knowledge-Based Systems 16, 29–36 (2003)
11. Menken M.: JClips - CLIPS for Java (2005), http://www.cs.vu.nl/~mrmenken/jclips/#clips
12. Mineau, G.W., Missaoui, R., Godix, R.: Conceptual Modeling for Data and Knowledge Management. Data and Knowledge Engineering 33, 137–168 (2000)

13. Preece, A., Flett, A., Sleeman, D., Curry, D., Meany, N., Perry, P.: Better Knowledge Management through Knowledge Engineering. IEEE Intelligent Systems 72 (January 2001)
14. Rumbaugh, J., Jacobson, I., Booch, G.: The Unified Modeling Language Reference Manual, 2nd edn. Addison-Wesley, Reading (2005)
15. Sowa, J.F.: Conceptual Graphs: Draft Proposed American National Standard. In: 7th International Conference on Conceptual Structures (ICCS 1999), pp. 1–65 (1999)
16. Srinivasan, S.: Design Patterns in Object- Oriented Frameworks. Computer 32(2), 24–32 (1999)
17. Sun Developer Network, MVC Pattern,
    `http://java.sun.com/blueprints/patterns/MVC-detailed.html`
18. Sun Developer Network, TransferObject Pattern,
    `http://java.sun.com/blueprints/corej2eepatterns/Patterns/`
    `TransferObject.html`
19. Sun Developer Network, DataAccessObject Pattern,
    `http://java.sun.com/blueprints/corej2eepatterns/Patterns/`
    `DataAccessObject.html`
20. The MPICH Implementation, `http://www-unix.mcs.anl.gov/mpi/mpich1/`
21. The Message Passing Interface standard, `http://www-unix.mcs.anl.gov/mpi/`

A Fuzzy Logic Based Approach to Feedback Reinforcement in Image Retrieval

Vincenzo Di Lecce[*] and Alberto Amato

DIASS, Politecnico di Bari 74100 Taranto, Italy
{v.dilecce,a.amato}@aeflab.net

Abstract. Nowadays, due to the spread of digital imaging technologies, the design of effective content based image retrieval (CBIR) systems is perceived by the research community as a primary problem. Various techniques such as clustering and relevance feedback were proposed to obtain a certain level of knowledge about a given image database. Often clustering techniques were used to obtain a first level characterization of the image database used to speed up the successive stage of queries. In this work the authors use the knowledge obtained using a fuzzy clustering algorithm to reinforce the user feedback. The system was tested on the Columbia Coil-20 image database and the obtained results seem to be encouraging.

Keywords: Intelligent image retrieval system, fuzzy clustering, knowledge management, relevance feedback.

1 Introduction

With the rapid diffusion of multimedia digital devices such as digital cameras and digital camcorders, image databases with thousands of digital photos are becoming more and more common. Often, the search for similar images into these databases is quite challenging for a human operator because it is time consuming and a burdensome work. On the other hand, human centricity is profoundly present in this kind of task, no matter how much the entire process could be automated, there is a pivotal role of a human user. The individual plays a primary role both as user and designer of the system. Digital albums need to be personalized: there should be an adjustment mechanism of their structure according to the individual needs of the user. As each user could have different preferences as to the same collection of images, this feature becomes important and may influence the users' opinion as to the performance of the system. At the same time while the relevance feedback is important, there should be efficient implementations of this feedback mechanism without mentioning a comprehensive suite of algorithms that support an automatic process of image annotation and formation of the overall structure. The automation of this process is an open challenge for the research community that is designing various image retrieval systems.

An image retrieval system can be seen as a system that uses a certain knowledge to efficiently index a given image database. From this point of view, Content Based

---

[*] Contacting author.

Image Retrieval (CBIR) systems try to use information automatically extracted from the images to build their own knowledge base.

This information refers to a description of images obtained using a set of visual features (often called signatures) that can be automatically extracted from them. There are many examples of low-level features such as: shape, color, and texture that are utilized to characterize images both in commercial (e.g., QBIC [1]) and in academic systems (e.g., Photo-book [2]). Other interesting features are those based on wavelet analysis [3] and those operating directly on compressed images [4]. The knowledge base, defined in this way, is used in every retrieval task. In order to evaluate the similarity between two images, a CBIR system computes the distance between their signatures. This value is considered as correlated to images similarity and it is used to answer user's query.

The main drawback of this approach is that low level features do not completely characterize all the semantic contents of an image. So, while user is searching for a semantic similarity among images, the system is searching for similarity among their visual features. This problem is known as semantic gap [5]. From the knowledge point of view, it is possible to say that the CBIR system is using a knowledge base characterized by a low semantic level.

There are approaches proposed by various authors that aim at rising the quality level of this knowledge base by using clustering algorithms and the user feedback (e.g. [6]).

Relevance feedback techniques are used in several ways, for example it can be used to modify the weights of a distance function [7] or to pose multi-image queries to the system as in [8] and in this work.

Clustering is an important technique used in discovering inherent structure present in a set of objects. In various CBIR systems, clustering techniques are used to enhance the efficiency and effectiveness of the retrieval process. Indeed, according to CBIR assumptions (perceptually similar images are situated close to each other within a connected region of a given space of visual features), each cluster is composed of similar images at least from the point of view of the low level features. So, each cluster conveys information about a neighborhood of the query space.

In this work the authors propose an advanced CBIR system implementing relevance feedback for multi-image queries and using the knowledge obtained by means of a fuzzy clustering algorithm to reinforce the user feedback.

The remaining part of this work is so organized: Section 2 briefly discusses related work, while section 3 describes the proposed system highlighting the adopted feature space and the used clustering method. Section 4 reports the conducted experiments and the obtained results. Conclusions and final remarks are presented in section 5.

2   Related Work

Content-based image retrieval (CBIR) has emerged as an important area in computer vision, multimedia computing, just to name a few representative examples. Typically in CBIR systems, images are described by using a collection of some visual features (often referred to as signatures) extracted from the individual images. In the literature there are many low-level visual features. Shape, color, and texture are those in

common use both in commercial systems (e.g., QBIC [1]) and in the academic community (e.g., Photo-book [2]). In [9] an improved version of color signature embedding some local information about the statistical and visual relevance (importance) of each pixel was proposed.

With the increasing diffusion of compressed images, there have been various attempts to extract their visual features in the compressed domain, see [10, 4]. The results obtained using more sophisticated and abstract features such as wavelets [3,11] are particularly interesting. Color, texture, shape, etc are used as visual descriptors into the MPEG-7 standard [12]. There are interesting approaches combining more than one visual feature to improve the effectiveness of the retrieval systems. For example in [13,14] the authors combine color and texture while in [15] they use color and shape building other aggregates. The study [5] provides a comprehensive overview of the theory, techniques and applications of content-based image retrieval.

Several image retrieval systems use clustering as a certain pre-processing step. In many studies, such as [16,17], clustering is used to segment images as an early step in the image understanding process. In other works, the underlying objective is to complete some organizations of the entire image collection into some homogeneous groups. In both approaches the processing time of the clustering algorithm is a critical point. Many works deal with this problem and a promising approach can be found in [18]. In [19] the authors use the clustering to build a CBIR system that makes the user aware of the overall image distribution into the database.

3   Proposed System

In this work the authors propose a CBIR system using fuzzy logic approach to reinforce the mechanisms of multi-image queries and relevance feedback. The information given by user with relevance feedback represents the knowledge at the highest semantic level into the system. Unfortunately, for practical reasons, this knowledge is related to a small portion of the image database (only for the images retrieved by the system as relevant for user query). For this reason, the authors propose to use the knowledge deriving from the FCM algorithm to enrich the previous human based knowledge.

Figure 1 shows a schematic overview of the proposed system. The phase of database population is highlighted using a grey shadow. It is composed of:

- the feature extraction module: it extracts the low level features from each image into image database, see section 3.1 for more details.
- the FCM module: it uses the Fuzzy C-Means algorithm to obtain a partition of the database. The obtained fuzzy clusters and the related partition matrix are stored into the fuzzy cluster DB, see section 3.2 for more details.

The query phase is highlighted using the cyan shadow. In this phase, the user can pose a query to the system using the query by example method. The system extracts the low level features from the query image (using the same feature extraction module used in database population phase) and sends them to the query engine module. This module computes distances between the query and the images into the database using the fuzzy knowledge as weight in the used distance function and then it shows the most relevant images to the user. At this point an iterative process of user feedback
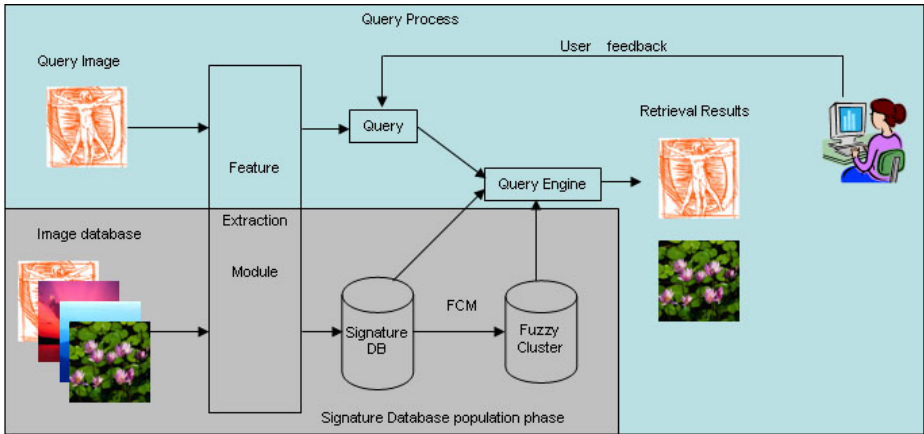
Fig. 1. A schematic overview of the proposed system

starts in which the user can select the images that he/she considers relevant to his/her similarity idea. These images have a high semantic value since they are considered relevant by the user. These images are used by the system to implement the multi-image queries computing the centroid of the signature associated to them (not necessarily coinciding with a real image) and using it as stimulus in the various retrieval steps in combination to the mechanism of fuzzy feedback reinforcement. This interactive visual query process leads to the creation of the final query and the final desired set of images extracted from the database.

3.1 Image Description

In this work a global feature was utilized to describe images, namely the angular spectrum signature [20]. The main hypotheses at the base of the formulation of this signature are:

- visual properties are mainly related to the largest objects in images;
- among the various object visual features, shape, texture and orientation play a major role in similarity assessment.

Following these hypotheses, the orientation of objects within an image and its shape can be key attributes in the definition of the similarity with other images.

The problem of finding the distribution of line directions within an image was faced by analyzing its Fourier's transform.

Before the forward Fourier's transform, each image is pre-processed applying a windowing function (2D Hamming function) in order to remove discontinuities at the edges which can complicate the interpretation of frequency spectra. The so obtained Fourier's transform is processed to obtain a polar representation of the image spectrum in which the zero-frequency component is shifted to center. This representation of the image spectrum is split in 10 angular sectors, each one is large 18° starting from $\theta = \pi/2$ down to $\theta = -\pi/2$. For each sector, 7 harmonics are evaluated and recorded in a histogram using the values of $\rho = 1, 2, \ldots, 7$. This histogram, composed of 7*10 values, is the angular spectrum signature (see figure 2.b). It is noteworthy that the angular
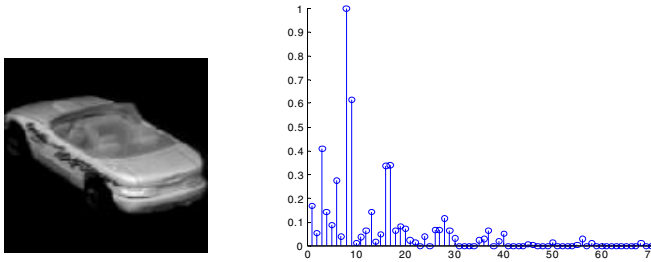
Fig. 2. (a) An example image, (b) The Angular Spectrum signature extracted from the image (a)

spectrum signature considers only the low frequencies. Indeed, the low frequencies correspond to the largest components within the image. These components are usually the most relevant ones for interpreting the image and can be assumed to be more frequently localized in central areas (e.g., foreground objects) or to be spread in the whole image (e.g., landscapes). Their contribution to the signature generation is almost completely preserved after the 2D Hamming function application.

High frequencies correspond to small details and fine-grain textures; a reduction of these components means focusing on the foreground components, located in the image central area, and ignoring the details of the peripheral contour.

3.2   Fuzzy Clustering

Clustering is a well-known technique used to discover inherent structure inside a set of objects. Clustering algorithms attempt to organize unlabeled pattern vectors into clusters or "natural groups" so that points within a cluster are more similar to each other than the points belonging to different clusters.

If the database images are successfully classified into different semantic clusters, a visual query can be quickly narrowed to a specific category. Therefore, image retrieval can proceed effectively and efficiently. However, different semantic clusters can have arbitrary shapes and overlaps, so, it may be impossible to detect and specify them accurately.

In this work, the clustering is not used just to speed-up the query process, but it is used to build a fuzzy knowledge base used to reinforce the user feedback in the query process. Fuzzy clustering was preferred to Boolean clustering due to the nature of the problem under discussion. Indeed, often it is difficult to assign an image to a single class because that image can contain many objects belonging to different classes [21]. For this reason, Boolean clustering algorithms, that attempt to label each patterns as "member" or "not member" of each class, are not suitable for this kind of applications.

In this work various experiments were carried out to define a suitably fuzzy knowledge of the image database using the FCM algorithm. The aim of these experiments was to find the values of the various FCM algorithm parameters maximizing the percentage of correctly classified patterns. In particular, the number of clusters was fixed to 20, while the value of the fuzzification coefficient (m) was found running the FCM

algorithm by using different values of m and choosing the one that offered the minimum classification error.

4   Experiments and Results

The experiments were carried out on Columbia Coil-20 image database [22]. This database comprises 1,440 grey images of 20 objects. There are 72 images per object where a photo of each image is taken at pose intervals of 5 degrees. The authors used this image database due to the acute lack in literature of a common standard database to be used as a test bed for CBIR systems. This database was often used in image retrieval and categorization e.g. [23, 24]. There are other image databases of general format. The most popular one is the Corel Image Database. This database is composed of a set of more than 800 Photo CDs where each of them consists of 100 images that are grouped along the same category. It seems that this database could solve the problem of the lack of common test bed for the validation of CBIR systems, however this is not true. In an interesting study [25] the authors show that it is very easy to reach different results, even in cases when the same collection of images is being used. Some selected examples of the Columbia image library are included in figure 3.

Fig. 3. 20 example images of the Columbia Object Image Library Coil-20

Two CBIR systems were implemented. They used the same image database and the same visual features and both implemented the same mechanism of relevance feedback and multi-image query. The only difference between the two CBIR systems is that the second one implements the proposed fuzzy logic approach to reinforce the mechanisms of multi-image queries and relevance feedback. Figure 4 reports the results of the comparative evaluation between the two systems. The shown results are the average performance values obtained posing 200 queries (10 images for each class) to both the CBIR systems. For each query 10 steps of relevance feedback were carried out using the method described in section 3. It should be highlighted the shape of the performance curve in figure 4.b representing the performance obtained using the proposed fuzzy clustering based feedback reinforcement method. In the second
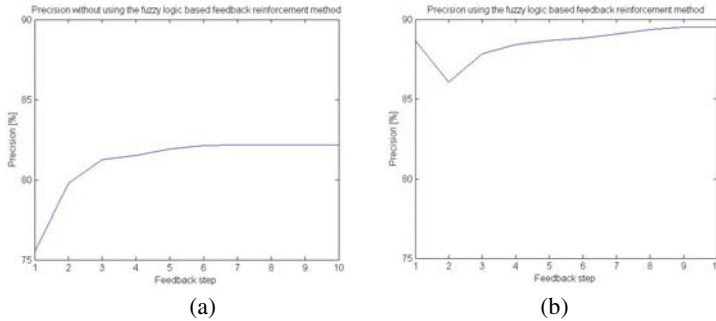
Fig. 4. The average performance in terms of precision obtained during the experiments without using the fuzzy logic based feedback reinforcement method (a) and while using it (b)

relevance feedback iteration there is a lower performance than that in the first step (where the query by example method is used). This fact is due to the initial combination of the multi-image query and the fuzzy clustering based feedback reinforcement method. After the initial performance loss, the combination of these methods works as expected and the retrieval precision grows.

In order to obtain the a-priori knowledge base of the image database the FCM algorithm with $m=1.6$ and $C=20$ was used. These values of fuzzification coefficient and several clusters are those that gave the better classification rate during the experimentation phase [26].

Various experiments were conducted to evaluate the retrieval performance of the proposed relevance-feedback algorithm in terms of precision.

Precision is defined as the ratio between the number of relevant images and the total number of images returned by the system. As figure 4 shows, the proposed system achieves an average precision increment of 15%.

5   Conclusions

Image retrieval systems are becoming ever more required due to the spread of digital imaging devices. In this work, the authors have presented an image retrieval system, which is able to improve the performances of classical CBIR system using a fuzzy logic approach to reinforce the user feedback.

From a knowledge point of view, the proposed approach tries to improve the high semantic level knowledge represented by the user feedback with an automatic extracted knowledge obtained with a FCM algorithm. The user knowledge is used to define the stimulus of the multi-image query, while the partition matrix and the clusters obtained by using the FCM are used to modify the weights of the distance function.

The proposed system was applied to the Columbia Coil-20 image database because it was used in various previous works allowing for a performance comparison among various systems. The obtained results have shown how the proposed system achieves an average precision increment of 15%.

References

1. Flickner, M., et al.: Query by Image and Video Content: The QBIC System. IEEE Computer 28(9) (1995)
2. Pentland, A., Picard, R.W., Sclaroff, S.: Photobook: Content-based Manipulation of Image Databases. In: Proc. SPIE Storage Retrieval Image Video Databases II, pp. 34–37 (1994)
3. Yumin, T., Lixia, M.: Image Retrieval Based on Multiple Features Using Wavelet. In: Proceedings of the 5th Int. Conf. on Computational Intelligence and Multimedia Applications, Xi'an China, pp. 137–143 (2003)
4. Bae, H.J., Jung, S.H.: Image Retrieval Using Texture Based on DCT. In: Proc. of Int. Conf. on Information, Comm. and Signal Processing, Singapore, pp. 1065–1068 (1997)
5. Gevers, T., Smeulders, A.W.M.: Image Search Engines: An Overview. In: Medioni, G., Kang, S.B. (eds.) Emerging Topics in Computer Vision. Prentice Hall, Englewood Cliffs (2004)
6. Amato, A., Calabrese, M., Di Lecce, V.: Relevance Feedback Oriented Cbir Interface For Semantic Discovery. Wseas Transactions on Computers 9(5), 1978–1985 (2006)
7. Santini, S., Jain, R.: The 'El Niño' Image Database System, In: IEEE International Conference on Multimedia Computing and Systems, Florence, Italy (1999)
8. Iqbal, Q., Aggarwal, K.: Feature Integration, Multi-image Queries and Relevance Feedback in Image Retrieval. Invited Paper. To Appear, 6th International Conference on Visual Information Systems (VISUAL). Miami, Florida. pp. 24–26 (2003)
9. Boujemaa, N., Fauqueur, J., Ferecatu, M., Fleuret, F., Gouet, V., Le Saux, B., Sahbi, H.: Ikona: Interactive Generic and Specific Image Retrieval. In: International 24 Workshop on Multimedia Content-Based Indexing and Retrieval (MMCBIR), Rocquencourt, France, pp. 25–28 (2001)
10. Lay, J.A., Guan, L.: Image Retrieval Based on Energy Histograms of the Low Frequency DCT Coefficients. In: IEEE Int. Conf. on Acoustics, Speech and Signal Processing, Phoenix USA, vol. 6, pp. 3009–3012 (1999)
11. Loupias, E., Sebe, N., Bres, S., Jolion, J.M.: Wavelet-based Salient Points for Image Retrieval. In: Int. Conf. on Image Processing. Vancouver BC Canada, vol. 2, pp. 518–521, 10–13 (2000)
12. Sikora, T.: The MPEG-7 Visual Standard for Content Description: an Overview. IEEE Transactions on Circuits and Systems for Video Technology 11(6), 696–702 (2001)
13. Zhang, D.: Improving Image Retrieval Performance by Using Both Color and Texture Features. In: Third Int. Conf. on Image and Graphics, Hong Kong, pp. 172–175 (2004)
14. Dorairaj, R., Namuduri, K.R.: Compact Combination of MPEG-7 Color and Texture Descriptors for Image Retrieval. In: Thirty-Eighth Asilomar Conference on Signals, Systems and Computers, Pacific Grove California, vol. 1, pp. 387–391, 7–10 (2004)
15. Gevers, T., Smeulders, A.W.M.: PicToSeek: Combining Color and Shape Invariant Features for Image Retrieval. IEEE Transactions on Image Processing 9(1), 102–119 (2000)
16. Frigui, H.: MembershipMap: Data Transformation Based on Granulation and Fuzzy Membership Aggregation. IEEE Transactions on Fuzzy Systems 14(6), 885–896 (2006)
17. Liew, A.W.C., Leung, S.H., Lau, W.H.: Segmentation of Color Lip Images by Spatial Fuzzy Clustering. IEEE Transactions on Fuzzy Systems 11(4), 542–549 (2003)
18. Eschrich, S., Hall, L.O., Ke, J.W., Goldgof, D.B.: Fast Accurate Fuzzy Clustering through Data Reduction. IEEE Transactions on Fuzzy Systems 11(2), 262–270 (2003)
19. Amato, A., Di Lecce, V.: A Knowledge Based Approach for a Fast Image Retrieval System. Image and Vision Computing 26(11), 1466–1480 (2008)

20. Di Lecce, V., Guerriero, A.: A Comparative Evaluation of Retrieval Methods for Duplicate Search in Image Database. Journal of Visual Languages and Computing 12, 105–120 (2001)
21. Shen, X., Boutell, M., Luo, J., Brown, C.: Multi Label Machine Learning and its Application to Semantic Scene Classification. In: Proceedings of the 2004 International Symposium on Electronic Imaging, pp. 18–22 (2004)
22. Nene, S.A., Nayar, S.K., Murase, H.: Columbia Object Image Library (coil-100), Tech. Rep., Department of Computer Science, Columbia University (1996),
    `http://www.cs.columbia.edu/CAVE/`
23. Le Saux, B., Boujemaa, N.: Unsupervised Robust Clustering for Image Database Categorization. In: Proc. 16th Int. Conf. on Pattern Recognition, Quebec, Canada, pp. 259–262 (2002)
24. Wang, Z., Feng, D., Chi, Z.: Comparison of Image Partition Methods for Adaptive Image Categorization Based on Structural Image Representation. In: IEEE 8th Int. Conf. on Control, Automation, Robotics and Vision, Kunming, China, pp. 676–680 (2004)
25. Müller, H., Marchand-Maillet, S., Pun, T.: The truth about corel - evaluation in image retrieval. In: Lew, M., Sebe, N., Eakins, J.P. (eds.) CIVR 2002. LNCS, vol. 2383, pp. 36–45. Springer, Heidelberg (2002)
26. Pedrycz, W., Amato, A., Di Lecce, V., Piuri, V.: Fuzzy Clustering with Partial Supervision in Organization and Classification of Digital Image. IEEE Trans. on Fuzzy System. 16(4), 1008–1026 (2008)

Dimension Reduction Using Semi-Supervised Locally Linear Embedding for Plant Leaf Classification

Shanwen Zhang[1] and Kwok-Wing Chau[2]

[1] Institute of Intelligent Machine, Chinese Academy of Science,
P.O. Box 1130, Hefei, Anhui, 230031, China
[2] Department of Civil & Structural Engineering, The Hong Kong Polytechnic University
`wjdw716@163.com`

Abstract. Plant has plenty use in foodstuff, medicine and industry, and is also vitally important for environmental protection. So, it is important and urgent to recognize and classify plant species. Plant classification based on leaf images is a basic research of botanical area and agricultural production. Due to the high nature complexity and high dimensionality of leaf image data, dimensional reduction algorithms are useful and necessary for such type of data analysis, since it can facilitate fast classifying plants, and understanding and managing plant leaf features. Supervised locally linear embedding (SLLE) is a powerful feature extraction method, which can yield very promising recognition results when coupled with some simple classifiers. In this paper, a semi-SLLE is proposed and is applied to plant classification based on leaf images. The experiment results show that the proposed algorithm performs very well on leaf image data which exhibits a manifold structure.

Keywords: Plant leaf image, Plant classification, locally linear embedding, Supervised locally linear embedding (SLLE), Semi-SLLE.

1   Introduction

Plants are the largest number of species and the most extensive form of life on earth. They are the most important genetic resources for human's survival and development and also the necessity of mankind living and production. Furthermore, plants play a crucial role of inhibiting desertification, improving the climate and Maintaining water and soil. According to the investigation, there are almost 300,000 kinds of plants on earth, among these, about 250,000 can be named and recorded by botanists. China is the world's second largest resource base of plants, only the higher plant has more 35,000 kinds of species. Along with the growing production activities of mankind, the ecosystem has being destroyed continuously and a lot of species have to face extinction, it has about 34,000 kinds of plants are on the edge of dying out, this was mentioned in the dissertation about plants' distribution, which was issued on Science in 2002 by doctor Alan K. Knapp [1]. And this point was supported by doctor J. Alan Pounds in 2004 with an article issued on Nature [2]. The extinction of a huge number of plants may have a serious impact on human beings and ecosystem. More and more evidences have made it clear that natural and artificial ecosystems are changing their

functions with the disappearance of biological variability. Wherefore, it is urgent to protect plant species, and an important aspect for the protection is plant classification.

Plants exist everywhere we live, as well as places without us. Many of them carry significant information for the development of human society. Whether for agriculture informatization or for ecological protection, study of plant classification is very necessary. Classification of plants is the basic research of botanical research and agricultural production. It has a vital signification for identifying and distinguishing the types of plants, exploring the relationship among plants, protecting plant species and the environment, avoiding further damage to the ecology, and for clarifying the evolution of the law in Plant system. Classification of plants can also set up databases and plants flora figure of plant resources, it is able to protect plant species variability, and, at the same time, it will allow more non-professionals to get to know plant resources, to enhance the public awareness of the plant resource protection. The urgent situation is that many plants are at the risk of extinction. So it is very necessary to set up a database for plant protection [2-6]. We believe that the first step is to teach a computer how to classify plants.

Comparing with the plant root, stem, flower, fruit and leather, the plant leaves in general on the shape of the structure are very stable and essentially flat in the state, suitable for two-dimensional image processing. In most of the time the plant leaves can be easily collected. For the leaves, their color, shape, texture and other characteristics can be used as a basis for classification. From the perspective of plant taxonomy, or images from the computer pattern recognition point of view, the plant classification based on leaf images is not only a most direct and simple and most effective method, but also the future of digital plant study of a natural trend.

Because the plant leaf images are more sensitive to the seasons, locations and illumination conditions, the leaf appearances do not have as regular size and shape as the industrial parts, and can in no way be described, many existing dimensional reduction methods are not effectively applied to plant classification based on leaf images.

Manifold learning is a very effective non-linear feature extraction and selection method, and it has been applied successfully in biometric identification technology, such as face and lip biometric identification. Because of the image performance of Plant leaves at the season, location and illumination conditions has the same performance characteristics of people face and lips at the expression, it is similar to the leaves as if the plants are the "face" or "lip", there is no doubt that manifold learning methods can also be applied to images based on the leaves of plant classification.

LLE is a nonlinear dimensionality reduction method. The main assumption behind LLE is that the data is sampled from a manifold, embedded in the high dimensional space (HDS). LLE is an unsupervised method, which avoids the local minima problems plaguing many competing methods. Some other advantages of LLE are that few parameters need to be set (selecting optimal values for these is discussed in [7,8]) and that the local geometry of high-dimensional data is preserved in the embedded space. To extend the concept of LLE to multiple manifolds, each representing data of one specific class, a supervised LLE (SLLE) was independently proposed in [7]. In this paper, Semi-SLLE is presented and applied to the plant classification based on leaf images.

The rest of the paper is organized as follows. Section 2 proposes a semi-supervised locally linear embedding. Section 3 introduces K-nearest neighbor classifier. Experimental results are illustrated in Section 4. Conclusion is given in Section 5.

2   Semi-Supervised Locally Linear Embedding

The Euclidean distance is often taken as a measure of the dissimilarity. If the Euclidean distance between two points is very large, we can say the probability of their dissimilarity is very high. Otherwise, they probably are similar to each other. The main purpose of manifold learning is to find the intrinsic geometry of the data, as captured in the geodesic manifold distances between all pairs of data points. The approximation of geodesic distance is divided into two cases. In case of neighboring points, Euclidean distance in the input space provides a good approximation to geodesic distance. In case of faraway points, geodesic distance can be approximated by adding up a sequence of "short hops" between neighboring points.

The novel distance metric introduced in Semi-SLLE algorithm is defined by

$$d(X_i, X_j) = \frac{\left\| X_i - X_j \right\|}{\sqrt{M(i) \cdot M(j)}} \tag{1}$$

where $\left\| X_i - X_j \right\|$ is Euclidean distance of X_i and X_j, $M(i)$ is the mean distance of X_i and X_j.

Based on the novel distance metric, we propose a semi-SLLE. The steps are described as follows [9,10]:

Step 1: For each data point $X_i \in X$, find the set Δ_i of k nearest neighbors of X_i by K-NN algorithm or $\varepsilon-ball$ algorithm in the geodesic distance metric. But in practice, it is hard to choose a proper ε -neighborhood. To avoid this difficulty, the method of k-nearest neighbors is always used instead of the δ-neighborhood in real-world applications.

Step 2: Define the distance metric by Eq.(1).

Step 3: Redefine the distance metric as follows,

$$D(X_i, X_j) = \begin{cases} \sqrt{1 - \exp(-\dfrac{d^2(X_i, X_j)}{\beta})} - \alpha, & \text{if } c_i = c_j \\[4mm] \sqrt{1 - \exp(-\dfrac{d^2(X_i, X_j)}{\beta})} & , \text{ if } x_i \text{ or } x_j \text{ is unlabeled}, \\[2mm] & \qquad x_i \in kNN(j) \text{ or } x_j \in kNN(i) \\[4mm] \sqrt{\exp(\dfrac{d^2(X_i, X_j)}{\beta})} , & \text{otherwise} \end{cases} \tag{2}$$

where β is a control parameter, which should depend on the "density" of the dataset, α is tuning parameter.

Step 4: Re-compute the k-nearest neighbors of each point by Eq.(2).

Step 5: Compute the reconstruction weights W_{ij} of the neighbors that minimize the error ε_i of reconstructing X_i, i.e., seek the best reconstruction weights.

$$\varepsilon_i(W) = \arg\min \left\| X_i - \sum_{X_j \in \Delta_i} W_{ij} X_j \right\|^2 \tag{3}$$

where $\|\bullet\|$ is the Euclidean norm. W_{ij} subject to two constraints, $\sum_{X_j \in \Delta_i} W_{ij} = 1$ and $W_{ij} = 0$ for any $X_j \notin \Delta_i$. In the first constraint, each point is represented as a convex combination of its neighbor; the second reflects that SLLE is a *local* method.

Step 6: Compute the low-dimensional embedding Y for X that best preserves the local geometry represented by the reconstruction weights. This means to solve the below equation (4),

$$\varepsilon(Y) = \arg\min \sum_{i=1}^{n} \left\| Y_i - \sum_{X_j \in \Delta_i} W_{ij} Y_j \right\|^2 \tag{4}$$

where Y_i subject to two constraints $\sum_{i=1}^{n} Y_i = 0$ and $\frac{1}{n}\sum_{i=1}^{n} Y_i Y_i^T = I$ (normalized unit covariance), 0 is a column vector of zeros and I is an identity matrix.

Based on the matrix W, we can define a sparse, symmetric, and positive semi-definite matrix M based on $M = (I - W)^T (I - W)$, the *Step* 6 is transformed to the solution of eigenvector decomposition which can be seen as follows:

$$\varepsilon(Y) = \arg\min \sum_i \left\| Y_i - \sum_{X_j \in \Delta_j} W_{ij} Y_j \right\|^2 \tag{5}$$

$$= \arg\min \|(I-W)Y\|^2 = \min tr(YMY^T)$$

In equation (6), by the Rayleigh-Ritz theorem [11], minimizing equation (4) can be performed by finding the eigenvectors with the smallest (nonzero) eigenvalues of the sparse matrix M. This corresponds to solving a sparse eigenvalue problem, using singular value decomposition (SVD) without performing a full diagonally matrix, to find the $d+1$ eigenvectors of M with the smallest $d+1$ eigenvalues.

3   K-Nearest Neighbor (K-NN) Classifier

Generally speaking, the number of samples (n) must be larger than the number of features (p) for good classification performance. The idea in feature selection is to find the ones with better discrimination ability and design a classifier using those features only. The K-NN classifier is a well-known nonparametric classifier. It is based on a simple and effective supervised classification technique [12]. K-NN is a most common and nonparametric method. To classify an unknown sample x, K-NN extracts k closest vectors from the training set using similarity measures, and makes decision for the label

of the unknown sample x using the majority class label of the k nearest neighbors. We adopt Euclidean distance to measure the similarity of samples. To classify a new input X, the K-NN is retrieved from the training data. Given an input vector X, K-NN extracts k closest vectors in the reference set based on similarity measures, and the input X is then labeled with the majority class label corresponding to the k-NN.

Pearson's coefficient correlation and Euclidean distance have been used as the similarity measure. When we have an input X and a reference set $D = \{d_1, d_2, ..., d_N\}$, the probability that X may belong to class c_j, $P(X, c_j)$ is defined as follows:

$$P(X, c_j) = \sum_{d_i \in kNN} S(X, d_i) P(d_i, c_j) - b_j \qquad (6)$$

where $S(X, d_i)$ is the similarity between X and d_i, and b_j is a bias term.

4   Experiments

In this project, we use 1-NN as a classification tool for its simplicity to verify the effectiveness of our feature reduction algorithm. The whole program is implemented in Matlab (version 6.5) environment.

Multi-class classification: In case of multi-class problem, we split the problem into k binary-classification problem, where k equals to the number of class, and each class is classified versus all other classes in the dataset. The classification accuracy is calculated by adding up the correctly classified samples over all classes.

Ten-fold cross validation: Cross validation is such a technique based on methods of re-sampling. On each data set, ten times ten-fold cross validation is run. That is, in each time, the original data set is randomly divided into ten equal-sized subsets while keeping the proportion of the instances in different classes. Then, in each fold, one subset is used as testing set and the union of the remaining ones is used as training set. After ten folds, each subset has been used as testing set once. One tenth of samples have not labels in training set. The average result of these ten folds is recorded. This procedure is repeated ten times and gets ten results for each compared algorithm. We will estimate the generalized classification results using ten-fold cross validation.

The SLLE is implemented using the coding provided by the original author. We select 20 species of plants, and each plant has 20 images of different leaf in our image database, as shown in Fig.2.

Fig. 1. 20 kinds of plants

These images were recorded at a resolution of 640×480 pixels, with a bit depth of 16 bits/pixel. Thus, 256 levels were available for each R, G, and B color plane. The colors of plant leaves are usually green. Moreover, the shades and the variety of changes of water, nutrient, atmosphere and season can cause change of the plant leaf color, so the color feature has low reliability. Thus, we decided to recognize various plants by the grey-level image of plant leaf before processing. Fig.2 shows 20 original leaf images of Katsura-tree.
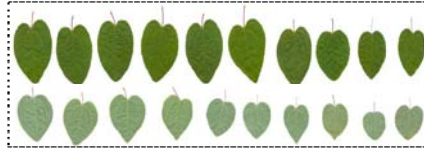

Fig. 2. 20 leaf images of Katsura-tree in different cases

When selecting the optimal neighborhood k and dimensionality d, it is crucial to determine k before d. The value of k is only used while performing the front step of the proposed algorithms, which determines the optimal weights for the nearest neighbors, while d is used in the final step of the proposed method. Too large k will cause elimination of small-scale structures in the manifold. In contrast, too small k may falsely divide the continuous manifold into disjointed sub-manifolds. So the optimal k for a given data first must be determined using an arbitrary value for d; then d is determined by using the optimal k value found in the previous experiment. The number of nearest neighbor k can be set to $k=m-1$, where m is the number of the training samples per class. The justification for this choice is that each sample should connect with the remaining $m-1$ samples of the same class.

In the first experiment, we choose 12 images from each class as training set, the rest as test set. We test the impact of k on the performance of Semi-SLLE. The maximal classification rates are shown in Fig.3, when $k=11$, the classification result attains the optimal value. In order to validate the effect of the proposed method, Table 1 shows the classification results by the region-based classification method [13], LLE [14] and supervised LLE (SLLE) [15].
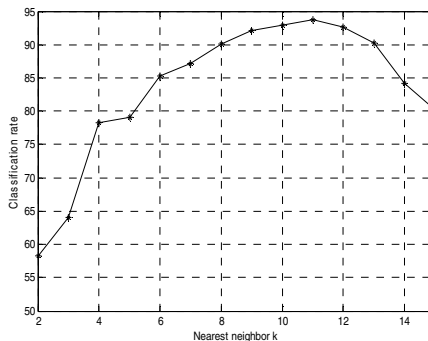

Fig. 3. Classification rates with varied k on leaf images by SLLE, where α =0.1, β =200

Table 1. Recognition rates of leaf images

Method	Reference [13]	LLE+K-NN [14]	SLLE+K-NN[15]	Semi-SLLE+K-NN
Recognition rate	85.86	86.45	88.78	90.26

From Fig.3 and Table 1, it is found that the proposed method outperforms the other methods.

5   Conclusion

This paper introduces a feature extraction approach for plant leaf recognition. The classifier K-NN is adopted for it has fast speed on training and simple structure. The features of plant leaf are extracted and processed by semi-SLLE to form the input vector of K-NN. The experimental results indicate that our algorithm is workable with a recognition rate greater than 90% on five kinds of plants. Compared with other methods, this algorithm is fast in execution, efficient in recognition and easy in implementation. Future work is under consideration to improve it. Future work should be directed to the following topics. (1) More difference leaf images in variety cases should be collected to generalize the proposed classification method. (2) More features of leaf should be included to improve the classification performance of the proposed method. Further study should be study a supervised robust feature extraction algorithm for plant classification. (3) The features of different leaf classes in the world should be taken into account to improve the classification accuracy. (4) The proposed method should be extended to handle distortion problems such as broken leaves and oriented leaves.

Future work should be directed to the following topics.

(1) More leaf images should be collected to generalize the proposed classification algorithm.

(2) More features of leaf should be included to improve the classification performance of the proposed method. For example, the texture of nervure can provide useful information. This study should extend to use shape and texture features for data classification.

(3) The occurrences of different leaf classes in the world should be taken into account to improve the classification accuracy.

(4) The proposed method should be extended to handle distortion problems such as broken leaves and oriented leaves

Acknowledgment

References

1. Alan, K.K., Philip, A., Fay, et al.: Rainfall Variability, Carbon Cycling, and Plant Species Diversity In A Mesic Grassland. Science 298, 2202–2205 (2002)
2. Alan Pounds, J., Robert, P.: Ecology: Clouded Futures. Nature 427, 107–109 (2004)
3. Du, J.X., Wang, X.F., Zhang, G.J.: Leaf shape based plant species recognition. Applied Mathematics and Computation 185, 883–893 (2007)
4. Ye, Y., Chen, C., Li, C.T., Fu, H., Chi, Z.: A computerized plant species recognition system. In: Proceedings of 2004 International Symposium on Intelligent Multimedia, Video and Speech Processing, Hong Kong (October 2004)
5. Miao, Z., Gandelin, M.H., Yuan, B.: An oopr-based rose variety recognition system. Engineering Applications of Artificial Intelligence 19 (2006)
6. de Oliveira Plotzede, R., Falvo, M., Pdua, J.G., Bernacci, L.C., Oliveira, M.L.C., Bruno, O.M.: Leaf shape analysis using the multiscale minkowski fractal dimension, a new morphometric method: a study with passifliora (passifloraceae). Canada Journal of Botany 83 (2005)
7. de Ridder, D., Duin, R.P.W.: Locally linear embedding for classification. Technical Report PH-2002-01, Pattern Recognition Group, Dept. of Imaging Science & Technology, Delft University of Technology, Delft, The Netherlands (2002)
8. Kouropteva, O., Okun, O., Pietikäinen, M.: Selection of the optimal parameter value for the locally linear embedding algorithm. In: Proc. of the 1st Int. Conf. on Fuzzy Systems and Knowledge Discovery, Singapore, pp. 359–363 (2002)
9. Horn, R.A., Johnson, C.R.: Matrix Analysis. Cambridge University Press, Cambridge (1990)
10. Wang, Y.H., Makedon, F.S., Ford, J.C., Pearlman, J.: HykGene: A Hybrid Approach for Selecting Marker Genes for Phenotype Classification using Microarray Gene Expression Data. Bioinformatics 21(8), 1530–1537 (2005)
11. de Silva, V., Tenenbaum, J.B.: Global Versus Local Methods in Nonlinear Dimensionality Reduction. NIPS, 705–712 (2002)
12. Huang, H., Li, J.W., Feng, H.L.: Face Recognition on Semi-supervised Manifold Learning. Computer Science 35(12), 220–222 (2008)
13. Lee, C.L., Chen, S.Y.: Classification for Leaf Images. In: 16th IPPR Conference on Computer Vision, Graphics and Image Processing, vol. 8, pp. 17–19 (2003)
14. Shi, C., Chen, L.H.: Feature dimension reduction for microarray data analysis using locally linear embedding. APBC162 16, 1–7 (2004)
15. Pillati, M., Viroli, C.: Supervised Locally Linear Embedding for Classification: An Application to Gene Expression Data Analysis. In: Proceedings of 29th Annual Conference of the German Classification Society (GfKl 2005), pp. 15–18 (2005)

DCGene: A Novel Predicting Approach of the Disease Related Genes on Functional Annotation

Yuan Fang[1] and Hui Wang[2]

[1] Shenzhen Institute of Information Technology, Shenzhen 518029, China
[2] Hubei Bioinformatics and Molecular Imaging Key Laboratory, Huazhong University of Science and Technology, Wuhan 430074, China
yuancopper@yahoo.com.cn

Abstract. Disease Candidate Genes (DCGene) is an advanced system for predicting the disease related genes, It is a novel computational approach by using the GO annotation information. The performance of the DCGene is evaluated in a set containing 1057 test samples, on both the local region and genome scale. In the local region scale, for 397 of 1057 (37.6%) samples, the disease-associated genes are at the top 1 of the out put gene prioritization list, and if the top 9 genes are all considered, 754(71.3%) disease-associated genes are included in the result. In the genome scale, 55% of the disease-relevant genes are included in the top scoring 3%, and 74% of the disease-relevant genes are included in the top 15%. The performance of the DCGene is demonstrated to be significant better than the others by comparison with the other systems and methods.

1 Introduction

Identifying the relevant genes of human genetic disorder, is not only the base of diagnosis and prevention of disease [1,2], but is also one of the major goals of the human genome project. The positional candidate strategy has become regular strategy for the identification of relevant genes of genetic disease. But there are hundreds of candidate genes in the located chromosome regions, especially for the linkage studies of complex genetic disorder, whose regions include 300 genes on average[3]. It is a time-consuming and expensive work to identify the gene from the interested regions if all genes are randomly selected for mutation analysis. Therefore, the crucial step is how to predict disease-relevant genes and prioritize the candidate genes for mutation analysis. With the accumulation of various genomic information, computational analyses have become the important methods for prioritizing the candidate genes according to their nosogenetic possibilities.

Some bioinformatics application tools for prioritizing candidate genes have been developed and released to public [4-10] in recent years. They can be classified into two broad, while not mutually exclusive categories. One is based on single type of genomic data (functional annotations, sequences, expression profiles et al.) [4-7], the other is based on the integration of various types of genomic data [8-10]. Franke, et al. [10] observed that GO[11] annotation is the most effective data resource, and

the accuracy based on GO was slightly improved by adding other types ofdata. There-fore, improving the accuracy of function annotation based methods will greatly help improve the accuracy of multi-information fusion methods.

In this study, we attempt to improve the accuracy of predicting disease-relevant genes by efficiently exploiting the GO functional annotation information. And a novel computational tool, called DCGene (Disease Candidate Genes), has been developed. Testing results demonstrated that the performance of DCGene is significantly better than those reported by the above mentioned approaches.

2   Material and Methods

In DCGene, a new approach is proposed to measure the similarity between any given candidate gene and the group of genes certified to be related to some disease, Specifi-cally, we explored how to make use of the structure of DAG to analyze function similarity for improving the performance of identifying disease genes. We search for candidate genes of similar functions with similar but not necessarily shared GO annotations.

2.1   The Functional Similarity between a Candidate Gene and the Cluster of Genes Related to the Disorder

In this study, the determination of functional similarity between a candidate gene and the group of disease –relevant genes is based on GO annotation information. And the definition of the similarity is divided into three different levels: GO-GO level, gene-gene level and gene-cluster level from the basic GO terms.

Functional Similarity at the GO-GO Level
Comparing the functional similarity of two GO terms is the basis for evaluating the functional similarity between two genes and/or two groups of genes. In previous methods[4,7], only shared GO-terms between two genes are considered, it means that two GO terms are considered as 'similar' only if they are identical. Consequently, it is improper to conclude that commonalities of function can not be shared by two different GO-terms in view of the DAG structure characteristic of GO,. For example, GO:0005267(potassium channel activity) and GO:0005262(calcium channel activity) are two different GO-terms, however, both of them have the parent "cation channel activity"(GO:0005261) according to the DAG structure of GO, so obvious functional similarity is existed between GO:0005267(potassium channel activity) and GO:0005262(calcium channel activity). As a result, the functional similarity can not be effectively revealed by the methods only considering the shared GO-terms between two genes . To tackle this problem, in this study, the following method which can exploit the complete DAG information is proposed to assess the functional similarity between two GO terms.

Firstly, the specificity and the ancestor set for each GO-term are determined. For any a GO term O_s, its specificity $O(O_s)$, is defined as:

$$O(O_s) = \frac{G_{NUM}(O_s)}{G_{NUM}(O)} \qquad (1)$$

where $G_{NUM}(O_s)$ is defined as the number of genes that have been annotated with O_s, $G_{NUM}(O)$ is the total number of genes annotated with a GO term belonging to the same category as O_s (Molecular function, biological process, or cellular component). Both $G_{NUM}(O_s)$ and $G_{NUM}(O)$ can be obtained from the GO Annatation database. For instance, the specificity of term GO:0005267, a child of "biological process", is 0.0026 (286/108120), where 286 and 108120 are the numbers of genes that are annotated with GO:0005267 and "biological process", respectively. Obviously, the value of $O(O_s)$ ranges over [0,1], and the lower the value is, the higher the specificity of the GO-term O_s is. The ancestor set of O_s, $O_{ANC}(O_s)$, is defined as a set including all the nodes that can reach O_s according to the DAG of GO. For example, the ancestor sets of node 7 and 12 are $O_{ANC}(7)=\{1,2,3,4,5,6,8,9,10\}$, and $O_{ANC}(12)= \{1,2,3,4,8,9,10,11,13,14\}$.
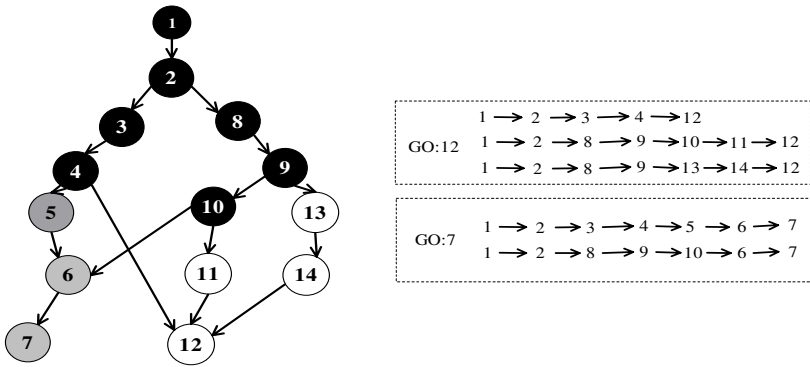

Fig. 1. An example to show the hierarchical nature of GO and the determination of ancestor set of a GO term

Based on the definition of the specificity and the ancestor set of GO-terms, the score that estimates the degree of functional similarity between any two GO-terms, O_s and O_t, can be defined as:

$$m'(O_s,O_t) = -\lg(\min\{O(O_{share})\}) \quad O_{share} \in \{O_{ANC}(O_s) \cap O_{ANC}(O_t)\} \quad (2)$$

where Oshare is the set of GO terms that are shared by the ancestor sets of O_s and O_t. The most specific common parent is used to measure the similarity of two GO-terms in this definition. This method is more seemly than the strategy which only considers the bottommost GO-terms.

The Association at Gene-Gene Level
With the definition of the functional similarity between two GO-terms, the functional similarity between any two genes, G_1 and G_2, can be defined as:

$$m(G_1,G_2) = \max\{m'(O_s \mid G_1, O_t \mid G_2)\} \quad (3)$$

where $\{O_s|G_1\}(s≥1)$ and $\{O_t|G_2\}(t≥1)$ denote the GO sets annotated to G_1 and G_2 respectively.

The Association at Gene- cluster Level

For the gene-cluster levels, functional similarity between a gene and a group of genes is defined as:

$$Score(G, D) = \frac{1}{n} \sum_{j=1}^{n} m(G, G_j) \tag{4}$$

where *G* and *D* denote the gene and the set of genes respectively, and *n* is the number of genes in *D*.

2.2   Prediction of the Disease-Relevant Genes

The procedure to predict the disease-relevant genes will be launched after inputting a set of disease-relevant genes and a set of candidate genes. First of all, calculate the functional similarity between each of the candidate genes and the group of disease-associated genes in different GO-terms categories. Next transform the functional similarity into a reliability value, and get the degree of association with the given disease for each candidate genes. Then rank all of the candidate genes in a descending manner, according to the degree of association.

2.3   Data Set

A data set containing 1057 human disorder entries is used to test the performance of DCGene. Firstly, all disorder entries contained in the OMIM[12] database are obtained from OMIM-morbid map (ftp://ftp.ncbi.nih.gov/repository/OMIM/ morbid-map), next the LocusLink database[13] of NCBI (table mim2loc and loc2ref) was used to screen these disorder entries , as a result, a list of 2715 disorder entries whose causative genes have been identified (September 2004) are acquired. however, the following three types of records are eliminated from this list: (1) 296 certain "non-disease entries ", mainly genetic variations that lead to apparently abnormal laboratory test values; (2) 615 entries whose causative genes can not be found, according to the appointed cytogenetic locations, from the Map Viewer of NCBI; (3) 747 entries that have only one contributing gene with GO annotations and therefore are not suitable for leave-one-out validation. Finally, the data set contains 1057 disorder entries. Each entry in this data set represents one disorder and consists of two parts: (1) a group of known genes associated with the disorder; (2) All the genes, together with their GO annotations, contained in the disease's cytogenetic locations given by OMIM. It is obvious that not all genes in these chromosome regions have been annotated with functional information, but only those genes with GO annotations are considered in our program.

3   Results

3.1   Predicting Disease-Relevant Genes in Their Located Chromosome Regions

A 5-fold cross-validation process was adopted to evaluate the performance of DCGene by using above data set, so as to predict disease-relevant genes in their

located chromosome regions, which is described as follows. (1) The data set containing 1057 disorder entries was randomly divided into five equal-sized partitions; (2) Each of the five partitions was selected as the testing set in turn, and the other four partitions were used to test the DCGene system; (3) For each entry in the testing set, firstly, its genuine nosogenetic gene is eliminated from the group of known relevant genes of this disease to guarantee the authenticity of prediction results; then, their possibilities of being relevant to the disease are scored with the methods given in last section for all the genes that are contained in the located chromosome region of this disease and have GO annotations; (4) finally a prioritized list of candidate genes for this disease can be obtained according to their sequence scores, in which the anterior genes are regarded to have higher possibility being relevant to the disease and might be preferentially selected for mutation analysis. Obviously, the more the genuine nosogenetic genes are predicted with high priority, the better the prediction performance will be. The 5-fold cross-validation results are given in Table 1. Of the 1057 disorder entries whose located chromosome regions contain 89 genes on average, the genuine nosogenetic genes of 397 cases (37.6%) rank first in the predicted lists of candidates, and in 754 cases (71.3%), the true solutions are within the top 9 of predicted candidates.

Table 1. The 5-fold cross-validation results of DCGene to predict the disease-relevant genes in 1057 located chromosome regions.

Test Set	Sample Size	Rank1 Size[top%]	Rank2 Size[top%]	Rank3 Size[top%]	Rank4 Size[top%]
Part 1	212	77(36.3%)	106(50.0%)	119(56.1%)	159(75.0%)
Part 2	212	79(37.3%)	98(46.2%)	114(53.8%)	148(69.8%)
Part 3	211	81(38.4%)	106(50.2%)	123(58.3%)	148(70.1%)
Part 4	211	79(37.4%)	102(48.3%)	114(54.0%)	150(71.1%)
Part 5	211	81(38.4%)	102(48.3%)	119(56.4%)	149(70.6%)
DCGene	1057	397(37.6%)	514(48.6%)	589(55.7%)	754(71.3%)

3.2   Predicting Disease-Relevant Genes on the Genome Scale

The ability of DCGene to predict disease-relevant genes on the genome scale has been tested so as to evaluate the performance of DCGene and compare it with the other approaches,. In this test, the same data set containing 1057 disorder entries and 5-fold cross-validation process as described above are adopted, and for each entry, all of the 12945 human genes that have GO annotations are scored and fell into place of their scores in a descending manner. Two parameters [the filter ratio (*FR*) and sensitivity (*Sn*)] are used to evaluate the performance of DCGene to predict disease-relevant genes on the genome scale. The filter ratio (*FR*) reflects the performance of efficiently selecting candidate disease-relevant genes from a number of genes. If the top *CN* genes in a prioritized gene list are considered as candidate disease-relevant genes, then *FR* can be defined as $FR = CN / GN$, where *GN* is the number of genes in the prioritized gene list (here, $GN = 12945$). The sensitivity (*Sn*) represents the probability that the selected candidates contain the true disease-relevant genes. For a test dataset containing *DN* disorder entries (here, $DN = 1037$), if there are *TN* entries for which the selected candidates contain true disease-relevant genes, then the sensitivity *Sn* can

be defined as $Sn = TN / DN$. Obviously, the smaller the filter ratio *FR* and the higher the sensitivity *Sn* are, the better the prediction performance is on the genome scale. The 5-fold cross-validation results of DCGene to predict disease-relevant genes on the genome scale are shown in Fig. 3. The possibility of genuine nosogenetic gene will be 90% if the top 50 of this DCGene list are regarded as candidated gene, and the possibility of genuine nosogenetic gene will be descend to 55% if the top 10 of this DCGene list were regarded as candidated gene. Consequently, the higher of *Sn* is, the better the prediction performance is, if the prediction performance has the same value of *FR*.
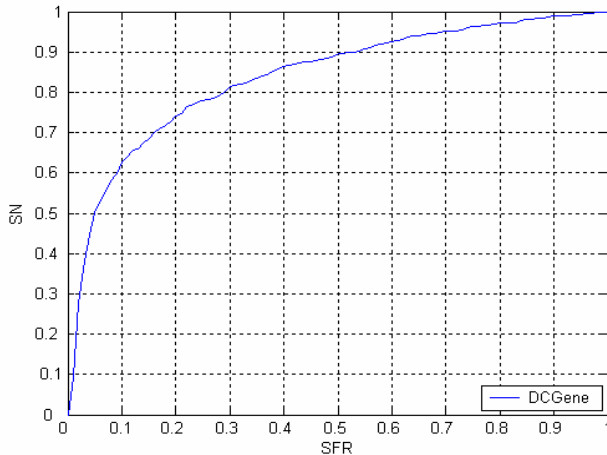

Fig. 2. The performance of DCGene to predict disease-relevant genes on the genome scale

3.3   Comparison with Other Candidate Gene Prediction Methods

As mentioned previously, three computational methods to predict disease-relevant genes using function information have been reported. However, different dataset and measures were used by different authors to evaluate the performance of their approaches. Specifically, Perez-Iratxeta et al. [4,5] used a dataset containing 100 known disease-associated genes to test their program G2D, Turner et. al [6] used a dataset containing 163 disease-associated genes to test their program POCUS, and Freudenberg and Propping [7] tested their method (F-P method) using a dataset containing 851 disease-relevant genes. As a result, we are not able to test and compare the performance of these approaches with a common dataset, because both the POCUS and the F-P method have not provided available programs. The only way is that compare the the result of DCGene's performance obtained with our dataset with the reported results of other researches. Furthermore, this indirect comparison does make sense to some extent, because the dataset used here contains more disease-relevant genes and would largely overlap the datasets mentioned above.

As the disease gene prediction program G2D was tested with 30 Mb chromosome regions nearby 100 known disease-relevant genes, the reported accuracy of G2D is

that about 47% of disease-relevant genes are among the 8 best scoring genes and about 62% of disease-relevant genes are among the 30 best scoring genes. According to the average distribution density of known human genes, it is estimated that there are about 300 genes in a 30 Mb chromosome region. The prediction accuracy of G2D is equivalent to the statistic that 47% of disease-relevant genes can be found within top 2.67% (8/300) of the prediction results, and 62% of disease-relevant genes can be found within top 10% (30/300). From Table 1, the accuracy of DCGene is that 48.6% of disease-relevant genes are within top 2.2% (2/89) of the prediction results, and 71.3% of disease-relevant genes are within top 10% (9/89). Obviously, this prediction accuracy is significantly better than that of G2D.

Freudenberg and Propping tested their F-P method on the genome scale by using a dataset containing 851 disease-relevant genes and 10627 candidate genes,. It is reported that 33.4% (284/851) of the disease-relevant genes were contained within the top scoring 3% of candidates (i.e. $Sn = 33.4\%$, $FR = 3\%$), , and 66.7% (568/851) of the disease-relevant genes were contained within the top scoring 15% of candidates (i.e. $Sn = 66.7\%$, $FR = 15\%$). The prediction sensitivity Sn of DCGene are 55% and 74%, respectively , if $FR = 3\%$ and $FR = 15\%$ (Fig.3). Obviously, the prediction accuracy of DCGene on the genome scale is much higher than that of the F-P method.

4   Discussion

4.1   The Performance Analysis about the DCGene

The strong similarity in functional annotation is existed in several genes in those different phenotypes of the same disorders, especially for those oligogenic diseases, As the involved proteins may interact directly or act indirectly at different positions or stages of the same pathway in these oligogenic disorders,so these coding genes may show certain degree of similarity in functional annotation [14]. DCGene can provide more powerful information of disease-relevant genes for those oligogenic diseases. For example, eight disorder entries of xeroderma pigmentosum [15], including ADPRT (MIM:173870), XPA (MIM:278700), XPC (MIM:278720), ERCC2 (MIM:126340), DDB2 (MIM:600811), DDB1 (MIM:600045), ERCC5 (MIM:133530), POLH (MIM:603968) are tested by the DCGene. The prediction result of level-one-out test shows that seven of these genes are at the first scoring of candidates, and only one is at the second scoring. The resean is due to those seven genes are similar with each other in the molecular function (i.e. the binding function of DNA, pathway of repairing of DNA). The similar prediction result can be obtained in the other oligogenic diseases, such as Bardet-Biedl syndrome, non-syndromic recessive deafness, Waardenburg syndrome.

However, prediction result seems to be poor in those complicated diseases. The reason maybe attributed to the poor functional similarity among their relevant genes of those complicated diseases. for example, in six relevant genes: DYT1 (MIM:605204), GCH1 (MIM:600225), SGCE (MIM:604149), TH (MIM:191290), DRD2 (MIM:126450), DRD5 (MIM:126453)[16], five genes all take par in the process of cell surface receptor linked signal transduction and synaptic transmission, only

DYT1 gene is quite different from the other five genes, therefore the rank of DYT1 gene is very low at the 106th of 139 candidates.

4.2 The Limitation of DCGene

Only the genes, which have been annotated by the GO-terms could be predicted by the DCGene due to the nature of its method. However, the annotation of the gene function is still not completed, furthermore, some of genes have not any function annotation, besides the annotation might be uncompleted for some annotated genes, As a result, this DCGene would be much more efficient in prediction of disease-relevant genes with the improvement of the GO annotation,. On the other hand, there are many other kinds of information such as the gene expression[17,18] and sequence information, which have been used in the prediction of disease-relevant genes[19,20]. Therefore, colligation of these different kinds information might improve the performance of prediction of disease-relevant genes.

Acknowledgement

References

1. Lander, E.S., Linton, L.M., et al.: Initial Sequencing and Analysis of the Human Genome. Nature 409, 860–921 (2001)
2. Venter, J.C., Adams, M.D., Myers, E.W., et al.: The Sequence of the Human Genome. Science 291, 1304–1351 (2001)
3. McCarthy, M.I., Smedley, D., Hide, W.: New Methods for Finding Disease-susceptibility Genes: Impact and Potential. Genome Biology 4, 119 (2003)
4. Perez-Iratxeta, C., Bork, P., Andrade, M.A.: Association of Genes to Genetically Inherited Diseases Using Data Mining. Nature Genetics 31, 316–319 (2002)
5. Perez-Iratxeta, C., Matthias, W., Bork, P., et al.: G2D: A Tool for Mining Genes Associated with Disease. BMC Genetics 6, 45 (2005)
6. Turner, F.S., Clutterbuck, D.R., Semple, C.A.: POCUS: Mining Genomic Sequence Annotation to Predict Disease Genes. Genome Biology 4, R75 (2003)
7. Freudenberg, J., Propping, P.: A Similarity-based Method for Genome-wide Prediction of Disease-relevant Human Genes. Bioinformatics 18(Suppl. 2), S110–S115 (2002)
8. Adie, E.A., Adams, R.R., Evans, K.L., Porteous, D.J., Pickard, B.S.: SUSPECTS: Enabling Fast and Effective Prioritization of Positional Candidates. Bioinformatics 22, 773–777 (2006)
9. Aerts, S.: Gene Prioritization through Genomic Data Fusion. Nat. Biotechnol. 24, 537–544 (2006)
10. Franke, L., Bakel, H., Fokkens, L., Jong, E.D., Egmont-Petersen, M., Wijmenga, C.: Reconstruction of a Functional Human Gene Network, with an Application for Prioritizing Positional Candidate Genes. Am. J. Hum. Genet. 78, 1011–1025 (2006)

11. Ashburner, M., et al.: Gene Ontology: Tool for the Unification of Biology. Nature Genetics 25, 25–29 (2000)
12. Hamosh, A., Scott, A.F., Amberger, J.S.: Online Mendelian Inheritance in Man (OMIM), a Knowledgebase of Human Genes and Genetic Disorders. Nucleic Acids Research 33, D514–D517 (2005)
13. Pruitt, K.D., Maglott, D.R.: RefSeq and LocusLink: NCBI Gene-centered Resources. Nucleic Acids Res. 29(1), 137–140 (2001)
14. Botstein, D., Risch, N.: Discovering Genotypes Underlying Human Phenotypes: Past Successes for Mendelian Disease, Future Approaches for Complex Disease. Nature genetics supplement 33, 228–237 (2003)
15. Lehmann, A.R.: The Xeroderma Pigmentosum Group D(XPD) Gene: One Gene, Two Functions, Three Diseases. Genes & Development 15, 15–23 (2001)
16. Nemeth, A.H.: The Genetics of Primary Dystonias and Related Disorders. Brain 125, 695–721 (2002)
17. Van Steensel, M.A., Celli, J., Van Bokhoven, J.H., et al.: Probing the Gene eXpression Database for Candidate Genes. European Journal of Human Genetics 7, 910–919 (1999)
18. Van Driel, M.A., Cuelenaere, K., Kemmeren, P.P., et al.: A New Web-based Data Mining Tool for the Identification of Candidate Genes for Human Genetic Disorders. European Journal of Human Genetics 11, 57–63 (2003)
19. Kondrashov, F.A., Ogurtsov, A.Y., Kondrashov, A.S.: Bioinformatical Assay of Human Gene Morbidity. Nucleic Acids Res. 32(5), 1731–1737 (2004)
20. Lopez-Bigas, N., Ouzounis, C.A.: Genome-wide Identification of Genes Likely to be Involved in Human Genetic Disorder. Nucleic Acids Res. 32(10), 3108–3114 (2004)

Image Processing Framework for Virtual Colonoscopy

Vitoantonio Bevilacqua[1,2,*] , Marco Cortellino[1,2], Michele Piccinni[1], Antonio Scarpa[1],
Diego Taurino[1], Giuseppe Mastronardi[1,2], Marco Moschetta[3],
and Giuseppe Angelelli[3]

[1] Politecnico di Bari - Dipartimento di Elettrotecnica ed Elettronica
Via Orabona, 4 – 70125 Bari, Italy
bevilacqua@poliba.it
[2] E.B.I.S. s.r.l. (electronic Business In Security), Spin-Off of Polytechnic of Bari,
Via Pavoncelli, 139 – 70125 Bari – Italy
[3] Di. M. I. M. P. - Sezione di Diagnostica per Immagini
Policlinico Universitario
Piazza Giulio Cesare 11 – Bari, Italy

Abstract. This paper describes a complete image processing framework for Virtual Colonscopy. The developed algorithms cover the entire process that allows a virtual navigation inside the colon lumen, starting from a dataset of axial CT slices. The implemented modules are: electronic colon cleansing, lumen segmentation, skeletonization, rendering and navigation. In particular for the centerline problem two different techniques are proposed and evaluated.

Keywords: Virtual Colonscopy, Segmentation, Colon cleansing, Skeletonization, Rendering, Image processing, CT Colonography.

1 Introduction

Colorectal cancer (CRC) is the third most common human malignancy and the second leading cause of cancer-related deaths in the United States [1]. Recently, *virtual colonoscopy* (VC) also refereed as CT-Colonography has been developed as an non-invasive method to evaluate the colon for polyps, that can degenerate to cancer. This new technology requires advanced image processing techniques that allow radiologists to virtually navigate through a patient's colon model reconstructed from helical abdominal computed tomography (CT) images. This paper presents a complete framework for image processing in VC. The first addressed problem is the removal of oral contrast fluid, that can be used to improve the interpretation of slices, then segmentation of the colon lumen is performed using morphological and threshold operations guided by prior anatomical knowledge. For the centerline extraction two different approaches are proposed: penalized distance based, and layered based. We have introduced a new metric for performance evaluation based on the mean distance from the colon boundary. Finally a virtual navigation interface has been implemented using the well know marching cube algorithm and surface rendering.

---

* Corresponding author.

2   Dataset and Library

The dataset of images that we use to test implemented algorithms is the public available Walter Reed Army Medical Center collection, that contains series of slice with resolution of 512x512 in DICOM format. We code the software in C++, with the support of the libraries ITK (Insight Segmentation and Registration Toolkit) and VTK (Visualization Toolkit).

3   Fluid Removal

To improve bowel preparation and reduce the misinterpretation probability, a contrast enhancing fluid (CEF) could be given to the patient to tag stool residual.

Tagged material should be subtracted from CT slices before the segmentation step: this operation is commonly referred as Electronic Colon Cleansing (ECC). A number of techniques are proposed for ECC: Lakare et al. [2] has employed a method called *Segmentation Rays*, in [3] a series of morphological operations and non linear transfer functions are used, a supervised learning strategy called *probabilistic boosting tree* is used in [4]; and in [5] a Gaussian mixture model is constructed in order to classify voxel materials, and a theoretical three transition model is studied in [6]. Our method is designed to produce clear results without using excessively complex algorithms.
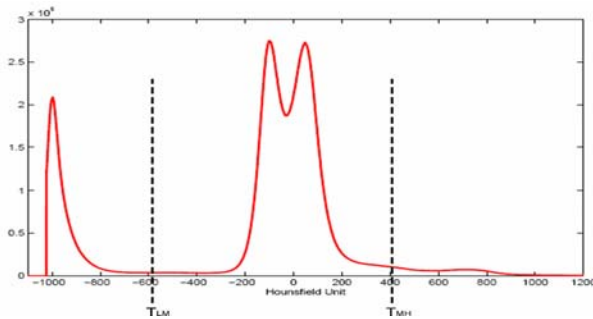

Fig. 1. Typical CT scan histogram with material threshold

Figure 1 shows a typical histogram profile of a CT scan in which, three main regions are recognizable: low values (air), medium values (fat and muscle) and high values (CEF and bones). Each region can be characterized by a peak value T_L, T_M and T_H, and two thresholds can be calculated as median peak values. This threshold classification is valid only if Partial Volume Effect (PVE) is not too influent: that is, on voxels in a homogeneous region. For transition voxels we first extract them with Canny operator [8], and for every boundary voxel a short linear neighborhood has been sampled in the gradient direction. The first and last voxel, referred to as M1 and M2, are classified with histogram threshold values and the vertical component of the gradient Δy is used for transition characterization:

a) $\Delta y < 0$, M1 = Air, M2 = Fluid → Air-Fluid transition
b) $\Delta y > 0$, M1 = Tissue, M2 = Fluid → Tissue-Fluid transition

Tissue-Air transition is not considered because it is not affected by the CEF removal algorithm. For transition types a) and b), two non-linear transfer functions are used to correct the PVE before fluid removal (Fig. 2).
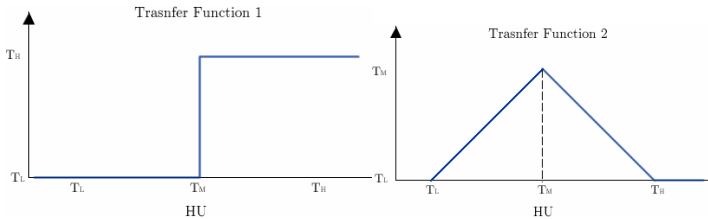

Fig. 2. Voxel transfer function for air-fluid transition (left) and fluid-tissue transition (right)
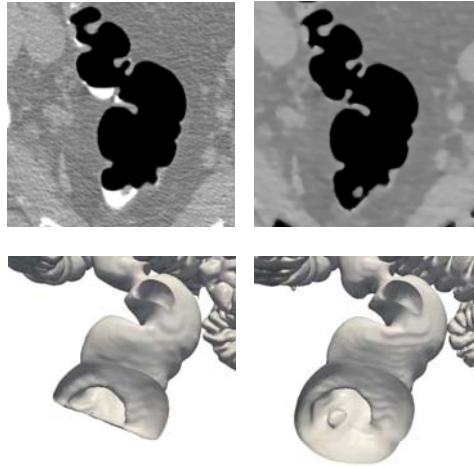

Fig. 3. An example of application of the cleansing algorithm the first row shows a 2D slice with a polyp before and after the electronic cleansing, the bottom row shows the 3D model

Once the PVE is corrected, the fluid can be accurately removed from the image by substituting the high intensity voxels with air voxels (around -1000 HU). One last post-process step is carried out in order to remove residual artifacts and consists of double noise filtering: morphological and anisotropic Gaussian [7].

4   Colon Segmentation

Several colon segmentation techniques are proposed in literature: in [19] the image gradient is used to identify colonic walls, since there is a sudden transition from high CT Number corresponding to a tissue and lower CT Number corresponding to colon

lumen. Other algorithms to segment colon, like in [20] and [21], use anatomical information to improve results, for example identifying regions in the volume where certainly there aren't colonic regions. More experimental approaches use geometric deformable models [22] to identify colonic wall or algorithms based on learning techniques [23]. Our approach is similar to [24] where colon segmentation is obtained estimating centerline. First background removal is performed with region growing in the three dimensional space, by placing the seed voxels in the eight vertices of the parallelepiped formed by the volume of images. The Hounsfield range used for growing is between -2000 and -300 HU that corresponds to air voxel.
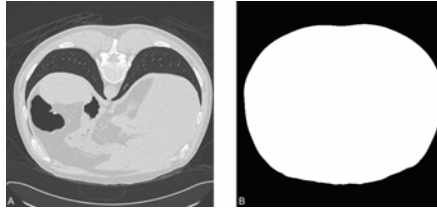

Fig. 4. Binary mask (B) representing the ROI of A, after background removal

For the lung removal, again we use the region growing algorithm with seed voxels placed in the first slice and the HU range between -2000 and -700. Finally air regions are extracted with a threshold filter. The threshold value, was selected analyzing histogram, and is set to -800 HU. Extra colonic material could be still present in segmented slices as shown in fig. 5, these regions have a limited volume compared with colon, so we set a volume threshold to delete region too small to be considered part of colon lumen. This operation, however, doesn't allow to exclude from analysis too large volumes that don't correspond to colon lumen, such as the intestine. Intestine removal is performed considering the anatomical information that is the average length of a colon. Calculation of the colon length can be performed using centerline (see next paragraphs). In this way, we can easily separate colon lumen from intestine and remove any extra-colon areas still present in the binary mask.
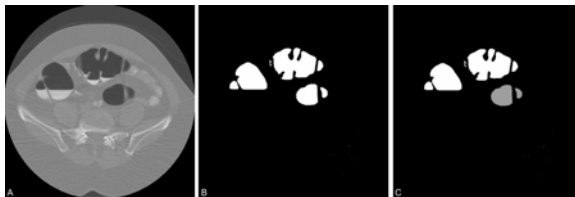

Fig. 5. Applying thresolding filter to the ROI, selected areas are colon lumen, intestine and small noise areas. Disconnected regions are tagged with different values (C)

5   Centerline Extraction

The extraction of a suitable centerline for path navigation is an important step of the VC image processing pipeline. The extracted line is required to be well centered
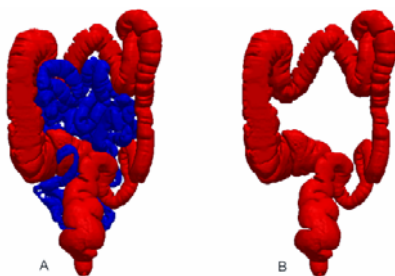
Fig. 6. Two examples in which is visible the improvement of results after segmentation refinement. In A the intestine is visible in blue, then excluded from the final result in B.

inside the lumen, accurate and smooth so that it can be used as guide path for a virtual camera.

The centerline extraction arise in the class of skeletonization problems, as defined in [9], and a lot of methods have been proposed in literature. A good number of approaches uses the notion of *Distance Map Transform* (DMT) as [10][11], a different class of algorithms as [12][13], uses an *onion peeling* approach, also referred as *topological thinning*, that consists in removing iteratively layers of voxel maintaining topological properties, until a single continue line remains. Bouix et al. [14] have proposed a new approach based on the mathematical concept of flux applied to a vector field of Euclidean distance from the object boundary. The well known level set algorithms [15], can also be applied to skeletonization problems, as shown in [16] in which an energy functional that describes the line is minimized through a fast marching approach. Skeletonization algorithms suffers of standardized and universally accepted evaluation methods lack. Common approach involve the use of phantom volumes, or comparison with human traced lines. We propose an intra-algorithm evaluation approach: two different algorithms have been developed and tested on the same datasets. The extracted line are compared, and the similarity between them could be a good indicator of both algorithms soundness.

6   Penalized Distance Algorithm

The first approach is summarized in Fig. 7, the algorithm consist of several step, including a final segmentation refining that uses the centerline and the anatomy knowledge of the colon to detect non colonic component and remove them. The input of the algorithm is a binary image representing the segmented colon, it is sub-sampled for improving the speed performance. Then for each connected component, the initial and the final point of the centerline are calculated with a double breadth first visit started in an arbitrary point inside the component. The farthest point reached with the first visit is considered the initial point, then starting from it, a new visit is conducted, so the last voxel reached is the final point. Knowing the initial and the final points, the centerline extraction problem is addressed with a modified version of the well known Disjktra algorithm [17], in which a penalty term is included in order to maintain the line far away from colon boundary.
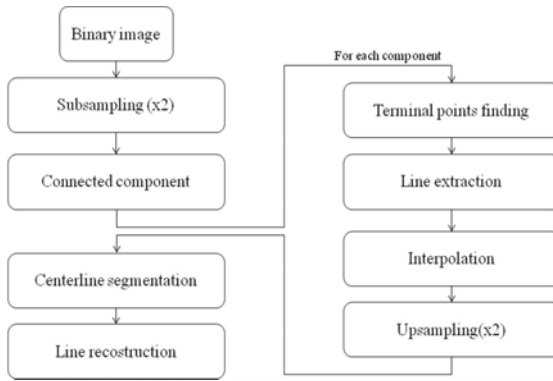
Fig. 7. Penalized distance algorithm flow chart

In particular for every new voxel *v* a penalty term is defined as:

$$\text{Penalty(v)} = N \exp(k*dist(v)) \tag{1}$$

Where *N* is a big integer, k is a constant (in the range of 0.5 – 0.7) and *dist(v)* is an approximation of the Euclidean distance from the object boundary, obtained with a sampling of the object in 26 directions around *v*). The function is designed to heavily penalize voxels too near to the object boundary without affect the centered voxels, in order to produce smooth lines. The path calculated with this algorithm, is subsampled with a factor of 7 and the remaining points are interpolated with a cubic B-Spline. Once that a centerline is calculated for every connected component, they are analyzed in order to reconstruct the full centerline. First it is needed to recognize colonic component from non colonic one that could be present in the segmented image. A set of rules derived from the anatomy knowledge of the colon are employed for this task, see [18] for more details.

6.1   Layers Based Centerline

In this second approach too, we use distance map to calculate centerline of the colon, but introducing an additional step that uses a modified region growing algorithm. The input from which to calculate centerline is the binary mask representing colon structure and the distance map, a matrix containing the distance of each voxel inside the colon lumen to the nearest voxel of the colon wall. Starting from the consideration that colon is essentially a tubular structure, to identify extension direction we apply region growing algorithm. Choosing seed voxels in one extremity of the colon, for simplicity the rectum, region containing seed voxels expands progressively from rectum to cecum. To reconstruct easily colon trend we modify region growing algorithm associating at each layer of growth a label corresponding to the current iteration number of the algorithm. This will virtually divides the colon into layers, each of which has a label with an incremental number from rectum to cecum.

There are a lot of benefits resulting from the division of the colon in layers. The most obvious is the simplicity to identify the start point and the end point of the centerline. The end of the centerline is very critical because if it is incorrectly calculated,
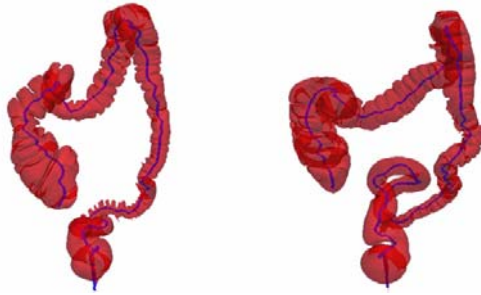
Fig. 8. Representation of how the algorithm works to calculate centerline from the colon stratification and distance map

Fig. 9. Layered based method: in B are visible key points, interpolated in C with a B-Spline

we could consider again regions already analyzed with a subsequent folding of the line on itself. Another advantage is to avoid cycles in the lines, because following the growth layers we can select the following point in the right direction from time to time. Finally, subdivision in layers permits to calculate centerline regardless of the presence of collapses. This problem is solved by jumping from the point where algorithm detect a colon shrinkage to the nearest point belonging to colon lumen. The counting of the layers will continue normally after the jump. Beginning with colon stratification, we select key points that will form the centerline: for each layer we choose a voxel belonging to the layer and farthest from colon edge, using distance map. Moreover, we select following voxel in the neighborhood of the current key point: in this way, we avoid that the line has sudden direction changes. For each key point we assign a label that allow to associate it to the belonging layer in order to reconstruct the centerline.

The final operation consists of building the full centerline from the calculated key points, using interpolation with a B-Spline. The B-Spline algorithm allows to obtain a line without sudden direction changes, useful characteristic for using the centerline in the virtual navigation of the colon 3D model.

6.2 Comparison Evaluation

Literature about colon skeletonization is quite large, but there isn't a universal accepted evaluation method. Common methods involve use of phantom dataset, or comparison with human traced centerline. We propose a new approach based on the mean distance of the line from the colon boundary as metric. Colon has a very irregular shape, but on the same colon, we expected that centerlines extracted with different algorithms produce the same metrics value.

Fig. 10. Two examples of calculated centerline using all the algorithms described

Moreover, this metric can be used also as algorithm performance indicator, because biggest means values, means more centered lines. We have evaluated the proposed algorithms, on a dataset of 7 segmented colon. Mean values of distance from the colon boundary are reported in Table 1.

Table 1. Algorithm evaluation based on the mean distance of the line from the colon boundary

Case #	Mean distance from boundary (voxel)	
	Penalized based	Layered based
1	13.12	12.17
2	12.23	11.46
3	12.67	10.58
4	12.34	12.26
5	8.65	8.30
6	12.03	11.34
7	13.56	13.61

The results shows that the penalized distance based approach performs slightly better, in all but one cases, by the way the distance values of both algorithms are quite consistent, so we can argue that their soundness is verified.

7   3D Reconstruction of the Colon

Reconstruction of a 3D model of the colon is achieved by using the Marching Cubes algorithm, which reconstructs surface of the interested object composing triangular meshes to create a connected surface Visualization of the colon 3D model can be from an external point or an internal point of view. **External view** is very useful to achieve an overall view, for example is possible to determine colon morphology and easily establish if colon presents collapses that can create problem in virtual navigation. Moreover, assigning the appropriate colors and transparency level to each object, it's possible to create a complex and realistic scenes, as the following pictures. The endoscopic view coincides with the **virtual colonoscopy**: a virtual camera moves in the colon following a path described by the centerline. This point of view of the colon 3D model simulates the optical colonoscopy to examine colon walls to detect adenomatous polyps.
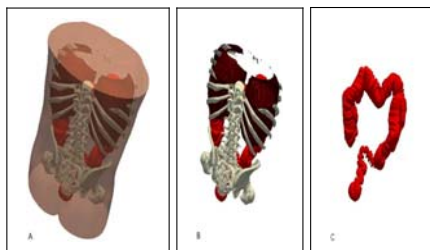
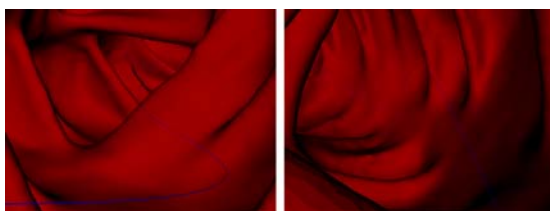**Fig. 11.** Examples of complex scene obtained using an external view

Fig. 12. Examples of Virtual colonoscopy

The navigation begins from the rectum and follows the centerline that represents a virtual track on which the virtual camera moves.

8 Conclusion

This paper presents a complete framework for image processing in Virtual Colonoscopy. A simple but effective fluid removal algorithm is proposed, it relies on analysis of gradient on the boundary between fluid and air and voxel classification. Segmentation of the colon lumen is performed with a sequence of morphological and threshold operations guided by anatomical knowledge. For the centerline extraction two different approaches are proposed: penalized distance based, and layered based. We have introduced a new metric for performance evaluation based on the mean distance from the colon boundary. Finally a virtual navigation interface has been implemented using the marching cube algorithm and surface rendering.

References

1. Greenlee, R.T., Murray, T., Bolden, S., Wingo, P.A.: Cancer statistics. Ca. Cancer J. Clin. 50, 7–33 (2000)
2. Lakare, S., Chen, D., Li, L., Kaufman, A., Liang, Z.: Electronic Colon Cleansing Using Segmentation Rays for Virtual Colonoscopy. Medical Imaging (2002)
3. Skalski, A., Socha, M., Zieliftski, T., Duplaga, M.: Colon Cleansing for Virtual Colonoscopy Using Non-linear Transfer Function and Morphological Operations. In: IEEE International Workshop on Imaging Systems and Techniques - IST (2007)
4. Tu, Z., Zhou, X(S.), Comaniciu, D., Bogoni, L.: A learning based approach for 3D segmentation and colon detagging. In: Leonardis, A., Bischof, H., Pinz, A. (eds.) ECCV 2006. LNCS, vol. 3953, pp. 436–448. Springer, Heidelberg (2006)

5. Wang, Z., Liang, Z., Li, X., Li, L., Li, B., Eremina, D., Lu, H.: An Improved Electronic Colon Cleansing Method for Detection of Colo-nic Polyps by Virtual Colonoscopy. IEEE Transactions on Biomedical Engineering 53(8) (2006)
6. Serlie, I.W.O., Truyen, R., Florie, J., Post, F.H., van Vliet, L.J., Vos, F.M.: Computed Cleansing for Virtual Colonoscopy Using a Three-Material Transition Model. In: Ellis, R.E., Peters, T.M. (eds.) MICCAI 2003. LNCS, vol. 2879, pp. 175–183. Springer, Heidelberg (2003)
7. Perona, P., Malik, J.: Scale-space and edge detection using anisotropic diffusion. IEEE Transactions on Pattern Analysis Machine Intelligence 12, 629–639 (1990)
8. Canny, J.: A Computational Approach To Edge Detection. IEEE Trans. Pattern Analysis and Machine Intelligence 8, 679–714 (1986)
9. Blum, H.: A transformation for extracting new parameter of shape. In: Models for the Perception of Speech and Visual Form. MIT Press, Cambridge (1967)
10. Bitter, I., Kaufman, A.E., Sato, M.: Penalized-Distance Volumetric Skeleton Algorithm. IEEE Transactions on Visualization and Computer Graphics 7(3) (2001)
11. Chen, D., Li, B., Liang, Z., Wan, M., Kaufman, A., Wax, M.: A tree-branch searching, multi-resolution approach to skeletonization for virtual endoscopy. In: SPIE Medical Imaging 2000 (February 2000)
12. Iordanescu, G., Summers, R.M.: Automated centerline for computed tomography colonography. Academic Radiology 10(11), 1291–1301 (2003)
13. Sadleir, R.J.T., Whelan, P.F.: Fast colon centreline calculation using optimised 3D topological thinning. Comput. Med. Imaging Graph. 29(251-258) (2005)
14. Bouix, S., Siddiqi, K., Tannenbaum, A.: Flux driven automatic centerline extraction. Medical Image Analysis 9(3), 209–221 (2005)
15. Sethian, J.A.: Level Set Methods and Fast Marching Methods. Cambridge University Press, Cambridge (1996)
16. Deschamps, T., Cohenb, L.D.: Fast extraction of minimal paths in 3D images and applications to virtual endoscopy. Medical Image Analysis 1 (2001)
17. Dijkstra, E.W.: A note on two problem in connexion with graphs. Numerische Mathematik 1, 269–271 (1959)
18. Frimmel, H., Nappi, J., Yoshida, H.: Centerline-based colon segmentation in CT-Colonography. Medical Physics 32(8) (August 2005)
19. Chen, D., Hassouna, M.S., Farag, A.A., Falk, R.: An improved 2D colonic polyp segmentation framework based on gradient vector flow deformable model. In: Yang, G.-Z., Jiang, T.-Z., Shen, D., Gu, L., Yang, J. (eds.) MIAR 2006. LNCS, vol. 4091, pp. 372–379. Springer, Heidelberg (2006)
20. Näppi, J., Dachman, A.H., MacEneaney, P., Yoshida, H.: Automated Knowledge-Guided Segmentation of Colonic Walls for Computerized Detection of Polyps in CT Colonography (2002)
21. Masutani, Y., Yoshida, H., MacEneaney, P.M., Dachman, A.H.: Automated Segmentation of Colonic Walls for Computerized Detection of Polyps in CT Colonography (2001)
22. Wyatt, C.L., Ge, Y., Vining, D.J.: Segmentation in virtual colonoscopy using a geometric deformable model (2006)
23. Tu, Z., Zhou, X(S.), Comaniciu, D., Bogoni, L.: A learning based approach for 3D segmentation and colon detagging. In: Leonardis, A., Bischof, H., Pinz, A. (eds.) ECCV 2006. LNCS, vol. 3953, pp. 436–448. Springer, Heidelberg (2006)
24. Näppi, J., Yoshida, H.: Centerline-based colon segmentation for CT colonography (2005)
25. Anisotropic mean shift based fuzzy C-means segmentation of dermoscopy images. IEEE Journal of Selected Topics in Signal Processing 3(1), 26–34 (2009)

Combined Use of Densitometry and Morphological Analysis to Detect Flat Polyps

Vitoantonio Bevilacqua[1,2], Marco Cortellino[1,2], Giuseppe Mastronardi[1,2], Antonio Scarpa[1], and Diego Taurino[1]

[1] Politecnico di Bari - Dipartimento di Elettrotecnica ed Elettronica
Via Orabona, Bari, Italy
[2] E.B.I.S. s.r.l. (electronic Business In Security), Spin-Off of Polytechnic of Bari,
Via Pavoncelli, 139 – 70125 Bari – Italy
`bevilacqua@poliba.it`

Abstract. This paper describes a CAD system to detect a particular type of colon cancer, flat polyps. The identification of suspicious regions is based on two type of analysis executed in succession: a densitometry analysis that researches contrast fluid on polyp surface and a morphological analysis that reduces number of false positives, calculating a curvature index of the surface.

Keywords: Virtual Colonoscopy, CAD, Flat Polyps, CT Colonography.

1 Introduction

The idea of developing CAD system for virtual colonoscopy (VC) is relatively recent, it is due to the advent of this new examination technique. In the case of CV, the CAD is related to a computerized scheme that automatically detects a series of suspicious polyps on CT images to report to the radiologists, who distinguish the colon lesions by false positives of system.

In particular, the developed system has the task for identifying automatically a particular type of polyps: flat polyps. This is a very difficult goal to be achieved because this type of polyps does not present morphological characteristics similar to common polyps.

Usually, the polyps are simple mucosal protuberances, contours are smooth and regular, flat polyps instead, as the name suggests, haven't these features but may appear as a slight swelling of the colonic surface [1]. The feature that distinguishes them from ordinary adenomatous polyps is the proportion between height and width: the height should be less than twice its diameter [2].

The scientific community devotes more attention to these kinds of polyps, because of the high difficulty in identifying and the high probability of turning into cancer (actually discussed yet). Because of their morphology, the CAD developed for normal polyps are not able to identify them (often because they are based on the research of Gaussian shape). In addition, flat lesions are difficult to detect by means of optical colonoscopy, so CAD systems can assist radiologists in this task [3].

In order to design an efficient CAD for flat polyps, morphological features are less important for CAD than other type of polyps. However, in order to maintain this minimum information is essential to perform very refined colon segmentation.

The existing CAD is mainly based on the thickness of the colon: the thickness increases near to the lesion. Other techniques, however, are designed to identify the base of the polyp, evaluating the extension [4].

The implemented CAD is based on contrast fluid used to mark fecal residue; it is able to temporarily hold on flat lesions of colon surface [5]. This is caused by the surface of the polyp which tends to retain the liquid around it. Therefore, flat polyps' detection initially starts identifying these small areas on colon surface much clearer than surrounding tissue. After that, obtained results were skimmed using morphological characteristics to distinguish better polyps from isolated regions of liquid: in this way the number of false positives has decreased considerably.

2   Dataset and Library

The dataset of images that we use to test implemented algorithm is the WRAMC (Walter Reed Army Medical Center) available from the National Cancer Imaging Archive (NCIA). This database, downloadable for research purposes, contains 813 image series: each scan consists of about 400 images with resolution of 512x512 in DICOM format. Pixels of an image are represented by a CT Number of 2 byte. The database contains, besides, a lot of information regarding the presence of polyps, their morphology and reports the distance of any polyps from the rectum, a useful information to identify the lesions in the colon. Distance from rectum is also important from the medical point of view, because this parameter is necessary before the start of removal operation.

Language used to implement the application is C++, with the support of the libraries ITK (Insight Segmentation and Registration Toolkit) and VTK (Visualization Toolkit), two open source and cross-platform libraries containing helpful functions for image processing and image visualization in 2D and 3D.

3   ROI Detection

The first step was to examine the slice available to select regions that rise to this type of lesion. Because they are only on inside surface of colon, the Region of Interest is the contour of the colon that we calculate from the binary mask obtained from colon segmentation. By limiting the analysis to these areas, rather than calculating it on the entire volume, it has several advantages: the most important is the decrease of error probability, identifying fewer false positives.

Therefore the algorithm starts with edge detection of the segmented regions obtained by means of colon skeletonization process. To detect voxels belonging to colon surface, for each voxel of the binary mask that represents colon lumen points, a check is carried out on the 3x3x3 neighborhood: if at least one neighborhood of the voxel is outside the lumen of the colon, it is classified as edge voxel.

In this case, however, interested area should be extended because, as suggested also in [6], near a flat polyp, there is an increase of thickness of the colon wall.

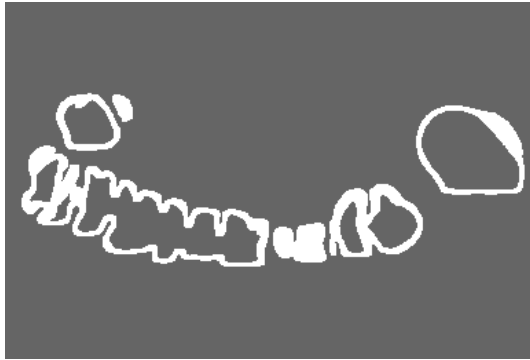**Fig. 1.** Original image (A) and binary mask representing colon lumen

Fig. 2. Selected ROI, calculated from binary mask of the colon lumen

So, to select completely each lesion, we execute an expansion of the edge area using a dilatation operator. This process leads to identify a new binary mask that specifies voxels to be analyzed in subsequent steps of the algorithm and that represents the region of interest.

4   Densitometry Analysis

The followed approach in this CAD is based on the presence of contrast fluid on surface of flat polyps. The fluid, that has the role of highlight residual stools, is easily identifiable because the values of HU are rather high: CT number value corresponding to contrast fluid is very similar bone CT numbers, but in the colon area there is no possibility of confusion [7].

Figure 3 shows a flat polyp covered with a thin layer of contrast fluid: it is possible to identify these areas more clearly by using different methodologies. The simplest is the thresholding, but executed tests carried out that a greater sensitivity is obtained by

using gradient value how comparison method. In fact, because the lesions are located on the inner surface of the colon, in these areas can be observed a clear passage between the fluid and the air. Therefore, attention is focused on the change of intensity values between adjacent voxel, rather than just the intensity values.

The entire mask representing the contour of the colon is examined to find slight variations in density due to the presence of flat polyps. The volume can be interpreted as a function of variables x, y and z, in which the value of a generic pixel is given by $f(x, y, z)$, then gradient is defined as the vector that has the partial derivative of the volume function: it represents the direction with maximum increment of the function.

Therefore, in order to differentiate into two classes (lesion and healthy tissue) parts of analyzed tissue, for each point of the mask is calculated the gradient value to highlight anomalies in the distribution of gray levels. So, to select points of relative maximum gradient we use a threshold filter: the threshold value is 300. After this, we delete isolated voxel because a polyp surface has a wider area. However, other regions present the same characteristics of flat polyps: an example is the interface between fluid deposed on the bottom of the colon and colon lumen that has high gradient values.
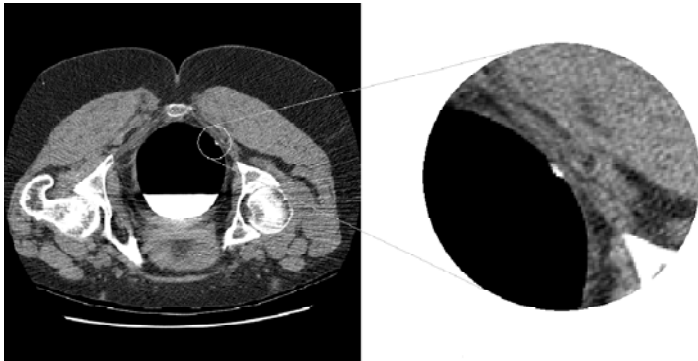

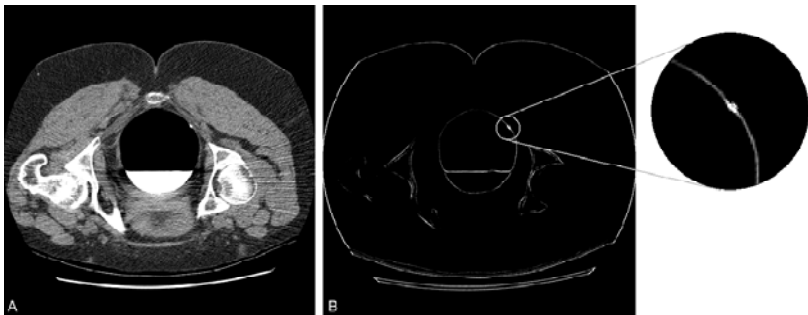**Fig. 3.** A polyp on the surface, covered by a thin layer of contrast fluid

Fig. 4. Gradient domain (B) and, in detail, the area containing the polyp

However, there is an obvious tendency of the deposited fluid to trace a perfectly horizontal line: in this way, we detect this interface. This implies an increase of the gradient in this region, especially along the y direction, and then we just isolate this component of the gradient, discarding areas with vertical gradient value greater than a fixed threshold.

In the next step, we isolate the suspicious areas corresponding to fluid on colon surface, excluding the liquid deposited on the bottom of the lumen of colon, but also small volume of fluid trapped in the colon folds, through recognition of the air-liquid interface.

5 Morphological Analysis

Densitometry analysis allows to have greatly limited the suspicious areas on colon surface, but the number of candidate polyps is still too high. To improve this result we must use other information, such as morphological ones, to reduce the number of false positives, despite flat polyps' present morphological characteristics less evident than normal adenomatous polyps [8].

Performing this analysis on entire colon surface, we would have found a huge number of false positives, because in this case we would interpret a normal fold or a normal surface as a flat polyp, introducing a lot of false positives.

By applying this kind of analysis in a limited number of areas, we can remove polyps characterized by shapes that deviate too much from a normal polyp shape [9].

The developed algorithm is based on the recognition of the surface shape by analyzing mesh obtained from colon segmentation. In particular, we consider the mesh orientation in the extracted surface.
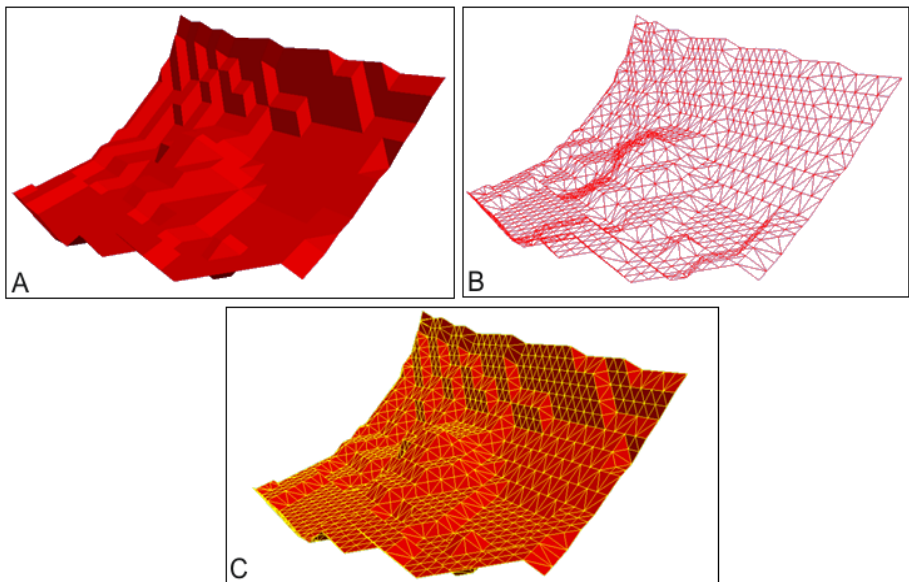

Fig. 5. (A) A flat polyp in the center of the surface. In B and C, are emphasized triangular meshes that compose the surface.

For each candidate polyp, is extracted a 13x13x13 volume centered on it to calculate the local surface through reconstruction with mesh, using the Marching Cubes algorithm. The analysis is based on a comparison between a normal wall surface and a colonial area where there is a polyp.

Considering the meshes that form extracted surface, we calculate direction of the normal vector for each mesh. To obtain a parameter that characterizes the mesh orientation for the entire analyzed area, we calculate variance of these normal vectors.

The normal vector variation is interpreted as a curvature index of the extracted surface: in this way we can exclude from analysis those areas with a high variance, corresponding to folds and surfaces with low variance, instead corresponding to normal colon walls.
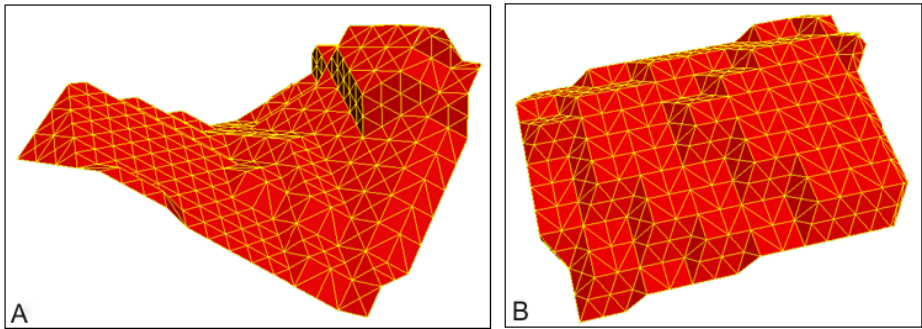

Fig. 6. Cases to be discarded because the variance of the normal is too high (A) or too low (B)

Selecting an intermediate range, we can discard those areas which don't contain polyps, reducing significantly the number of false positives.

The calculation of normal vectors is made by extracting the three vertices for each triangular mesh that forms the surface. Starting from these three vertices, we locate the plane passing through these points, and from this the normal vector. The variance of the normal vector is so calculated:

$$\sigma^2 = \frac{\sum_{i=1}^{n}\left(\begin{bmatrix} N_x \\ N_y \\ N_z \end{bmatrix} - \begin{bmatrix} \overline{N}_x \\ \overline{N}_y \\ \overline{N}_z \end{bmatrix}\right)^2}{n}, \tag{1}$$

where N_x, N_y and N_z represent the normal components of the vector along x, y and z, while $\overline{N}_x$, $\overline{N}_y$ and $\overline{N}_z$ components are the corresponding averages of the entire surface.

The variance is a measure of dispersion because it is null only when all values are equal with each other (and therefore equal to their average) and it grows with the increase of differences between values. In the examined case, variance is null where

surface is completely flat, while areas with higher values present meshes oriented in different directions.

Flat polyps have intermediate values of variance, which allow us to distinguish polyps and normal areas. We performed, therefore, a detailed analysis to identify the best threshold to increase CAD accuracy. Then, we normalized the obtained value of variance in the interval 0-1: in this way we excluded polyps with normalized variance less than 0,2 and more than 0,6.

6   Results

We test implemented CAD on 10 scans in the available database containing information regarding the presence of flat polyps that help us to evaluate the algorithm performance.

The analysis of the results is performed by comparing candidate polyps resulting from CAD that uses only the densitometry analysis and the CAD that uses also geometrical information. We can assert that the number of detected polyps in the second methodology is lower for all cases: furthermore, we verified that the discarded polyps through morphological information really represent false positives.

The following is a summary of performed tests:

Table 1. Number of FP obtained with only densitometry information and with also morphological information

ID	Number of flat polyps	Densitometry analysis (detected polyps)	Morphological and densitometry analysis (detected polyps)	False positive correctly discarded
SD VC-058M	2	17	4	76 %
SD VC-093M	3	112	23	79 %
SD VC-238M	3	43	19	56 %
SD VC-279M	3	13	4	69 %

The table shows the benefits of the combined use of morphological and densitometry information: the percentage of false positives rejected in the second step of the algorithm is about 70%.

Therefore, we can assert that the joint use of both kind of analysis is positive, improving the accuracy of the CAD system.
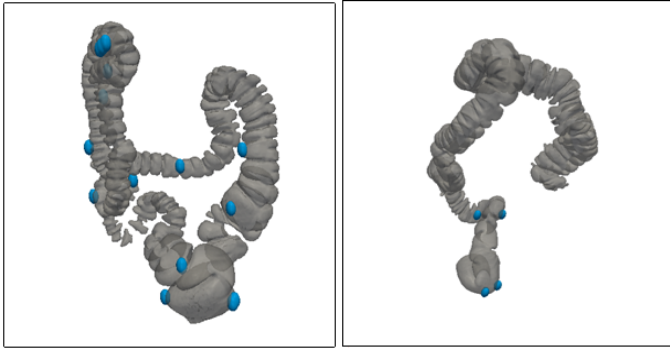
Fig. 7. Results of the developed CAD: blue spheres highlight suspicious areas in the case SD VC-238M and SD VC-279M

7   Conclusion

Analyzing the results we can state that the number of false positives returned by the system is acceptable, because of inherent difficulty of the problem. Although the presence of false positives, the attention of the radiologist is focused on a small number of points in the colon that are the suspected areas.

This result is very important because, by reducing the number of regions to be submitted to radiologists, both the time of analysis and errors probability can be reduced in detecting a flat lesion.

References

1. Fidler, J.L., Johnson, C.D., MacCarty, R.L., Welch, T.J., Hara, A.K., Harmsen, W.S.: Detection of flat lesions in the colon with CT colonography 27(3), 292–300 (2002)
2. The Paris endoscopic classification of superficial neoplastic lesions: esophagus, stomach, colon. Gastrointest Endosc 58, S3–S43 (2003)
3. Zhao, L., et al.: Lines of Curvature for Polyp Detection in Virtual Colonoscopy. IEEE Transactions on visualization and computer graphics 12(5), 885–892 (2006)
4. Taylor, S.A., Suzuki, N., Beddoe, G., Halligan, S.: Flat neoplasia of the colon: CT colonography with CAD. Abdominal Imaging 34(2), 173–181 (2008)
5. O'Connor, S.D., Summers, R., Choi, J.: Oral contrast adherence to polyps on CT colonography. Journal of computer assisted tomography 30(1), 51–58 (2006)
6. Van Uitert, R.L., Summers, R.M.: Colonic wall thickness using level sets for CT virtual colonoscopy visual assessment, polyp detection. In: Manduca, A., Hu, X.P. (eds.) Medical imaging: physiology, function, and structure from medical images, Proceedings of SPIE, vol. 6511, p. 65110S (2007)
7. Zhao, L., et al.: Lines of Curvature for Polyp Detection in Virtual Colonoscopy. IEEE Transactions on Visualization and Computer Graphics 12(5), 885–892 (2006)

8. Taylor, S.A.: CT colonography: Computer-aided Detection of Morphologically Flat T1 Colonic Carcinoma. Acad. Radiol. 14, 287–300 (2008)
9. Chen, D., Hassouna, M., Farag, A., Falk, R.: Principal Curvature Based Coloinc Polyp Detection. International Journal of Computer Assisted Radiology and Surgery 2(supp. 1), 6–8 (2007)
10. Yao, J., et al.: CT colonography computer-aided polyp detection using topographical. In: IEEE International Conference, Image Processing, ICIP 2007, vol. 5, pp. V21–V24 (2007)

Relevant Measurements for Polyps in 3D Virtual Colonoscopy

Vitoantonio Bevilacqua[1,2], Marianna Notarnicola[1], Marco Cortellino[1,2],
Antonio Scarpa[1], Diego Taurino[1], and Giuseppe Mastronardi[1,2]

[1] Politecnico di Bari - Dipartimento di
Elettrotecnica ed Elettronica
Via Orabona, 4 – 70125 Bari, Italy
[2] E.B.I.S. s.r.l. , Spin-Off of Polytechnic of Bari,
Via Pavoncelli, 139 – 70125 Bari – Italy
bevilacqua@poliba.it

Abstract. Virtual Colonoscopy is an innovative method to discover colon neo-
plasias created in order to alleviate patients aches generated by the standard
colonoscopy. For the same reason, we have realized an automatic process final-
ized to find polyps into the lumen through the extraction of colon centerline and
the calculation of polyps distance from anus. This paper contains the descrip-
tion of what is implemented. In particular, the developed algorithms build up
following steps: colon lumen segmentation starting from a dataset of axial CT
slices, 3D rendering, centerline extraction and evaluation of polyps distance
from anus.

Keywords: Virtual Colonoscopy, Segmentation, Skeletonization, Rendering,
Image processing, CT Colonography, Polyp distance.

1 Introduction

Colorectal cancer (CRC) is the third most common human malignancy and the second
leading cause of cancer-related deaths in the United States [1]. CRC is caused by a
malign proliferation of the cells that constitute the inner lining of the colon, and repre-
sents an high social-impact pathology in industrialized countries. Most colorectal can-
cers are believed to arise within benign adenomatous polyps that gradually develop
over the course of many years [2]. Evidence-based guidelines recommend the screen-
ing of adults who are at the age of average risk for colorectal cancer, that is 50 years
old, since the detection and removal of adenomas has been shown to substantially re-
duce the incidence of cancer and cancer-related mortality. Several methods for colo-
rectal screening are available such as colonoscopy, sigmoidoscopy, barium enema and
fecal occult blood testing. While currently available optical colonoscopy is the most
accurate diagnostic procedure for detecting polyps, it is expensive, invasive, uncom-
fortable, time consuming and carries a slight risk of perforation and death. Patients are
usually reluctant for early examination [15]. A minimally invasive, safe, and low-cost
method to evaluate the colon for colorectal cancer would be preferred by most

patients. Recently, virtual colonoscopy also refereed as CT-Colonography has been developed as an alternative method to evaluate the entire colon for polyps [3, 4, 5]. This new technology requires advanced image processing techniques to allow radiologists to virtually navigate through a patient's colon model reconstructed from helical abdominal computed tomography (CT) images.

2   Dataset and Library

The dataset of images that we use to test implemented algorithm is the WRAMC (Walter Reed Army Medical Center) available from the National Cancer Imaging Archive (NCIA). This database, downloadable for research purposes, contains 813 image series: each scan consists of about 400 images with resolution of 512x512 in DICOM format. Pixels of an image is represented by a CT Number of 2 byte. The database contains, besides, a lot of information regarding the presence of polyps, their morphology and reports the distance of any polyps from the rectum in cm, a useful information to identify the lesions in the colon. Distance from rectum is also important from the medical point of view, because this parameter is necessary before

the start of removal operation. Language used to implement the application are MatLab, especially its image processing tool and C++ with the support of the libraries ITK (Insight Segmentation and Registration Toolkit) and VTK (Visualization Toolkit), two open source and cross-platform libraries containing helpful functions for image processing and image visualization in 2D and 3D.

3   Colon Segmentation

In literature exists several techniques to segment the colon. Most of these methodologies is based simply on the combined use of image analysis tools, for example applying in sequence image filters. For example in [23] the image gradient is used to identify colonic walls, since there is a sudden transition from high CT Number corresponding to a tissue and lower CT Number corresponding to colon lumen. Other algorithm to segment colon, how in [24] e [25], anatomical information is used to improve results, for example identifying regions in the volume where certainty there aren't colon regions. More experimental approaches use geometric deformable models [26] to identify colonic wall or algorithms based on learning techniques [27]. Approach that we adopt in this work is similar to [28] where colon segmentation is obtained estimating centerline: in particular, we use centerline to delete extra-colic areas. Implemented algorithm consists of the following four steps, background removal, lung removal, selection of air regions in patient body, extra-colic areas removal. These steps will be analyzed in the following paragraphs.

3.1   Background Removal

In a lot of algorithms developed for this aim, the first step to analyze TC images is to remove the background. In fact, areas around the patient don't contain relevant information and they represent only a payload to be removed. The purpose of this step is

to build binary masks that represent the Region Of Interest (ROI) in the images: this operation is necessary to reduce or to avoid possible misinterpretation of the algorithm, by restricting the analyzed areas. This is possible because the following steps will not applied to the entire volume, but only on the region really occupied by the patient: this areas is often very smaller than the whole volume. One of the most utilized techniques used to remove large connected areas with homogeneous characteristics is the region growing algorithm: in this case we use the Hounsfield Unit (HU) assigned to a voxel to detect connected areas. In region growing algorithm, we have to indicate a collection of seed pixels from which start, adding recursively their adjacent pixels if they satisfy the adopted similarity condition. We apply region growing in the three dimensional space, by placing the seed voxels in the eight vertices of the parallelepiped formed by the volume of images. Chosen similarity condition select adjacent voxels only if their CT Number is between -2000 and -300 HU. This range of HU corresponds to voxel that can be considered air voxel: in this way, the region growing ends when reaches the limit of the patient body.
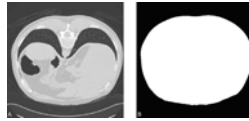

Fig. 1. Binary mask (B) representing the ROI of (A), after background removal

3.2 Lung Removal

Lung is a organ with density characteristics very similar to colon, since it is less dense than other tissues. For this reason, before proceeding to identify the colon in the patient, we restrict further the ROI deleting voxels belonging to lungs. Adopted technique uses the region growing algorithm in the three dimensional space: seed voxels are positioned in left lung and in right lung. In this way, since colon and lung are visually and physically disconnected, the region growing algorithm selects only the lungs in the body. We select seed voxels in the first slice (if the last slice correspond to rectum): in fact, in the first slice considering only the ROI region, voxels that have CT Number with a low value are voxels belonging to lungs. Similarly to the precedent application of the region growing algorithm, the similarity condition is the belonging to a range of CT Number, in particular in the range between -2000 and -700. Using region growing algorithm, we infuse again robustness to the process, allowing a very fast detection of the lungs.
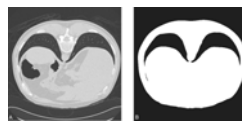

Fig. 2. Removal of the lungs from ROI. The mask is again represented by the white pixels in the image B.

3.3   Extra-Colon Volumes Removal

At this point, considering selected ROI, voxels related to air regions match to colon lumen, but are included other regions too, how for example intestine and small extra-colon areas corresponding to noise in images. To select all these areas, we use simply a thresholding filter: the threshold that we choose is optimized by analyzing carefully available cases to select the entire colon and less extra-colon areas is possible. The threshold, chosen analyzing histogram too, is -800 to distinguish areas that refer to tissues from organ containing air.
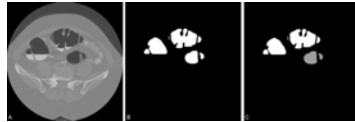

Fig. 3. Applying thresolding filter to the ROI, selected areas are colon lumen, intestine and small noise areas. Disconnected regions are tagged with different values (C).

In figure B is possible to observe how areas outside the colon are noise regions caused by the presence of artifacts in the image. In fact, not all the scans show this characteristic: artifacts may be caused by a bad preparation of the contrast fluid. To avoid mistakes in following steps of the implemented algorithm, the noise removal is performed, however, on all volumes. Regions belonging to binary mask and corre-sponding to noise have a limited volume, if compared with colon volume.

Then, for every disconnected region in the three dimensional space, we calculate the volume in number of voxels: we set a threshold to delete region too small to be considered part of colon lumen. This removal, however, don't allow to exclude from analysis too large volumes that don't correspond to colon lumen, such as the intestine. Intestine removal is performed considering the anatomical information that is the av-erage length of a colon. Calculation of the length of a colon can be performed using centerline. Detection of the centerline in a colon is described in the following para-graphs. We observed that a colon can be collapsed in several point, then we have to distinguish if a colon shrinkage is a collapses or if is simply the end of the colon. The anatomy of the colon is characterized by a gradually increasing of diameter from rec-tum to cecum, then collapse probability is higher in the first section, near to rectum. Then, we set a threshold, considering this observation: we estimate a threshold related to distance of examined point from the rectum. If we are under this threshold, a shrinkage of the colon lumen is interpreted how a collapse, else we consider the colon ended. In this way, we can easily separate colon lumen from intestine and remove any extra-colon areas still present in the binary mask.
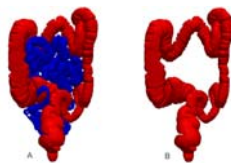

Fig. 4. Two examples in which is visible the improvement of results after segmentation refine-ment. In A the intestine is visible in blue, then excluded from the final result in B.

4   3D Reconstruction of the Colon

The 3D Rendering is a process that leads to the visualization of 3D object. Colon 3D realization is obtained using the Matlab function *isosurface* which implements Lorensen "Marching Cubes" algorithm.

The starting data are:

- a three-dimensional space S;
- a function *Val : R^3 $\rightarrow$ Z* that associates each voxel $v_{x,y,z}$, belonging to S, to a value *Val($v_{x,y,z}$)=i*;
- a value *a* belonging to Z.

The *isosurface* is defined as the whole of voxels that have same value. The algorithm runs cubes (voxels) which S is constituted of and for each of them, it verifies if voxel is crossed by isosurface or if it's out from isosurface, building one or more polygons that represent the correspondent isosurface part. At the end, melting all polygons, searched surface is realized. During segmentation, the volumetric dataset V is transformed in a 3D scalar camp, where a voxel has value 1 if belongs to colon else it has value 0. The scalar camp is indicated whit C. Colon isosurface contains voxels such that *Val(v) = 1*. Marching Cubes realizes isosurface approximation with a triangular mesh. To perceive the three-dimensionality of created structure *shading* techniques are used. These consist in to insert lighting sources inside the scene, establishing with their own colours and their own positions, the colour of each surface's point reproducing real vision mechanism; moreover to round off mesh corners, that are too spiked, *smoothing* operation is applied. Following figure shows the two possible cases that we can find after 3D reconstruction: a collapsed colon, that is to say constituted by two disconnected parts, and an integral one.

Fig. 5. 3D model of a collapsed (left) and an integral (right) colon

5   Centerline Extraction

The extraction of a suitable centerline arise in the class of skeletonization problems, as defined in [13], and a lot of methods are been proposed in literature. A good

number of approaches uses the notion of *Distance Map Transform* (DMT) as [14][15], in which every voxel is the radius of a maximal sphere inscribed in the object, so that an approximation of the line could be extracted as the set of local minima in the DM image. A different class of algorithms as [16][17], uses an *onion peeling* approach, also referred as *topological thinning*, that consists in removing iteratively layers of voxel maintaining topological properties, until a single continue line remains. Such algorithms, are typically very slow and depth optimizations are required. Bouix et al. [18] have proposed a new approach based on the mathematical concept of flux applied to a vector field of Euclidean distance from the object boundary. The well known level set algorithms [19], can also be applied to skeletonization problems, as shown in [20] in which an energy functional that describe the line is minimized through a fast marching approach. Skeletonization algorithms suffers of standardized and universally accepted evaluation methods lack. Common approach are the use of phantom volumes, or comparison with human traced lines. We propose the following algorithm.

5.1 Penalized Distance Algorithm

The algorithm consist of several step, including a final segmentation refining that uses the centerline and the anatomy knowledge of the colon to detect non colonic component and remove them. The input of the algorithm is a binary image representing the segmented colon, it is subsampled with a factor of 2, for improving the speed performance, during this operation topological structure of the image is preserved. Ideally the colon should be a single connected component, but due to poor bowel preparation, it can be collapsed in one or more points, resulting in multi component object, that are separately treated. For each component, the initial and the final point of the centerline are calculated, with a double breadth first visit started in an arbitrary point inside the component. The farthest point reached with the first visit is considered the initial point, then starting from it, a new visit is conducted, so the last voxel reached is the final point. Knowing the initial and the final points, the centerline extraction problem is addressed with a modified version of the well known Disjktra algorithm [21], in which a penalty term is included in order to maintain the line far away from colon boundary. In particular for every new voxel v a penalty term is defined as:

$$Penalty(v) = N \exp(k*dist(v)) \tag{1}$$

Where *N* is a big integer, k is a constant (in the range of $0.5 - 0.7$) and *dist(v)* is an approximation of the Euclidean distance from the object boundary, obtained with a sampling of the object in 26 directions around *v*). The function is designed to heavily penalize voxels too near to the object boundary without affect the centered voxels, in order to produce smooth lines. The path calculated with this algorithm, is subsampled with a factor of 7 and the remaining points are interpolated with a cubic BSpline. Once that a centerline is calculated for every connected component, they are analyzed in order to reconstruct the full centerline. First it is needed to recognize colonic component from non colonic one that could be present in the segmented image. A set of rules derived from the anatomy knowledge of the colon are employed for this task, see [22] for more detail. The colonic component lines are then connected with straight lines from the cecum to the rectum.

Fig. 6. In B are visible calculated key points, interpolated in C using a B-Spline

6   Polyp-Anus Distance Valuation

The evaluation of this parameter is very important for the surgeon who must remove polyps from colon lumen. Our system succeeds to calculate polyp-anus distance even if colon is collapsed. The case that follows, considers just this eventuality.

Using the DICOM dataset which corresponds to the following colon and centre-line, thanks to the radiologist we have individuated polyps position (in yellow) in four slices :

Fig. 7. Colon and its centerline

Slice 243   Slice 318

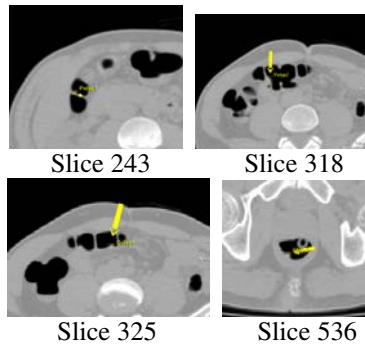Slice 325   Slice 536

Fig. 8. Polyps in axial CT slices

To evaluate polyps positions we have exploited the key points previously calculated. Practically it's summed, piece by piece, the distance between the point in the considered slice and the successive until to reach anus.

The table shows results obtained:

Table 1. Polyp-anus distance calculation

Slice	Pixels Distance	mm Distance
243	3.954,1	1.046,2
318	3.039,4	804,2
325	3.685,3	975,2
536	180,6	47,8

From table it's pointed out the fact that slices utilized are sequential (at first corresponds lungs final part, at last anus) so paying attention on the second row, calculated distance is smaller than the value on the third row. This is due at colon shape. In fact it isn't linear but presents handles that sometimes are accentuated; At the case analyzed corresponds following situation:
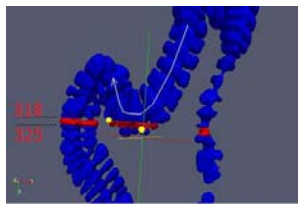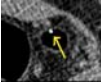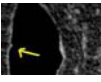

Fig. 9. 3D reconstruction in which slices position is indicated

Image shows how the polyp in the 318° slice is nearest to anus than the one in the 325°.So results found are verisimilar. The table below shows the other results obtained (for shortness only the slice and the distances are pointed out) :

Table 2. Results

Colon	Slice	Pixels Distance	mm Distance
2	250	4.233,8	1.120,2
2	317	3.404,6	900,8
2	443	2.117,6	560,3
3	464	1.588,9	420,4
3	573	414,6	109,7

7   Conclusion

This paper presents a method to individuate polyps into human colon. The first addressed problem is the segmentation of the colon lumen that is performed with a sequence of morphological and threshold operation guided by prior anatomical knowledge. For the centerline extraction the algorithm based on penalized distance is used. We have introduced a new metric for performance evaluation based on the mean distance from the colon boundary. Finally the calculation of polyp-anus distance has been implemented.

References

1. Greenlee, R.T., Murray, T.S., Bolden, P.A.W.: Cancer Statistics. Ca. Cancer J. Clin. 50, 7–33 (2000)
2. Pickhardt, P.J., Choi, J.R., Wang, I.H., Butler, J.A., Puckett, M.L., Hildebrandt, H.A., Wong, R.K., et al.: Computed Tomographic Virtual Colonoscopy to Screen for Colorectal Neoplasia in Asymptomatic Adults. The New England Journal of Medicine 349(23), 2191–2200 (2003)
3. Hong, L., Kaufman, A., Wei, Y.C., et al.: 3-D virtual Colonoscopy. In: Loew, M., Gershon, N. (eds.) Proc. Biomedical Visualization, Altanta, GA, pp. 26–33 (1995)
4. McFarland, E., Brink, J., Loh, J.: Visualization of Colorectal Polyps with Spiral CT Colography: Evaluation of Processing Parameters with Perspective Volume Rendering. Radiology 205, 701–707 (1997)
5. Vining, D.J., Gelfand, D., Bechtold, R., et al.: Technical Feasibility of Colon Imaging with Helical CT and Virtual Reality. Ann. Meeting Amer. Roentgen Ray. Soc., 104 (1994)
6. Lakare, S., Chen, D., Li, L., et al.: Electronic Colon Cleansing Using Segmentation Rays for Virtual Colonoscopy. Medical Imaging (2002)
7. Socha, S.M., Zieliftski, T., Duplaga, M.: Colon Cleansing for Virtual Colonoscopy Using Non-linear Transfer Function and Morphological Operations. In: IEEE International Workshop on Imaging Systems and Techniques - IST (2007)
8. Tu, Z., Ss, X., Zhou, D., et al.: A Learning Based Approach for 3D Segmentation and Colon Detagging
9. Wang, Z., Liang, Z., Li, X., et al.: An Improved Electronic Colon Cleansing Method for Detection of Colonic Polyps by Virtual Colonoscopy. IEEE Transactions on Biomedical Engineering 53(8) (2006)
10. Serlie, I., Truyen, R., Florie, J., et al.: Computed Cleansing for Virtual Colonoscopy Using a Three-Material Transition Model
11. Perona, P., Malik, J.: Scale-space and Edge Detection using Anisotropic Diffusion. IEEE Transactions on Pattern Analysis Machine Intelligence 12, 629–639 (1990)
12. Canny, J.: A Computational Approach To Edge Detection. IEEE Trans. Pattern Analysis and Machine Intelligence 8, 679–714 (1986)
13. Blum, H.: A Transformation for Extracting New Parameter of Shape. In: Models for the Perception of Speech and Visual Form. MIT Press, Cambridge (1967)
14. Bitter, I., Kaufman, A.E., Sato, M.: Penalized-Distance Volumetric Skeleton Algorithm. IEEE Transactions on Visualization and Computer Graphics 7(3) (2001)
15. Chen, D., Li, B., Liang, Z.: A Tree-branch Searching, Multi-resolution Approach to Skeletonization for Virtual Endoscopy. In: SPIE Medical Imaging 2000 (Feburary 2000)

16. Iordanescu, G., Summers, R.M.: Automated Centerline for Computed Tomography Colonography. Academic Radiology 10(11), 1291–1301 (2003)
17. Sadleir, R.J.T., Whelan, P.F.: Fast Colon Centreline Calculation using Optimised 3D Topological Thinning Computerized Medical Imaging and Graphics 29(251–258) (2005)
18. Bouix, S., Siddiqi, K., Tannenbaum, A.: Flux Driven Automatic Centerline Extraction. Medical Image Analysis 9(3), 209–221 (2005)
19. Sethian, J.A.: Level Set Methods and Fast Marching Methods. Cambridge University Press, Cambridge (1996)
20. Deschamps, T., Cohenb, L.D.: Fast Extraction of Minimal Paths in 3D Images and Applications to Virtual Endoscopy. Medical Image Analysis 1 (2001)
21. Dijkstra, E.W.: A Note on Two Problem in Connexion with Graphs. Numerische Mathematik 1, 269–271 (1959)
22. Frimmel, H., Nappi, J., Yoshida, H.: Centerline-based Colon Segmentation in CTColonography. Medical Physics 32(8) (August 2005)
23. Chen, D., Hassouna, M.S., Farag, A.A., Falk, R.: An Improved 2d Colonic Polyp Segmentation Framework Based on Gradient Vector Flow Deformable Model (2006)
24. Janne, N., Dachman, A.H.: MacEneaney, Peter; Yoshida, Hiroyuki, Automated Knowledge-Guided Segmentation of Colonic Walls for Computerized Detection of Polyps in CT Colonography (2002)
25. Masutani, Y., Yoshida, H., MacEneaney, P.M., Dachman, A.H.: Automated Segmentation of Colonic Walls for Computerized Detection of Polyps in CT Colonography (2001)
26. Wyatt, C.L., Ge, Y., Vining, D.J.: Segmentation in Virtual Colonoscopy using a Geometric Deformable Model (2006)

Characterization of Endomicroscopic Images of the Distal Lung for Computer-Aided Diagnosis

Aurélien Saint-Réquier[1], Benoît Lelandais[1], Caroline Petitjean[1],
Chesner Désir[1], Laurent Heutte[1], Mathieu Salaün[2], and Luc Thiberville[2]

[1] Université de Rouen, LITIS EA 4108
BP 12, 76801 Saint-Etienne-du-Rouvray, France
[2] CHU de Rouen, LITIS EA 4108, 76031 Rouen, France
{caroline.petitjean,chesner.desir,laurent.heutte,
mathieu.salaun,luc.thiberville}@univ-rouen.fr

Abstract. This paper presents a new approach for the classification of pathological vs. healthy endomicroscopic images of the alveoli. These images, never seen before, require an adequate description. We investigate two types of feature vector for discrimination: a high-level feature vector based on visual analysis of the images, and a pixel-based, generic feature vector, based on Local Binary Patterns (LBP). Both feature sets are evaluated on state-of-the-art classifiers and an intensive study of the LBP parameters is conducted. Indeed best results are obtained with the LBP-based approach, with correct classification rates reaching up to 91.73% and 97.85% for non-smoking and smoking groups, respectively. Even though tests on extended databases are needed, first results are very encouraging for this difficult task of classifying endomicroscopic images of the distal lung.

1 Introduction

The lungs are the essential respiration organ. They are divided into two anatomic and functional regions: (i) the air conduction system, that includes the trachea, bronchi, and bronchioles, and (ii) the gas-exchange region, or lung parenchyma, made of alveolar sacs. These sacs are made up of clusters of alveoli, tightly wrapped in blood vessels, that allow for gas exchange. Whereas the conduction airways can be explored in vivo during bronchoscopy, the alveolar region was until recently unreachable for in vivo morphological investigation. Therefore, the pathology of the distal lung is currently assessed only in vitro, using invasive techniques such as open lung biopsies. No real time imaging was available.

A new endoscopic technique, called Fibered Confocal Fluorescence Microscopy (FCFM), has recently been developed that enables the visualisation of the more distal regions of the lungs in-vivo [7]. The technique is based on the principle of fluorescence confocal microscopy, where the microscope objective is replaced by a fiberoptic miniprobe, made of thousands of fiber cores. The miniprobe can be introduced into the 2mm working channel of a flexible bronchoscope to produce in-vivo endomicroscopic imaging of the human respiratory tract in real-time.
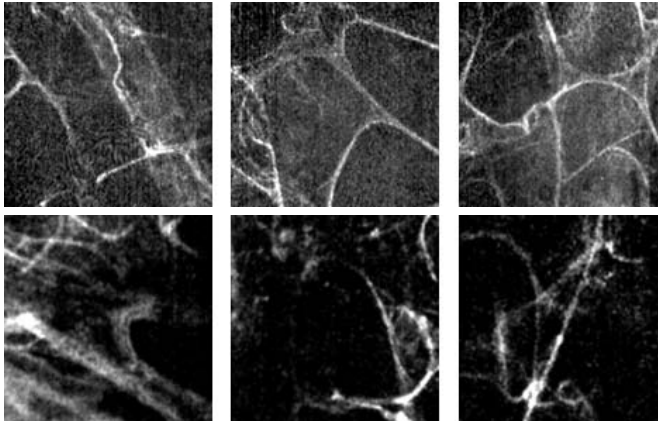
Fig. 1. FCFM images of healthy cases (above) and pathological cases (below)

Real-time alveolar images are continuously recorded during the procedure and stored for further analysis. This very promising technique could replace lung biopsy in the future and might prove to be helpful in a large variety of diseases, including interstitial lung diseases [8].

A clinical trial is currently being conducted that collects FCFM images in several pathological conditions of the distal lungs. This trial also includes a control group of smoker and non smoker healthy volunteers. This strategy provides a dataset of normal images, that can be compared to pathologic ones.

The images recorded within the alveolar regions of the lungs have not been very well described so far. These images represent the alveolar structure, made of elastin fiber, with an approximate resolution of 1μm per pixel. This structure appears as a network of (almost) continuous lines. This elastic fiber framework can be altered by distal lung pathologies and as one can see on Figure 1, images acquired on pathological subjects differ from the ones acquired on healthy subjects. The great complexity of these new images justifies the development of reproductible software tools for computer aided diagnosis, that enables automatic image description for diagnosis and follow up of pathological situations.

We wish to conceive and develop methods for the automatic analysis of FCFM images, so as to classify them as healthy cases or pathological cases. As usual when designing such a pattern recognition system, the most crucial step is the choice of a suitable feature space in which healthy and pathological subjects can be discriminated. However, as it is shown in Figure 1, some images present strong similarities where pathological cases can be visually misclassified for healthy ones and vice versa. The choice of suitable features is therefore critical. Furthermore, to the best of our knowledge, there is no reference work upon which we can rely to find the best way of characterizing these totally new images. Therefore, as it is commonly done when designing the feature extraction stage, two opposite strategies can be investigated to build the feature vector that will describe an image. One can first rely on the visual analysis of the images, searching

for general properties that enable to discriminate classes, transforming them numerically into a feature vector. In this case, it boils down to characterizing images by knowledge-based features that one can a priori assume that they will be adapted to the problem. On the other hand, one can design a set of totally unspecific, low-level, pixel-based features that have been proved to be efficient for characterizing a large variety of images. Because we are currently looking for the best way to describe our FCFM images, we propose in this paper to investigate these two opposite strategies for image characterization: problem-dependent, knowledge-based features vs general-purpose, low level features.

Note that a knowledge-based feature vector has already been designed and evaluated on a preliminary database of FCFM images [5]. Our alternative to design a general-purpose feature vector for characterizing our FCFM images has been to investigate the LBP operator, which is a generic tool to describe an image texture [3]. The aim of this work is thus twofold: (i) to study the influence on classification performance of the parameters of an LBP-based characterization, and (ii) to compare performance obtained with generic features and knowledge-based features.

The remaining of this paper is organized as follows: our classification method, including the description of the two feature sets, is described in Section 2, and results and discussion are provided in Section 3. Section 4 concludes and draws some perspectives for this work.

2 Image Classification Method

2.1 Feature Extraction

LBP-based feature set. The LBP operator is a powerful means of texture description [3]. It consists in locally identifying and coding some binary patterns and in using their distribution to characterize the image. The original invariant against gray-scale LBP analyzes an eight-pixel neighborhood, using the value of the center pixel as a threshold [3]. The LBP code for a neighborhood is obtained by combining the thresholded neighborhood with powers of two and summing up the result (Figure 2). Histogram of the LBP codes can then be used as a texture descriptor.

The method is extended in [4] by defining a circularly symmetric neighbor set (Figure 2). The number of neighbors and the radius are denoted P and R, respectively. The couple (P, R) is denoted scale factor.

The $LBP_{P,R}$ operator can be defined as:

$$LBP_{P,R} = \sum_{p=0}^{P-1} s(g_p - g_c).2^p$$

where g_c is the gray value of the center pixel, g_p defines the gray value of neighbor p and $s(x)$ is the thresholding function defined as:

$$s(x) = \begin{cases} 1 \text{ if } x \geq 0 \\ 0 \text{ otherwise.} \end{cases}$$
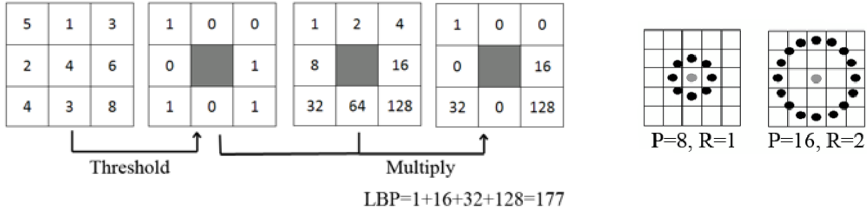
Fig. 2. Circularly symmetric neighbor set for different values of (P, R) (left), the original LBP operator (right)

In order for the LBP code to be rotation invariant, each code is rotated back to a reference position having the smallest LBP code. The rotation-invariant LBP operator, denoted as $LBP_{P,R}^{ri}$, is then defined as:

$$LBP_{P,R}^{ri} = \min\{ ROR(LBP_{P,R}, i) \mid i = 0, 1, ..., P - 1 \}$$

where $ROR(x, i)$ circularly shifts the $P - bit$ binary number x i times to the right and $\min(\cdot)$ is the minimum function.

Let us now denote by U the number of spatial transitions in a pattern. For example, patterns "00000000" and "11111111" have no spatial transition ($U = 0$), while pattern "00001111" has two transitions in the pattern ($U = 2$). Patterns with U values less than or equal to 2 are considered as "uniform". It has been observed that a vast majority of patterns falls into this category [4]. This allows to define a new operator denoted as $LBP_{P,R}^{riu2}$ that outputs only $P + 2$ distinct values:

$$LBP_{P,R}^{riu2} = \begin{cases} \sum_{p=0}^{P-1} s(g_p - g_c) & \text{if } U(LBP_{P,R}) \le 2 \\ P + 1 & \text{otherwise} \end{cases}$$

with:

$$U(LBP_{P,R}) = |s(g_{P-1} - g_c) - s(g_0 - g_c)|$$
$$+ \sum_{p=1}^{P-1} |s(g_p - g_c) - s(g_{p-1} - g_c)|$$

This $LBP_{P,R}^{riu2}$ operator is a gray-scale and rotation-invariant measure of the spatial structure of the image texture. For sake of simplicity, the $LBP_{P,R}^{riu2}$ operator is denoted by $LBP_{P,R}$ in the remaining of the paper.

Texture can also be characterized by its contrast, that can be measured with local variance :

$$VAR_{P,R} = \frac{1}{P} \sum_{p=0}^{P-1} (g_p - \mu), \text{ where } \mu = \frac{1}{P} \sum_{p=0}^{P-1} g_p$$

$VAR_{P,R}$ is by definition invariant against rotation. Note that this variance measure has a continuous-valued output, which requires quantization. Quantization intervals are defined by assigning, between each cut value, the same amount

Table 1. LBP-based feature vectors used to characterize FCFM images

Resolution levels	P, R	Feature number		
		$LBP_{P,R}$	$VAR_{P,R}$	$LV_{P,R}$
	8,1	10	15	25
1	16,2	18	15	33
	24,3	26	15	41
	8,1+16,2	28	30	58
2	8,1+24,3	36	30	66
	16,2+24,3	44	30	74
3	8,1+16,2+24,3	54	45	99

of values from the total distribution of variance. As suggested in [4], the number of intervals is empirically determined. Number of intervals between 10 and 22 have been tested on our training set of images and a number of 15 intervals for quantization of the variance values has finally been chosen, as it provided the best recognition rates on our training set of images.

As shown above, the characterization of images through LBP and VAR operators finally requires to choose only the value of the scale factor, i.e. the number P of neighbors and the radius R. However, to the best of our knowledge, there is no commonly agreed rule upon which one can rely to fix its value according to the characteristics of the images we want to classify. Therefore, in our experiments, we investigated several ways to characterize our images in order to determine how these two parameters act on classification performance. For example, one can first test each operator LBP and VAR independently, with varying values for the scale factor (P, R): (8,1), (16,2), (24,3). Note that a multiresolution analysis can also be performed by concatenating values obtained with operators having varying values of the scale factor. For example, a 2-level LBP operator can be obtained by concatenating $LBP_{8,1}$ values and $LBP_{16,2}$ values. Doing so, three levels of multiresolution analysis can be investigated. Moreover, as $LBP_{P,R}$ and $VAR_{P,R}$ operators allow to respectively characterize the spatial structure and the contrast, one can note that they are of complementary nature. Their combination is thus expected to be a more complete measure of local texture. A new operator denoted by $LV_{P,R}$ can then be created and tested by concatenating the two single operators. Finally, Table 1 lists all the feature vectors that have been tested for comparison and gives their respective feature numbers.

Knowledge-based feature set. Several general characteristics can be observed from the visual analysis of the images. As shown in Figure 1, the alveolar structure in healthy subjects can be described as contrasted continuous lines and curves. On the opposite, in the pathological subset, the disorganization of the meshing is illustrated by the numerous irregularities and the tangle of the fibered structures (see Figure 1). Differences are mostly visible for the structure shape, image texture and contrast. A total of 148 features are computed for this knowledge-based feature vector. Number of features are given in parenthesis.

The **structure contrast** seems to be an important property of the images we deal with. For example, we could suppose that pathological images will have higher values of densities than healthy ones because of an emphasized disorganization of the meshing in pathological images. Therefore the structure contrast can be characterized by studying first order pixel gray level distribution and computing pixel densities. The features retained for characterizing the image contrast are: (i) first order statistics on the image histogram (5), (ii) pixel densities obtained on binarized images using Otsu thresholding (1), (iii) the sum of the image gradient values, obtained using Prewitt operator (1).

The **complexity of the structure shape** can be characterized by studying the image skeleton. After skeletonization [1] obtained on the binary image, the number of junction points is computed. One can suppose that on clearly organized, healthy images, this number will be small, contrary to pathological images where the meshing mess will induce a higher number of points (1).

The **image texture** can be characterized by Haralick parameters computed from co-occurrence matrix [2]. Co-occurrence matrix provides the joint distribution of gray-level intensities between two image points. These two points are located according to several configurations, that represent different distances and rotation angles. We chose the following classical 10 translation vectors: [0 1], [-1 1], [-1 0], [-1 -1], [0 2], [-1 2], [-1 -2], [-2 1], [-2 0], [-2 -1]. From the features originally proposed by Haralick, we retain the following ones: energy, contrast, correlation, variance, inverse different moment, entropy, sum average, sum entropy, sum variance, difference entropy, difference variance, and two information measures of correlation. The only discarded feature is the maximum correlation coefficient, which is too computationally expensive. To these 13 parameters we added dissimilarity, a measure of homogeneity [6]. All these 14 parameters are computed over the 10 co-occurrence matrices (140).

2.2   Classifier

On the previously cited feature sets two standard classifiers have been implemented. First a 1-Nearest Neighbour (1-NN) classifier is used to assess the discriminating power of the features. Due to the high computational cost of the 1-NN classifier, we have also implemented a Support Vector Machine (SVM) classifier on our features [9]. SVM is one of the most performing and most used classification algorithm. The support vector machine classifier is a binary classifier algorithm that looks for an optimal hyperplane as a decision function in a high-dimensional space. A classical choice for the kernel is the cubic polynomial kernel.

2.3   Experimental Protocol

Because of the large difference between non-smoker and smoker images, experiments have been conducted separately on those two groups. In non-smoker, FCFM images exclusively represent the elastin framework of the alveolar ducts. In smokers, tobacco-tar induced fluorescence allows to observe the alveolar walls
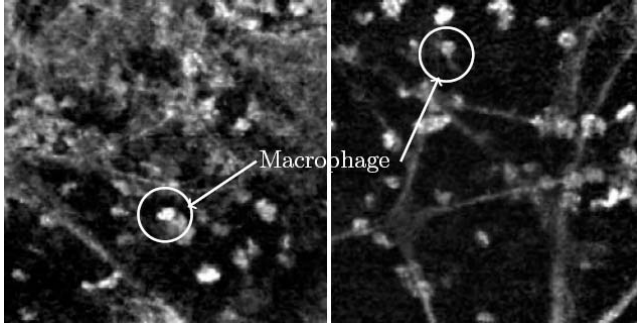
Fig. 3. FCFM images of smoker, healthy (left) and pathological (right) cases. Notice the presence of macrophages.

and the presence of macrophages (cells which digest cellular debris), as shown in Figure 3. Note that a histogram equalization has been applied to the images, initially quite dark.

The non-smoker database includes 133 images, 31 originating from healthy subjects, and 102 from pathological ones. The smoker database includes 93 images, 60 originating from healthy subjects, and 33 from pathological ones. Because of the relatively small number of images in the non-smoker and the smoker bases, a leave-one-out cross validation process is used, which ensures unbiased generalization error estimation. It consists in extracting one sample image from the image base for validation, the rest of the base being used for learning. Classification rate is computed over all the samples.

3   Results

The influence of the LBP method parameters, i.e. scale factor (P, R), operator combination and number of resolution levels, is first assessed with a 1-NN classifier. Then, LBP-based features and knowledge-based (K) features are compared.

3.1   Influence of LBP Parameters

Correct classification rates are provided in Table 2, for LBP, VAR and combination of both operators, on the non-smoker and smoker groups using a 1-NN classifier. Analyzing the results allows to underline the following points:

- The LBP operator is better than the VAR operator, in all cases, when comparing single operators. Recognition rate reaches up to 95.7% for the smoking group and 91.73% for the non-smoking group, whereas the VAR operator only reaches 89.25% and 85.71% on the two groups respectively.
- Combining LBP and VAR operators yields better or equal results than the LBP alone, in 10 out of 14 cases (71%). A more precise analysis of the results shows that the non-smoking group especially takes benefit of the

Table 2. Classification rates (%) on non-smoker and smoker databases

P,R	Non-smoker database			Smoker database		
	$LBP_{P,R}$	$VAR_{P,R}$	$LV_{P,R}$	$LBP_{P,R}$	$VAR_{P,R}$	$LV_{P,R}$
8,1	**91,73**	85,71	90,98	93,55	86,02	92,47
16,2	87,97	82,71	89,47	94,62	83,87	94,62
24,3	84,21	81,20	88,72	94,62	75,27	93,55
8,1+16,2	90,98	84,21	90,98	**95,7**	89,25	94,62
8,1+24,3	83,46	84,21	90,98	94,62	87,1	94,62
16,2+24,3	89,47	82,71	90,98	94,62	83,87	94,62
8,1+16,2+24,3	87,21	83,46	**91,73**	94,62	86,02	**94,62**

concatenation of both complementary type of vectors. This highlights the difference of content between non smoking and smoking images. Presence of macrophages could explain the perturbation of local contrast.

– Increasing the number of resolution levels does not improve the performance. In 20 out of 24 cases (83%) where resolution is increased, recognition rates are decreased or remain unchanged. This suggests that information contained in the image is not sensitive to the scale factor.

Note that the 3-level $LV_{P,R}$ operator obtains the better compromise of performance for both smoking and non-smoking groups. This has led us to choose it for comparison with the knowledge-based feature vector in the next section.

3.2 Comparative Study between LBP-Based Features and Knowledge-Based Features

In this section, performance obtained with the knowledge-based feature set and the $LV_{P,R}$, 3-level operator, are compared. Results provided in Table 3 include recognition and error rates, as well as number of false positive (FP) and false negative (FN). False negative is the number of healthy instances that are erroneously reported as pathological and false positive is the number of pathological cases considered healthy.

LBP-based features yield better results on both databases, and whatever the classifier used. Note that, as observed in the previous study, results on the smoker

Table 3. Comparison between results obtained by 1-NN and best results obtained by the cascade for the KFV and LBPFV

	Non-smoker database				Smoker database			
	LBP (99)		K (148)		LBP (99)		K (148)	
	1-NN	SVM	1-NN	SVM	1-NN	SVM	1-NN	SVM
Recognition rate (%)	**91,73**	**91,73**	86,47	85,71	94,62	**97,85**	92,47	96,77
Error rate (%)	8,27	8,27	13,53	14,29	5,38	2,15	7,53	3,23
FP (number of images)	5	6	7	10	1	1	3	2
FN (number of images)	6	5	11	9	4	1	4	1

group are better than those obtained on the non smoker group. For the smoker group, the 2 recognition rates provided by the 2 feature sets are pretty close for a given classifier. Performance are increased by about 1.6%, vs 5.7% for the non-smoker group. This might suggest that smoker images could benefit from a higher level description in addition to the low-level, LBP-based feature vector. Nonetheless, the image database needs to be extended, and FP and FN are still quite high, which is encouraging us to continue our work.

4   Conclusions

We presented in this paper an original system for the automatic classification of normal versus abnormal endomicroscopic images of the respiratory alveolar system with an emphasis on feature extraction. The first evaluated feature vector does not require any a priori knowledge and is pixel-based. It includes a local description of spatial structure (LBP operator) and contrast (VAR operator) of the image texture. Several configurations are possible with these two operators, which have been considered for the experiments (operator combination, number of resolution levels, different values for the scale factor). The second feature set is a combination of descriptors (histogram, density, gradient, skeleton, cooccurrence matrices) deduced from a visual inspection of healthy and pathological images.

Recognition rates obtained after SVM based classification reach 91.73% and 97.85% with the LBP approach on non-smoker and smoker databases respectively, whereas they are 85.71% and 96.77% with the knowledge-based feature vector. These results show that the LBP-based feature vector is more adapted to the description of FCFM images, especially the non-smoker images. In particular, the study of the different configurations shows that combining LBP and VAR operators increases correct classification rates. This could suggest to further investigate combination of LBP and VAR operators, by computing for instance their joint distribution, or combining LBP and VAR with different values of P and R. The behavior difference of the two image databases (smoking and non-smoking) suggests that different feature sets could be also required for these two groups.

Future work will concern rendering the process real-time, so as to aid the clinician during examination in real time. Classification methods could also give information about which part of the image is the most discriminant or which part of the structure might be more altered by pathologies. A future goal will also be to discriminate between different pathologies : interstitial lung diseases (abestosis, systemic sclerosis, fibrosis, sarcoidosis), carcinomatous lesions etc.

References

1. Dibajaa, G.S., Thiel, E.: Skeletonization algorithm running on path-based distance maps. Image and Vision Computing 14, 47–57 (1996)
2. Haralick, R.M., Shanmugam, K., DinsteinI: Textural Features for Image Classification. Systems, Man and Cybernetics 3(6), 610–621 (1973)
3. Ojala, T., Pietik, M., Harwood, D.: A comparative study of texture measures with classification based on featured distribution. Pattern Recognition 29(1), 51–59 (1996)

4. Ojala, T., Pietikäinen, M., Mäenpää, T.: Multiresolution gray-scale and rotation invariant texture classification with Local Binary Patterns. IEEE Transactions on Pattern Analysis and Machine Intelligence 24(7), 971–987 (2002)
5. Petitjean, C., Benoist, J., Thiberville, L., Salaün, M., Heutte, L.: Classification of In-vivo Endomicroscopic Images of the Alveolar Respiratory System. In: IAPR Conference on Machine Vision Applications, Yokohama, Japan (2009)
6. Pratt, W.K.: Digital Image Processing, 2nd edn. John Wiley & Sons, Chichester (1991)
7. Thiberville, L., Moreno-Swirc, S., Vercauteren, T., Peltier, E., Cave, C., Bourg-Heckly, G.: In-vivo imaging of the bronchial wall microstructure using fibered confocal fluorescence microscopy. American Journal of Respiratory and Critical Care Medicine 175(1), 22–31 (2007)
8. Thiberville, L., Bourg-Heckly, G., Salaün, M., Dominique, S., Moreno-Swirc, S.: Human in-vivo confocal microscopic imaging of the distal bronchioles and alveoli. Chest Journal 132(4), 426 (2007)
9. Vapnik, V.: The nature of statistical learning theory. Springer, Heidelberg (1995)

DDoS Attack Detection Method Based on Linear Prediction Model

Jieren Cheng[1, 2], Jianping Yin[1], Chengkun Wu[1], Boyun Zhang[3], and Yun Liu[1]

[1] School of Computer, National University of Defense Technology,
410073 Changsha, China
[2] Department of Mathematics, Xiangnan University, 423000 Chenzhou, China
[3] Department of Computer, Hunan Public Security College, 423000 Changsha, China
`cjr22@163.com`

Abstract. Distributed denial of service (DDoS) attack is one of the major threats to the current Internet. The IP Flow feature value (FFV) algorithm is proposed based on the essential features of DDoS attacks, such as the abrupt traffic change, flow dissymmetry, distributed source IP addresses and concentrated target IP addresses. Using linear prediction technique, a simple and efficient ARMA prediction model is established for normal network flow. Then a DDoS attack detection scheme based on anomaly detection techniques and linear prediction model (DDAP) is designed. Furthermore, an alert evaluation mechanism is developed to reduce the false positives due to prediction error and flow noise. The experiment results demonstrate that DDAP is an efficient DDoS attacks detection scheme with more accuracy and lower false alarm rate.

Keywords: Network security; distributed denial of service; attack features; linear prediction; ARMA model.

1 Introduction

DDoS (Distributed Denial of Service) attack is one of the main threats that the Internet is facing. The defense of DDoS attacks has become a hot research topic. The DDoS attack makes use of many different sources to send a lot of useless packets to the target in a short time, which will consume the target's resource and make the target's service unavailable. Among all the network attacks, the DDoS attack is easier to carry out, more harmful, hard to be traced, and difficult to prevent, so its threat is more serious.

The detection of attacks is an important aspect of defense of DDoS attack, and the detection results can affect the overall performance of attack defense. Recently, the DDoS attacks are tending to use true source IP address to perform an attack [1], and can make serious destroy on the victims by using periodically non-suspicious low-rate attack flows [2-5]. So it has become more difficult to distinguish between the normal network flow and the attack flow, which makes an early and accurate detection more difficult. Most current detection methods employ some specific attack feature to identify an attack [6-13], which can improve the detection accuracy. However, they usually fail when there is a change in attacking strategies and techniques.

This paper proposes the IP Flow feature value (FFV) algorithm based on the essential features of DDoS attacks, such as the abrupt traffic change, flow dissymmetry, distributed source IP addresses and concentrated target IP addresses. A simple and efficient detection scheme based on anomaly detection techniques and ARMA linear prediction model (DDAP) is designed. Furthermore, an alert evaluation mechanism is developed to reduce the false positives due to prediction error or flow noise.

2   Related Work

Previous methods for DDoS attack detection can be classified into three categories as: abrupt traffic change based, flow dissymmetry based and distributed source IP addresses based. The methods [6-9] based on abrupt traffic change are the most popular ones, as DDoS attacks will cause obvious deviation from normal flows in traffic volume.However, organizing many different attacking sources, the attacker can send out attack packets at different time, different types of flow, different size of packets and different rate, which will follow the statistical features of normal flow and avoid being detected. Furthermore, the traffic offset can not distinguish DDoS attack flow from normal network congestion flow. The methods [10-11] are based on flow dissymmetry. However, the coming-in and going-out rate of normal flow is not balanced sometimes, such as audio/video flows. What is more, attackers may use random cheating source IP addresses, send out the equivalent amount of SYN packets and FIN/RST packets, or simulate normal flow to send out attack packets by many attack sources without losing the power of attack.The methods [12-13] made use of the distribution of attacking sources. However, methods of this type are insensitive to the attack which has the characters of small amount of source IP addresses and big enough attack traffic volume, and are easy to be disturbed by normal flows.

Based on the analysis above, it can be concluded that using a single feature of attack flows is not enough for an accurate and efficient detection. The approaches based on multi-features of DDoS attack flows can detect DDoS attacks more effectively [14].

3   DDoS Attack Features Analysis

A typical DDoS attack employs many hosts to flood the target in a distributed and cooperative way, as depicted figure 1. Thus it usually has the following four features:

1. A burst in traffic volume. There will be a burst in traffic volume in a certain period no matter the attack is high-rate or low-rate. The low-rate DDoS attack will have the burst in an extremely short time periodically, but the average traffic volume rate is similar with that of normal conditions. Thus compared with high-rate DDoS ones, the low-rate DDoS attacks are more difficult to be detected and distinguished from network congestion.
2. Dissymmetry. The mapping relationship between the set of source IP addresses and destination IP addresses will appear to be a many-to-one mapping for an attack with many different attack sources.

3. Distributed Source IP Addresses. To avoid being detected and traced, the attack sources will use the forging varies source IP addresses, thus cause a change of the distribution of source IP   addresses, which provide another clue for detection.
4. Concentrated Destination IP Addresses. DDoS attacks will lead to a concentrated set of destination IP addresses.

Above four features will be quite different under different attacks and applications. Thus the FFV algorithm is proposed in this paper, which considers all of them.
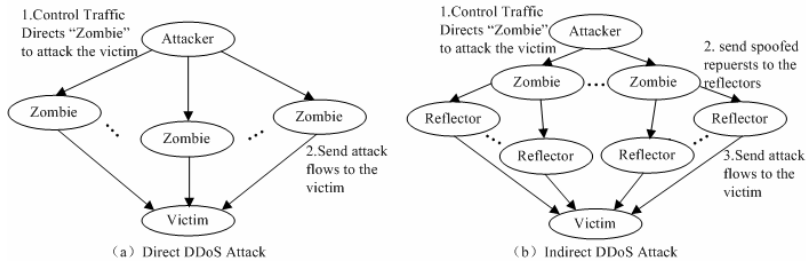
Fig. 1. Typical DDoS attack

4   Flow Features Algorithms

The FFV algorithms reduce firstly the interference of normal traffic based on attack features, and then use a feature value to reflect the mentioned four features, which can be used to detect DDoS attacks with different kinds of attack strategies.

Definition 1 (flow Feature Value). The network flow F in the certain time span T is given in the form of $<(t_1, S_1, D_1, P_1), (t_2, S_2, D_2, P_2),\ldots, (t_n, S_n, D_n, P_n)>$. For the ith packet p, t_i is the time, S_i is the source IP address, D_i is the destination IP address, and P_i is the destination port. Classify all the packets by source IP address and destination IP address, which means all packets in a certain class share the same source IP address and destination IP address. A class which is consisted of packets with a source IP address A_i and a destination IP address A_j is noted as $SD(A_i, A_j)$.

Carry out the following rules one by one to the above mentioned classes:

If there are two different destination IP address A_j, A_k, which makes class $SD(A_i, A_j)$ and class $SD(A_i, A_k)$ both nonempty, then remove all the class with a source IP address A_i.

If there is only one class $SD(A_i, A_j)$ containing the packets with a destination IP address A_j, then remove all the classes with a destination IP address A_j.

Assume that the remaining classes are $SDS_1, SDS_2,\ldots, SDS_l$, classify these classes by destination IP address, that is all the packets with the same destination IP address will be in the same class. The class made up of packets with the same destination IP address A_j is noted as $SDD(A_j)$, these classes are $SDD_1, SDD_2,\ldots, SDD_m$, the FFV(Flow Features Value, FFV) is defined as:

$$IFFV_F = \frac{1}{m}(\sum_{i=1}^{m} W(SDD_i) - m)$$ (1)

Where $W(SDD_i) = SIP(SDD_i) + \theta_1 \sum_{j=1}^{SIP(SDD_i)} OA(Pack(A_j)) + (1-\theta_1)OB((Port(SDD_i)-1)), (0 \leq \theta_1 \leq 1)$,

and $SIP(SDD_i)$ is the number of different source IP addresses in the class SDD_i

$$OA(x) = \begin{cases} Pack(A_j) & Pack(A_j)/\Delta t \geq \theta_2 \\ 0 & Pack(A_j)/\Delta t < \theta_2 \end{cases},$$

$Pack(A_j)$ is the number of packets with a source IP address A_j in the class SDD_i.

$$OB(x) = \begin{cases} Port(SDD_i) & Port(SDD_i)/\Delta t \geq \theta_3 \\ 0 & Port(SDD_i)/\Delta t < \theta_3 \end{cases},$$

Δt sampling time interval, θ_2, θ_3 are thresholds, $Port(SDD_i)$ is the number of different port in the class SDD_i. The value of θ_1, θ_2 and θ_3 turns out to be important, and they need to be set according to the maximal values obtained by a statistical way in the normal network flows and application condition.

In order to analyze the state features of the network flow F more efficiently and exclude the interference of normal flow, the FFV-based algorithm classifies the packets of F by source IP address and destination IP address. According to the dissymmetry of attack flows, the source IP address A_i is in an one-to-many mapping if there are different destination IP addresses A_j and A_k which make the classes $SD(A_i, A_j)$ and $SD(A_i, A_k)$ nonempty, thus the classes with source IP address A_i can be removed. After that, if there is a destination IP address A_k makes A_i and A_j in $SD(A_i, A_k)$ and $SD(A_j, A_k)$ the same, then the destination IP address A_k is in a one-to-one mapping, thus the class with packets going to the destination IP address A_k can be removed. Then classes with the same destination IP address SDS_i will be merged into the class SDD_j as a single class, which can reflect the asymmetry of DDoS attack flow as well as a decrease in the disturbance of normal flow. DDoS attack is a kind of attack that sends useless packets to the attack target from many different sources in the hope of exhausting the resources of the target. This act can produce lots of new source IP addresses in a short time, which will lead to an abnormal increase of $SIP(SDD_i)$ for some classes of F; the occurrence possibility of DDoS attack increases when many different sources are sending useless packets with a lot of different destination ports to a destination IP address, and $Port(SDD_i)$ will increase abnormally; $Pack(A_j)$ will increase abnormally when in a certain time period a large amount of packets are sent from a source IP address to a destination IP address. Thus we obtain $W(SDD_i)$ by weighted $SIP(SDD_i)$, $Port(SDD_i)$ and $Pack(A_j)$ in the class of SDD_i. As the DDoS attacks will increase the ratio of the sum of $W(SDD_i)$ and the number m of different destination IP addresses dramatically in a short time, thus the FFV will form a new flow state different from the normal one. In summary, FFV can reflect the characteristics of DDoS attack including the burst in the traffic volume, asymmetry of the flow, distributed source IP addresses and concentrated destination IP address.

5   DDoS Attack Detection Based on ARMA Prediction Model

The attackers can simulate the normal flows without losing its attacking ability, which will reduce the abnormal changes of Attack flow states. Moreover, the normal traffic will have abnormal changes at some times. So we employ the anomaly detection techniques and linear prediction model based on FFV to detect DDoS attacks.

5.1   Characteristics of FFV Time Series

Sampling the network flow F with a time interval Δt, then calculate the FFV of each sampling. After N samplings, we can get the time series FFV of sample A, $A(N,\Delta t)=\{FFV_i, i=1,2,\ldots,N\}$, where N is the length of the series.

Let $FFV_i=a_i$, $i=1,2,\ldots,N$, then the k lags autocorrelation coefficients (AC) of sample A can be represented as:

$$\rho_k = \frac{\sum_{i=1}^{N-k}(a_i-\bar{a})(a_{i+k}-\bar{a})}{\sum_{i=1}^{N}(a_i-\bar{a})^2}, \quad k = 0,1,2,\ldots,N-1 \tag{2}$$

where $\bar{a}=\sum_{i=1}^{N}a_i/N$ is the mathematical expectation of sample A.

The k lags partial autocorrelation coefficients (PAC) of sample A can be represented as:

$$\phi_{k,k} = \begin{cases} \rho_1 & k=1 \\ \dfrac{\rho_k - \sum_{j=1}^{k-1}\phi_{k-1,j}\rho_{k-j}}{1-\sum_{j=1}^{k-1}\phi_{k-1,j}\rho_j} & k=2,3,\ldots,N-1 \end{cases} \tag{3}$$

Where $\phi_{k,j} = \phi_{k-1,j} - \phi_{k,k}\phi_{k-1,k-j}$, $j=1,2,\ldots,k-1$.
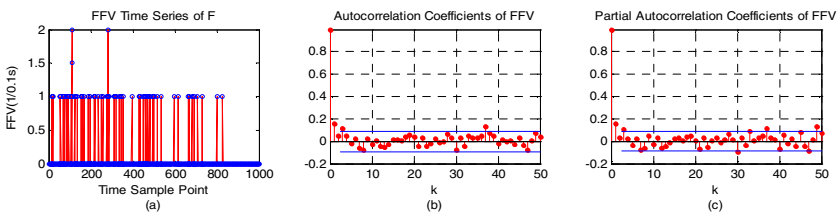

Fig. 2. FFV time series of Network Flow

We obtain an arbitrary network flow F from the MIT datasets[15], the FFV time series of sampling is depicted in figure 2(a), the 50 lags autocorrelation coefficients is depicted in figure 2(b), The 50 lags partial autocorrelation coefficients is depicted in figure 2(c). The stationary of the FFV time series can be judged by the autocorrelation coefficients, and the model identification and order determination can be concluded by the autocorrelation coefficients and the partial autocorrelation coefficients. Figure 2 shows that the FFV time series of F is a stationary time series, and the curves of AC

and PAC are tailing respectively, thus the correlation of FFV time series can be analyzed efficiently by ARMA model with small prediction error variance, which is suitable for short time prediction. In this paper, we employ the ARMA($2p$, $2p$-1) model to analyze the FFV time series of network flows, in which p is order. If p is too large, the computation cost will be too high. To perform a real-time detection, we assign p the value 1, thus the ARMA(2,1) model.

5.2 Smoothing FFV Time Series

Assume that the slide window size is n, and the sampled FFV time series are a_1,…, a_i,…, a_n. To filter random white noise from the data, to make the FFV time series smoother and establish the ARMA model, a_1,…, a_i,…, a_n are transformed into A_1,…, A_i,…, A_n by moving average smoothing method before the ARMA model is established to predict the (n+1)th FFV. In real-time applications, the sliding window will move forward once when one prediction is completed.

5.3 ARMA Modeling and Forecasting

The processed time series are A_1,…, A_i,…, A_n, establish the ARMA model [16] as follows:

$$\varphi(B)\, A_i = \psi(B)\varepsilon_i \tag{4}$$

where B is lag operator, ε_i is Gaussian white noise, its average is zero, its variance is σ_ε^2. Thus AR coefficient polynomial is $\varphi(B)=1-\varphi_1 B-\varphi_2 B^2$, MA coefficient polynomial is $\psi(B)=1-\psi_1 B$, where φ_1, φ_2, ψ_1 are estimated parameters. Considering the precision of estimation of parameters and computational cost, we estimates φ_1, φ_2, ψ_1, σ_ε^2 by least-squares estimators, and then the ARMA(2, 1) model is as follows:

$$A_t=\varphi_1 A_{t-1}-\varphi_2 A_{t-2}+\varepsilon_t-\psi_1\varepsilon_{t-1} \tag{5}$$

For decreasing the false positives due to prediction error, we use the inverse function to establish the 1-step-ahead prediction model. The inverse function of ARMA(2,1) is as follows: $\beta_1=\varphi_1 - \psi_1$; $\beta_2=\varphi_2+\beta_1\psi_1$; $\beta_j=\beta_{j-1}\psi_1,(j\geq3)$, and then The 1-step-ahead-prediction model is represented as:

$$A_t(1) = \sum_{j}^{m}\beta_j A_{t-j} \tag{6}$$

where m is the number of observed values before A_t, it is assigned according to the requisite for a precision, the prediction error is as follow: $e_t=-\sum_{j}^{m}\beta_j A_{t-j}$, $\beta_0=-1$.

For increasing the prediction precision, we propose a statistical correction method for prediction error and compensation. Considering the features of FFV time series, we replace the negative prediction value with the minimal statistical FFV of the current network flow and then round the prediction value to the nearest integer.

5.4   Abnormal Detection and Alarm Valuation Mechanism

In the DDAP method, we use the FFV time series of normal flows accumulated in the sliding window time to estimate the model parameters and establish the ARMA(2,1) 1-step prediction model, then use the model to predict the arriving FFV value of test flows at time t, and compare it with the real FFV value of test flows. We can detect anomaly by judging whether the absolute difference between them has exceeded the predefined threshold. Assume the threshold is U, and the absolute difference between real FFV value A_t and the predicted FFV value A'_t is $y_t=|A_t - A'_t|$,   when $y_t - b > U$, there is anomaly in the network flow,   where $b=E[e_t^2]$, is the prediction error.

The cause of abnormal changes in FFV states of network flows includes the DDoS attack as well as the network flow noise, congestion and some other reasons. So the number of false positives will increase if we do not make a study of the reason of abnormal changes of FFV states. Frequent false positives will decrease the detection efficiency and disturb normal application, so it's very important to decrease the number of false positives. In this paper, we propose a simple alert evaluation mechanism based on the alert frequency and time interval, which will only generate alerts when the $C(C≥1)$ anomalies are detected in a designated time interval $\Delta T(\Delta T≥0)$. The value of ΔT and C turns out to be very important, larger ΔT and C can decrease the risk of false positives, such as flash traffic, but the time efficiency will be decrease too. They need to be set dynamically according to the network security situations when in application.

6   Experiments and Results

The experiment used the normal flow data in 1999 and DDoS flow data LLDoS2.0.2 in 2000 from MIT Lincoln lab [15]. The normal flow samples were from normal flow data, and the attack flow samples were from DDoS flow data.
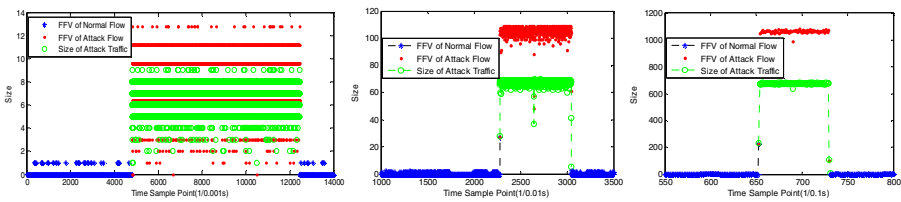

Fig. 3. FFV time series of LLDoS2.0.2 network flow

When the sampling interval Δt is 0.001s, 0.01s, 0.1s, the FFV time series obtained by multiple sampling and calculation is depicted in figure 3; When the sampling time interval Δt is 0.001s, 0.1s, the FFV time series obtained by multiple sampling and calculation is depicted in figure 4. From figure 3 & figure 4, we can see that, FFV time series are sensitive to attack flows and they can magnify size of the attack traffic using randomized destination ports. Meanwhile, they are steady when the network flow is normal. Thus the FFV can well reflect different state features of normal network flow and DDoS attack flow.
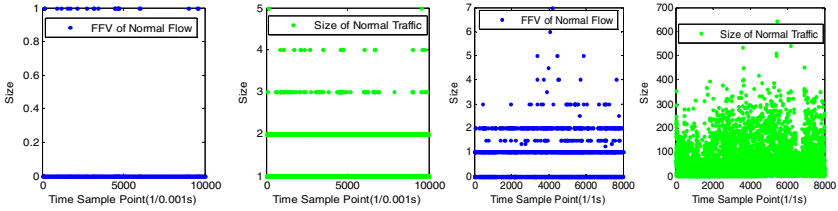
Fig. 4. FFV time series of 1999 normal flow

The solid line in figure 5 shows the FFV time series of a normal flow when the sampling time interval Δt was 0.1s, used the first 400 FFV values to estimate the model parameters, and established the ARMA(2,1) 1-step prediction model, then used the model to predict the following FFV values, the result of 100 times prediction is depicted by the dashed line. We can see from figure 5 that the predicted FFV time series match the real values well, which argues that the ARMA(2,1) model work well as a short time related prediction model, and can be used as a real-time prediction model.
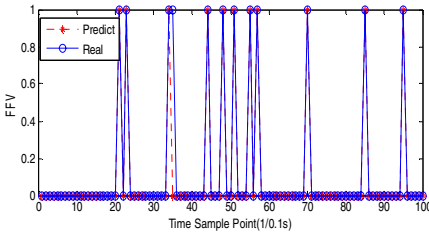

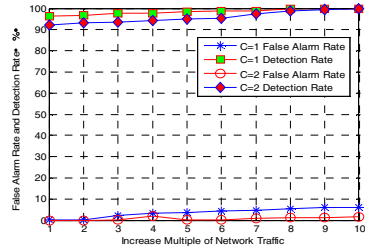**Fig. 5.** FFV time series of ARMA predictions   **Fig. 6.** DDAP detection method

To show the advantages of DDAP method, we set the threshold θ_1 as 0.5, θ_2, θ_3 as 3/ms, Δt=0.01s, U=2, ΔT=0.02s and did multiple experiments with C=1 and C=2, respectively. When C=1, (1) sampled 200 FFV values of normal network flows to estimate the model parameters, and established the ARMA(2,1) 1-step prediction model. (2) Mixed the attack flows and normal flows with a ratio 1/10, sampled and calculated the FFV time series , then used the DDAP method to detect, and got a group of detection rate; (3) Remained the normal flows, duplicated the attack traffic volume, and got another group of detection rate using DDAP method; (4) Repeated the step 3 for 8 times and got 8 groups of detection rate; (5) Sampled and calculated the FFV time series of normal flows and used the DDAP method to detect then obtained a group of false alarm rate; (6) Duplicated the normal traffic volume and got another group of false alarm rate by DDAP method; (7) Repeated step 6 for 8 times and got 8 groups of false alarm rate.

When C = 2, used the above procedures to do 20 groups of experiments. The results are depicted in figure 6.

Figure 6 shows that, the detection rate of DDAP method increases from 96.4% to 100% when the attack traffic volume increases, and the average detection rate is 98.5% in 10 groups of experiments. This means that DDAP can identify the DDoS flows with a high precision, and be sensitive to the abnormal phenomenon caused by attack flows, which will help an early discovery of attacks. The false negatives in the experiments come mainly from two aspects: (1) At first, the normal flow is greater than the attack flow, which makes the FFV states lean to normal ones, thus makes the difference too small for detection. (2) The increase of attack traffic volume and random noise deviate the network flows' states. The false alarm rate of DDAP method increases slowly from 0.1% to 5.9% with the increase of normal traffic volume, the average false alarm rate is 3.6% in 10 groups of experiments. This means that the DDAP method can identify the normal flows correctly and exclude the interference of the increase of normal traffic volume. The false positives mainly come from the prediction error of FFV and random noise. When C = 2, the detection rate of DDAP method increase from 92.0% to 100%, the false alarm rate increase from 0.0% to 1.5%, among all the 10 groups experiments, the average detection rate is 96.1% and the average false alarm rate is 0.8%. We can see the threshold settings and attack density influence the detection results, in which the setting of threshold is the key. To increase the detection precision, the noise and requirements of certain applications should be considered in general and then set the threshold appropriately.

In summary, FFV can well reflect the features of DDOS attacks; the ARMA(2,1) has a high prediction precision and can predict the FFV values of network flows in real-time; the DDAP method can decrease the number of false positives and detect the DDoS attacks efficiently.

7   Conclusions

In this paper, we propose the FFV algorithm based on the the abrupt traffic change, flow dissymmetry, distributed source IP addresses and concentrated target IP addresses, which can reflect four essential features of DDoS attacks, then we present a novel DDoS detection method DDAP based on FFV linear prediction. DDAP makes use of the FFV time series of network flows to describe the essential features of network flow states that are changing with time. To detect DDoS attacks is to identify the state features of FFV time series. Finally, we employ an anomaly detection method based on ARMA(2,1) prediction model and an alert evaluation mechanism to realize DDoS attack detection.

Experiment results argue that, FFV can well reflect the different state features of DDoS attack flows which are different from normal flows, the DDAP method can identify the anomaly in network flow caused by DDoS attacks efficiently, and increase the detection rate; it can also distinguish between normal network flow and abnormal network flow containing DDoS attacks, and decrease the false alarm rate.

In our future work, we will make a detail study of how to set all kinds of parameters in different application scenarios adaptively.

Acknowledgement

References

1. Handley, M.: DoS-resistant Internet Subgroup Report. Internet Architecture WG. Tech. Rep. (2005),
   `http://www.communications.net/object/download/1543/doc/`
   `mjh-dos-summary.pdf`
2. Kuzmanovic, A., Knightly, E.W.: Low-rate TCP-targeted Denial of Service Attacks and Counter Strategies. IEEE-Acm Transactions on Networking 14(4), 683–696 (2006)
3. Macia-Fernandez, G., Diaz-Verdejo, J.E., Garcia-Teodoro, P.: Evaluation of a Low-rate DoS Attack against Iterative Servers. Computer Networks 51(4), 1013–1030 (2007)
4. Macia-Fernandez, G., Diaz-Verdejo, J.E., Garcia-Teodoro, P.: Evaluation of a Low-rate DoS Attack against Application Servers. Computers & Security 27(7-8), 335–354 (2008)
5. Kumar, V.A., Jayalekshmy, P.S., Patra, G.K., et al.: On Remote Exploitation of TCP Sender for Low-Rate Flooding Denial-of-Service Attack. IEEE Communications Letters 13(1), 46–48 (2009)
6. Cheng, C.M., Kung, H.T., Tan, K.S.: Use of Spectral Analysis in Defense against DoS Attacks. In: Proceedings of IEEE GLOBECOM, pp. 2143–2148 (2002)
7. Manikopoulos, C., Papavassiliou, S.: Network Intrusion and Fault Detection: A Statistical Anomaly Approach. IEEE Commun. Mag., 76–82 (2002)
8. Lakhina, A., Crovella, M., Diot, C.: Diagnosing Network-Wide Traffic Anomalies. In: Proceedings of ACM SIGCOMM, Portland, Oregon, USA (2004)
9. Sanguk, N., Gihyun, J., Kyunghee, C., et al.: Compiling Network Traffic into Rules Using Soft Computing Methods for the Detection of Flooding Attacks. Applied Soft Computing, 1200–1210 (2008)
10. Keunsoo, L., Juhyun, K., Ki, H.K., et al.: DDoS Attack Detection Method Using Cluster Analysis. Expert Systems with Applications, 1659–1665 (2008)
11. Abdelsayed, S., Glimsholt, D., Leckie, C., et al.: An efficient Filter for Denial-of Service Bandwidth Attacks. In: Proceedings of the 46th IEEE GLOBECOM, pp. 1353–1357 (2003)
12. Lakhina, A., Crovella, M., Diot, C.: Mining Anomalies Using Traffic Feature Distributions. In: Proceedings of ACM SIGCOMM, Philadelphia, Pennsylvania, USA (2005)
13. Peng, T., Leckie, C., Kotagiri, R.: Proactively Detecting Distributed Denial of Service Attacks Using Source Ip Address Monitoring. In: Proceedings of the Third International IFIP-TC6 Networking Conference, pp. 771–782 (2004)
14. Cheng, J., Yin, J.P., Liu, Y., et al.: DDoS Attack Detection Algorithm Using IP Address Features. In: Proceedings of FAW 2009. LNCS. Springer, Heidelberg (2009)
15. `http://www.ll.mit.edu/mission/communications/ist/corpora/`
    `ideval/data/index.html`
16. Ljung, L.: System Identification: Theory for the User. Prentice-Hal PTR, Upper Saddle River (1999)

Modified AES Using Chaotic Key Generator for Satellite Imagery Encryption

Fahad Bin Muhaya, Muhammad Usama, and Muhammad Khurram Khan

Prince Muqrin Chair for IT Security (PMC)
King Saud University, Riyadh
Kingdom of Saudi Arabia
Fmuhaya@ksu.edu.sa, khurram.khan@scientist.com

Abstract. In this paper, we propose a new modified version of Advanced Encryption Standard (AES) using chaotic key generator for satellite imagery security. We analyze and examine the Modified AES and chaotic key generator to enhance the key space and sensitivity, performance, and security level for reducing the risk of different attacks. The chaotic key generator utilizes multiple chaotic maps named as Logistic, Henon, Tent, Cubic, Sine and Chebyshev. The proposed algorithm presents numerous interesting and attractive features, such as a high level of security, large enough key-space with improved key sensitivity, pixel distributing uniformity and an acceptable encryption and decryption speed. The presented algorithm is ideal for real-time applications to deal with redundant, bulky, complex and stubborn satellite imagery.

1 Introduction

The space science and technologies along with fast growing Information technologies (IT) have recently attracted scientists, researchers, and industrial communities. Rapid development and advancements in these technologies present new opportunities for research, development, and improvement. These technologies together play a vital role in increasing productivity with new modernization of technologies, reducing cost, and creating practical virtual collaborative environments for addressing the emerging challenges. IT and Space Science scientists have grappled with the challenges for a decade or more and have identified the utility of satellite imagery as major sources of consistent, continuous data for research, academics, industries, military, map production, agriculture, forestry, planning of national land, atmospheric and various fields at a variety of spatial and temporal scales. The satellite imagery along with GIS, remote sensing and geospatial technologies contribute for preserving and protecting the public, private and nation's sensitive and critical infrastructure and respected data [1]. The unlawful, unofficial, unauthorized and illegal usage/access/intrusion into satellite imagery have raised the importance of information security, in order to keep protected critical and confidential satellite imagery storage and transmission process secure, dependable, trustworthy and reliable.

The cryptography by far the most widely adopted security mechanism is used to provide the data storage and transmission process (usually based on CD/DVD-ROM

hardcopy or on shared network environment) secure and reliable from unauthorized access and illegal use [2]-[4]. The objective of cryptography is to develop or build up a secure cryptosystem, which must have two basic functions encryption and decryption (reverse of encryption).

To fulfill such security and privacy needs, satellite imagery encryption is important. There are number of encryption algorithms available such as DES, Triple-DES, International Data Encryption Algorithm (IDEA) and RSA (developed by Rivest, Shamir and Adleman) [5]. These traditional algorithms are very complex for implementation and have weakness; especially DES and Triple-DES have small key size. They require extensive computation power and not very fast to handle or deal with large sized, redundant, bulky, complex and stubborn satellite imagery.

In order to accept these problems as a challenge the systems based on AES, using the Rijndael algorithm were proposed. AES is very fast symmetric-key block cipher which operates with any combination of block and key size 128, 192 or 256 bits. The AES key size is long enough to resist against brute force attack. It would take millions of years to break the key on the fastest computers presently available. AES is flexible, reliable and efficient for its hardware implementation as well [6]-[9]. AES performs well for various applications that require high-speed, computation and processing e.g. flash drives, smart cards, cellular phones, image-video encryption etc. [10]

This paper proposes new satellite imagery encryption technique as a modification of AES algorithm. The modification is done by adding chaotic key generator based on multiple chaotic maps named as Logistic, Henon, Tent, Cubic, Sine and Chebyshev for improving the key space, key sensitivity, performance and security. Chaotic systems are very suitable for cryptography because of sensitivity, ergodicity, random behavior and unstable periodic orbits with long periods. The desirable diffusion and confusion properties for cryptosystem can be achieved through iterations using chaotic system [14]. Initial work on chaos-based cryptosystems was based on chaotic dynamical system. The concepts of chaotic dynamical system were associated with synchronization of two chaotic systems and controls [16]. Several methods and techniques have been proposed in this domain to synchronize chaotic systems. Some typical forms have been brought up, which includes chaotic masking, shift keying and modulation using inverse systems [16]-[20].

2   Modified AES Algorithm Using Chaotic Key Generator

In cryptosystem, the confusion and diffusion properties are used to reduce the correlation among pixel values and to change the original location of input pixel values, respectively. In other words, diffusion changes the location of the pixels which were modified in confusion process. To achieve these properties number of symmetric-key chaos-based data encryption algorithms have proposed since 1990s [21]-[30]. These algorithms mainly based on single chaotic map (for generating secrete key) in encryption/decryptions process.

The proposed satellite imagery encryption algorithm is a modified version AES algorithm which uses multiple chaotic maps for key generation of variable length e.g. 128, 192 and 256 bits as shown in Figure 1. The idea of using multiple chaotic maps in key generation process is to improve the security level of the AES by incorporating

the large size and complex secret key. Due to this, AES becomes more complex and secure from attacks. The algorithm takes satellite image and secret key as input and produces scrambled and unreadable satellite image. Similarly for decryption process, it takes encrypted satellite image along with secret key for getting original satellite image back in original form. The detailed description of Modified AES using chaotic key generator is given in the next subsections.
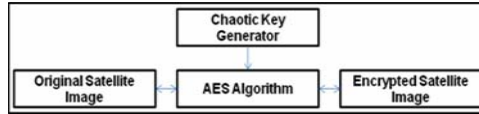

Fig. 1. Block Diagram of Modified AES Algorithm

2.1   AES Satellite Image Encryption

AES is an iterative symmetric-key block cipher for data encryption developed by Joan Daemen and Vincent Rijmen and further known as Rijndael algorithm. The AES algorithm is flexible to support variable block and key size 128, 192 or 256 bits. AES supports 128, 192 or 256 bits input block, output block and State. The State is divided into four operational blocks and organized in 4x4 bytes of array where number of columns in state matrix is $NC = 4$ and with key size K of length $NK = 4$, 6 or 8 respectively, which reflects the number of 32-bits or number of columns in the Key.

Table 1. Key-Block-Round Combinations

	Key Length (NK words)	Block Size (NC words)	No. of Rounds (NR)
AES-128	4	4	10
AES-192	6	4	12
AES-256	8	4	14

The key size suggests number of rounds NR required by algorithm to perform encryption/decryption operation. If $NR = 10$ then $NK = 4$, if $NR = 12$ then $NK = 6$ and if $NR = 14$ then $NK = 8$. The supported key size, block size and number of rounds combinations are given in Table 1. The AES key size is long enough to resist against the attacks like brute force. AES is flexible, reliable and have good performance for hardware implementation as well [7],[10]. AES performs well for real time applications that need high level of security on resource hungry environment [11]. AES uses round function for encryption/decryption. The round function put together four transformation operations Byte substitution transformation, Shifting rows transformation, MixColumns transformation and AddRoundKey transformation. The decryption structure has exactly the same sequence with inverse transformation as the one in the encryption structure. The encryption procedure consists of several steps as shown by Figure 2.
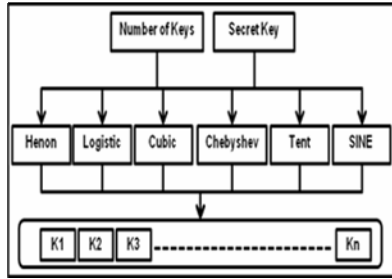
Fig. 2. Block Diagram of Chaotic Key Generator

2.2 Chaotic Key Generator

The chaotic key generator utilizes six chaotic maps named as Logistic map, Tent map, Henon map, Sine map, Cubic map and Chebyshev map. It supports variable length secret key of sizes 128, 192 and 256 bits as supported by AES algorithm. The secret key can be represented as:

$$K = K_1 K_2 K_3 K_4 K_5 .... K_{S/8},$$
(1)

Where S is the key size for the secret key K. For example, if key size S is 256 bits then secret key K is equals 32 bytes (that is equal to 256/8). Table 2, shows the experimented chaotic maps respected governing equations, and their parameter values. The chaotic sequences of each chaotic map is in real number, so first generator transform these real numbers into sequence of bits (0s and 1s) and than in bytes format for getting secret key.

The hexadecimal mode is used to define the secret key for initial condition IC (for each chaotic map) which generates different bit sequences from real numbers. The secret keys are generated by the following equations:

$$N = \sum_{i=1}^{S/8} (K_i / 256)$$
(2)

$$IC = N - \lfloor N \rfloor,$$
(3)

where K_i is the ith key value in decimal equivalent of the secret key, $\lfloor N \rfloor$ is the floor of the value N, S is the key size and IC is the initial condition value which is transformed back into real numbers. The block diagram of chaotic key generator is shown in Figure 3. The generator takes two parameters, first is secret key and second is n which is number of keys (where n must be greater or equal to one). Suppose $n = 10$ than proposed algorithm automatically creates ten keys for each map using single input secret key and parameters (as given in Table 2). To combine these keys, key generator performs XOR operation as:

$$K_i = BK_i \oplus LK_i \oplus CK_i \oplus SK_i \oplus HK_i \oplus TK_i,$$
(4)

Where $i = 1,2,3,..............,n$ and $n \geq 1$ and

Table 2. Governing equations and system parameters values in chaotic range for the chaotic maps used in the proposed algorithm

Chaotic	Governing Equations	Parameter
Chebyshev	$x_{n-1} = \cos(\lambda \cos^{-1}(x_n))$	$\lambda = 4$
Logistic	$x_{n+1} = \lambda x_n(1 - x_n)$	$\lambda = 4$
Cubic	$x_{n+1} = \lambda x_n(1 - x_n^2)$	$\lambda = 2.59$
Sine	$x_{n+1} = \lambda \sin(\pi x_n)$	$\lambda = 0.99$
Henon	$x_n = 1 + \lambda(x_{n-2} - x_{n-3}) + ax_{n-2}^2$	$\lambda = 0.3$ 1.07 $\leq a \leq$ 1.09
Tent	$x_{n+1} = \begin{cases} x_n/\mu & if & x_n \leq \mu \\ 1 - x_n/1 - \mu & if & x_n \geq \mu \end{cases}$	$\mu = 0.4$

The modified AES algorithm is key dependent, so the encryption time will always depend on time taken to generate n number of keys before performing actual encryption/decryption process. For analysis purpose, we have calculated the encryption time using ten keys (for Boston satellite image shown in Figure 3.

Chebyshev =	*BK* =	bk_1,	bk_2,	....	bk_n
		$\oplus$,	$\oplus$,	....	$\oplus$
Logistic =	*LK* =	lk_1,	lk_2,	....	lk_n
		$\oplus$,	$\oplus$,	....	$\oplus$
Cubic =	*CK* =	ck_1,	ck_2,	....	ck_n
		$\oplus$,	$\oplus$,	....	$\oplus$
Sine =	*SK* =	sk_1,	sk_2,	....	sk_n
		$\oplus$,	$\oplus$,	....	$\oplus$
Henon =	*HK* =	hk_1,	hk_2,	....	hk_n
		$\oplus$,	$\oplus$,	....	$\oplus$
Tent =	*TK* =	tk_1,	tk_2,	....	tk_n
		=,	=,	....	=
Keys =	*K* =	k_1,	k_2,	....	k_n

3   Statistical Analysis of Proposed Algorithm

Claude E. Shannon stated in his masterpiece paper [31], "It is possible to solve many kinds of ciphers by statistical analysis," and therefore he suggested two methods diffusion and confusion for preventing realization of the great influence in statistical analysis. In next subsection, it is demonstrated that the AES image encryption algorithm has good confusion and diffusion properties. This has been shown by performing histogram and information entropy analysis on original and encrypted images of Boston.

3.1   Histogram Analysis

To prevent the leakage of information to attackers, it is important to ensure that encrypted and original images do not have any statistical similarities. The histogram analysis clarifies that, how the pixel values of image are distributed. We performed analysis on several satellite images having different contents and sizes using proposed algorithm. One typical example among them is shown in Figure 4 of Boston satellite image. The histogram of original image contains large sharp rises followed by sharp declines as shown in Figure 3(b). And the histogram of the encrypted image as shown in Figure 4(b), has uniform distribution which is significantly different from original image, and has no statistical similarity in appearance. Therefore, the proposed algorithm does not provide any clue for statistical attack.
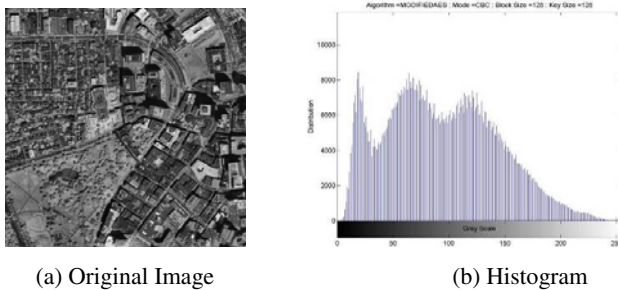

(a) Original Image   (b) Histogram

Fig. 3. Boston Original Image

(a) Encrypted Image   (b) Histogram of Encrypted Image (c) Decrypted Image
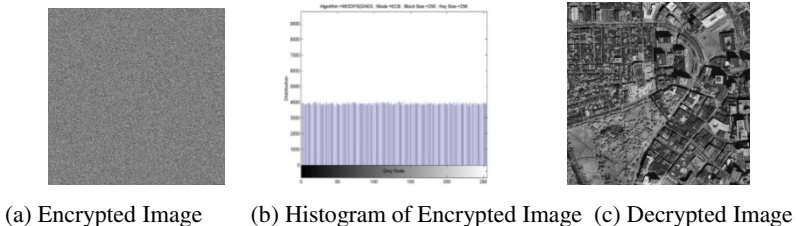
Fig. 4. Application of Modified AES Algorithm to Boston image in ECB Mode

3.2   Information Entropy

The Shannon entropy or information entropy is a measure of the uncertainty associated with a random variable which was first introduced by Claude E. Shannon in his 1948 paper "A Mathematical Theory of Communication" [32]. It quantifies the information contained in a data, usually in bits or bits/symbol. The entropy is concerned with error-correction, data compression, cryptography, communications systems, and related topics. To calculate the entropy $H(m)$ of a source m, we have: [32]

$$H(m) = \sum_{i=0}^{2N-1} P(m_i) \log_2 \frac{1}{P(m_i)},$$

(5)

Where $P(m_i)$ represents the probability of symbol m_i and the entropy is expressed in bits. Let us suppose that the source emits 2^8 symbols with equal probability, i.e. $m = \{m_1, m_2, ...m_{2^8}\}$. After evaluating above equation, we obtain its entropy $H(m) = 8$, corresponding to a truly random source. Let us consider the experimental results of the proposed algorithm on original encrypted and decrypted gray-scale Boston satellite image, the number of occurrence of each outcome data is recorded and the probability of occurrence is computed. The calculated information entropy values are given in Table 3.

Table 3. Computed Information Entropy using proposed algorithm

Block Size	Key Size	Mode	Entropy of Encryption Process	Entropy of Decryption Process
256	256	CBC	7.9998	7.5588
256	256	CFB	7.9998	7.5588
256	256	ECB	7.9998	7.5588

4   Security Analysis of Proposed Algorithm

The Modified AES algorithm is implemented using Microsoft VB.Net programming language and MATLAB, and experimental results are observed on a Pentium-IV 1.8 MHz PC with 1.46 GB RAM. Results of some experiments are given to prove the efficiency and security for satellite imagery. For the brevity, we have not given all the results here. We use the gray-scale Boston satellite image of size 1000 x 1000, depicted in Figure 3. The secret key "123456GHIJKLMNOPQRSTUV WXYZ[\]^_ `" (in ASCII) is used for encryption and decryption having size 256-bit (32 Bytes). The encrypted and decrypted images using modified AES are depicted in Figure 5. As shown that, the encrypted images are totally scrambled and unreadable.

4.1   Key Space Analysis

To secure the sensitive and critical satellite imagery from brute-force and similar attacks, the encryption algorithm should have large key space and sensitive to secret key. The secret key generate by chaotic key generator for modified AES algorithm is $2^{128}, 2^{192}$ and 2^{256}. The key space is large enough to reduce the risk of brute-force attack. The key space is sufficient for practical use for real-time applications of satellite imagery.

4.2   Key Sensitivity Analysis

The encryption algorithm should be sensitive to the secret key and original satellite image to resist against brute-force attacks. High level of key sensitivity is required in cryptosystem. Therefore, slight change (of a single bit) in secret key or original image should produce completely different and unpredictable results. The experiment results of the modified AES algorithm suggest that algorithm is very sensitive to the secret key and input satellite image because of chaotic key generator. For the proposed algorithm, key sensitivity analysis have been carefully performed and completely carried out, with results summarized as follows:

- An original image in Figure 5(a) is encrypted by using the secret key "12345678901234567890123456789012" (in ASCII) and the resultant image is referred as encrypted image A as shown in Figure 5(b).
- The same original image is encrypted by making the slight modification in the secret key i.e. "22345678901234567890 0123456789012" (in ASCII) (the most significant bit is changed in the secret key) and the resultant image is referred as encrypted image B as shown in Figure 5(c).
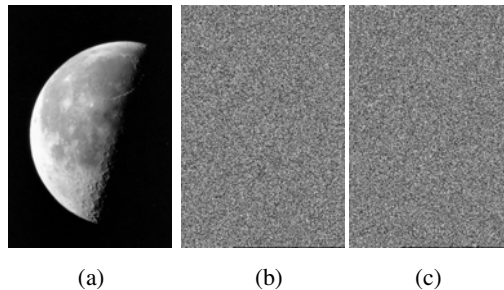

<div align="center">(a)   (b)   (c)</div>

Fig. 5. Key sensitive test results of Proposed Algorithm (a) Original Image (b) Encrypted image with key "12345678901234567890123456789012" (c) Encrypted image with key "22345678901234567890123456789012"

Key sensitivity analysis shows that changing one bit in secret key will result completely different output image. It is clear from images that the decryption process fails to recover the original image completely because of slight change in secret key. Therefore, proposed algorithm is highly key sensitive.

4.3 Performance Analysis

The Boston satellite image (of size 1000 x 1000, gray-scale (0-255) as depicted in Figure 4 has been used in the simulations. The Modified AES Algorithm is tested and evaluated under the three operational modes ECB, CBC and CFB with 256 bits key and block size on a 1.8GHz Pentium IV computer with 1.46 RAM. The algorithm is very fast as compare to other traditional data encryption algorithms. The experimental results of encryption and decryption are given in Table 4.

Table 4. Comparative speed test of experiment results on Boston satellite image. (Time in seconds).

	Simple DES	Triple DES	Modified AES
Avg. time taken for Encryption	5.42	5.76	0.3
Avg. time taken for Decryption	5.47	5.73	0.29
Max. time taken for Encryption	5.53	6.14	0.53
Max. time taken for Decryption	5.61	6.10	0.42
Min. time taken for Encryption	5.31	5.52	0.22
Min.time taken for Decryption	5.39	5.50	0.22

5 Conclusion

We have presented a modified version of AES Algorithm by introducing a new concept of chaotic key generator, to protect the high resolution and multi-spectral Satellite Imagery from unauthorized and illegal use/access/intrusion. The chaotic key generator utilizes six chaotic maps named as Logistic, Tent, Henon, Sine, Cubic and Chebyshev. The idea of using multiple chaotic maps in key generation process is to achieve high level of confusion and diffusion properties and key sensitivity in encryption process. The variable length, large size and complex secret key increases the security level of AES and enhances the resistance level from attackers. Detailed analysis and experimental results have shown that the proposed algorithm offers high level of security and performance. The proposed algorithm is useful for real-time applications of Satellite Imagery, in order to keep the storage and transmission process secure and reliable. Our future work will focus on performing the detailed security and performance analysis of the presented algorithm.

Acknowledgment

References

1. Ray, A.W.: Remote Sensing and Transportation Security. Pecora 15/Land Satellite Information IV/ISPRS Commission I/FIEOS Conference Proceedings (2002)
2. Schneier, B.: Applied Cryptography: Protocols, Algorithms and Source Code in C. Wiley and Sons, Chichester (1996)
3. Menezes, A.J., Oorschot, P.C.V., Vanstone, S.A.: Handbook of applied cryptography. CRC Press, Boca Raton (1997)
4. Stallings, W.: Cryptography and Network Security: Principles and Practice. Prentice-Hall, Upper Saddle River (1999)
5. Stinson, D.R.: Cryptography: Theory and Practice, 2nd edn. Chapman & Hall/CRC, Boca Raton (2002)
6. Gaj, K., Chodowiec, P.: Fast Implementation and Fair Comparison of the Final Candidates for Advanced Encryption Standard Using Field Programmable Gate Arrays. CT-RSA, 84–99 (2001)
7. Hodjat, A., Verbauwhede, I.: A 21.54 Gbits/s Fully Pipelined AES Processor on FPGA. In: Proceedings of the 12th Annual IEEE Symposium on Field-Programmable Custom Computing Machines (FCCM), pp. 308–309 (2004)
8. Janvinen, K., Tominisko, M., Skytta, J.: A Fully Pipelined Memorlyess 17, 8 Gpbs AES-128 encryptor. In: International Symposium of Field programmable Gate arrays, pp. 207–215 (2003)
9. Mclone, M., McCanny, J.V.: Rijindael FPGA Implementations Utilizing Look-Up Tables. Journal VLSI signal process, system 34, 261–275 (2003)
10. Federal Information Processing Standards Publication, FIPS PUB 197: Advanced Encryption Standard (AES) NIST (2001)
11. Bourbakis, N., Dollas, A.: Scan-Based Compression-Encryption Hiding for Video on Demand. IEEE Multimedia Mag. 10, 79–87 (2003)
12. Shiguo, L., Jinsheny, S., Zhiquan, W.: A Block Cipher Based a Suitable of the Chaotic Standard Map. Chaos, Solutions and Fractals 26, 117–129 (2005)
13. Shujun, L., Xuan, Z., Xuanqin, M., Yuanlong, C.: Chaotic encryption scheme for real time digital video. In: Real-Time Imaging, SPIE, vol. 4666, pp. 149–160 (2002)
14. Usama, M., Khan, M.K.: Classical and Chaotic Encryption Techniques for the Security of Satellite Images. In: International Symposium on Biometrics and Security Technologies, pp. 1–6 (2008)
15. Jiang, Z.P.: A Note on Chaotic Secure Communication Systems. IEEE Transactions on Circuits and Systems-I: Fundamental Theory and Applications 49, 92–96 (2002)
16. Mohammad, Z.H., Parvez, M.S.: A Cost Effective Symmetric Key Crypto-graphic Algorithm for Small Amount of Data. In: Proceedings of the 9th IEEE International Multi-topic Conference, pp. 1–6 (2005)
17. Behnia, S., Akhshani, A., Mahmodi, H., Akhavan, A., El Naschie, M.S.: A novel algorithm for image encryption based on mixture of chaotic maps. Chaos, Solutions and Fractals 35, 408–419 (2008)
18. Fridrich, J.: Image Encryption based on Chaotic Maps. In: IEEE International Conference on Systems, Man, and Cybernetics, Computational Cybernetics and Simulations, vol. 2, pp. 1105–1110 (1997)
19. Fridrich, J.: Symmetric Ciphers based on Two-Dimensional Chaotic Map. Int. J. Bifurcation and Chaos 8, 1259–1284 (1998)
20. Li, S., Mou, X., Cai, Y.: Improving Security of a Chaotic Encryption Approach. Physics Letters A 290, 127–133 (2001)

21. Baptista, M.S.: Cryptography with Chaos. Physics Letters A 240, 50–54 (1998)
22. Kotulski, Z., Szczepanski, J.: Discrete Chaotic Cryptography (DCC): New Method for Secure Communication. In: Proc. NEEDS 1997 (1997)
23. Matthews, R.: The Derivation of a Chaotic Encryption Algorithm. Cryptologia XII, 29–42 (1989)
24. Habutsu, T., Nishio, Y., Sasase, I., Mori, S.: A secret key cryptosystem by iterating a chaotic map. In: Davies, D.W. (ed.) EUROCRYPT 1991. LNCS, vol. 547, pp. 127–140. Springer, Heidelberg (1991)
25. Kotulski, Z., Szczepanski, J., Gorski, K., Paszkiewicz, A., Zugaj, A.: Application of Discrete Chaotic Dynamical Systems in Cryptography–DCC Method. Int. J. Bifurcat. Chaos 9, 1121–1135 (1999)
26. Alvarez, E., Fernandez, A., Garcia, P., Jimenez, J., Marcano, A.: New Approach to Chaotic Encryption Physics Letters A 263, 373–375 (1999)
27. Wong, W.K., Lee, L.P., Wong, K.W.: A modified chaotic cryptographic method. Computer Physics Communication 138, 234–236 (2000)
28. Wong, K.W.: A Fast Chaotic Cryptography Scheme with Dynamic Look-up Table. Physics Letters A 298, 238–242 (2002)
29. Wong, K.W., Ho, S.W., Yung, C.K.: A Chaotic Cryptography Scheme for Generating Short Ciphertext Physics Letters A 310, 67–73 (2003)
30. Pareek, N.K., Patidar, V., Sud, K.K.: Discrete Chaotic Cryptography Using External Key. Physics Letters A 309, 75–82 (2003)
31. Shannon, C.E.: A Mathematical Theory of Communication. Bell System Tech. J. 27, 379–423 (1948)
32. Xiang, T., Liao, X.F., Tang, G.P., Chen, Y., Wong, K.W.: A Novel Block Cryptosystem Based on Iterating a Chaotic Map. Physics Letters A 349, 109–115 (2006)

Group-Based Proxy Re-encryption Scheme

Chunbo Ma and Jun Ao

School of Information and Communication,
Guilin University of Electronic Technology,
Guilin, Guangxi, 541004, P.R. China

Abstract. Recently, proxy re-encryption scheme received much attention. A proxy re-encryption used for divert ciphertext from one group to another without revealing underlying plaintext is proposed in this paper. The scheme is bidirectional and any member can independently decrypt the ciphertext encrypted to its group. The security of the proposed scheme is discussed and proofs are given to show that the scheme withstands chosen ciphertext attack in standard model.

Keywords: group-based; proxy; re-encryption; standard model; V-DDH assumption.

1 Introduction

Mambo and Okamoto introduced the technique for delegating decryption right in [1]. Later, Blaze et al. [3] presented the notion of "atomic proxy cryptography" in 1998. In a proxy re-encryption scheme, proxy is allowed to transform a ciphertext corresponding to Alice's public key into one that can be decrypted by Bob's private key. The proxy in this scheme can't obtain any information about the plaintext or the private key used to decrypt the ciphertext. Generally speaking, proxy re-encryption scheme can be divided into two categories by proxy functions, namely bidirectional and unidirectional [2]. In a bidirectional scheme, the proxy secret key can be used to divert ciphertexts from Alice to Bob and vice versa. Obviously, a mutual trust relationship between Alice and Bob is needed, otherwise, some security problem will arise [4]. In a unidirectional scheme, the proxy secret key is allowed to be used to divert ciphertexts from Alice to Bob, and from Bob to Alice is not permitted.

The proxy re-encryption scheme has many applications. For example, in traditional storage system [12][13], the Server who housing information sometimes just semi-trusted and some added means should be used to ensure its security. In 2005, Ateniese et al. [4] designed an efficient and secure distributed storage system in which the proxy re-encryption scheme is employed. There are some other applications, such as secure email forwarding, and so on [3][6].

Group communication is a useful primitive for sharing message in a specifically group and has been widely used in unbalanced networks, for example, clusters of mobile devices [17]. Ma et al. [5] designed an encryption scheme to ensure the privacy of the messages shared in the group. In the scheme, anyone can encrypt a

message and distribute it to a designated group and any member in the designated group can decrypt the ciphertext. There exists proxy re-encrypted problem in two different groups. For example, due to the change of duty, some work managed by group A has been assigned to group B such that some encrypted documents sent to group A should be decrypted by group B. In such scenario, proxy re-encryption technique can be used to realize this transformation.

To date, most of the researches of proxy re-encryption emphasize two entities communication. For example, a proxy transforms a ciphertext computed under Alice's public key into one that can be opened by Bob's secret key. However, few literatures present approach to deal with proxy re-encryption for group communication. Motivated by above mentioned, we present a group-based proxy re-encryption scheme for group communications in this paper. It is a bidirectional scheme, i.e. the proxy using one secret key can divert ciphertext from group A to group B and vice versa. Moreover, since the secret value of the group public key can't be deduced from the re-encryption key and the member's private keys, the scheme is secure against collude attack.

The rest of paper consists of following sections. In section 2, we introduce some related works. In section 3, we give the security model and complexity assumptions. The proposed group-based proxy re-encryption scheme is presented in section 4. In section 5, we discuss the security of the proposed scheme in standard model. Finally, we draw the conclusions in section 6.

2   Related Works

The notion of "atomic proxy cryptography" was presented by Blaze et al. [3] in 1998. It provides securer and more efficient way than usual to deal with the scenario in which a proxy decrypts a ciphertext using Alice's private key and then encrypts the result using Bob's public key. They depict two examples: one for encryption, and another for signature. However, the two examples presented in this paper were proved to have low security guarantees. Their approach is only useful when the trust relationship between Alice and Bob is mutual. In addition, it is not suitable for group communication since the proxy has to preserve n re-encryption key for n group members.

In 2003, Ivan and Dodis [2] designed proxy encryption for ElGamal, RSA, and an IBE scheme using secret sharing technique. In their ElGamal based scheme, Public Key Generator (PKG) generates encrypt key EK and decrypt key DK for each user, and then DK is divided into two parts x_1 and x_2, which satisfy DK$= x_1 + x_2$. Moreover, they designed unidirectional and bidirectional proxy encryption scheme. These "secret-sharing" schemes don't change ciphertexts for Alice into ciphertext for Bob in the purest sense, the delegate decryption by requiring Bob to store additional secret that maybe difficult for him to manage.

Following the work of Ivan and Dodis, Ateniese et al. [4] presented an improved proxy re-encryption scheme, and employed it in distributed storage system. In their re-encryption scheme, the proxy only preserves a discrete value to prevent the collude attack. The advantage of the method presented in [2] is that it is feasible to design a unidirectional proxy encryption. Whereas it is very difficult to extend the scheme to

group communication since overload stems from the secret sharing technology. Thus why the scheme proposed in [4] is not very practical.

Canetti and Hohenberger [6] proposed a proxy re-encryption scheme secure against chosen ciphertext attack. They discuss its security in standard model. In their paper the bilinear pairing technology is used to design proxy re-encryption scheme. Although their approach is just suitable for two entities, some method can be used to design group communication.

Libert and Vergnaud [18] proposed a proxy re-encryption scheme which comes from Canetti-Halevi-Katz's [20] scheme and can be seen as a natural extension of the Canetti-Hohenberger definition to the unidirectional case. Their scheme is unidirectional, i.e. only allows the proxy to divert ciphertexts form Alice to Bob. However, some messages on Alice such as public key have been preserved in the ciphertext generated in the phase of ReEnc. An attacker maybe uses these messages to recognize the original target of the ciphertext. furthermore, the scheme may be menaced by malleability.

There are some other re-encryption schemes, such as Jakobsson's quorum controlled asymmetric proxy re-encryption [7], and the identity-based scheme presented by Green and Ateniese [8]. There are some investigations on proxy signature schemes [9][10]. However, we have not seen any group-based proxy re-encryption scheme.

3 Background

3.1 Preliminaries

Let G_1 be a cyclic multiplicative group generated by g, whose order is a prime q and G_2 be a cyclic multiplicative group of the same order q. Assume that the discrete logarithm in both G_1 and G_2 is intractable. A bilinear pairing is a map $e : G_1 \times G_1 \rightarrow G_2$ and satisfies the following properties:

- Bilinear: $e(g^a, p^b) = e(g, p)^{ab}$. For all g, $p \in G_1$ and $a,b \in \mathbb{Z}_q^*$, the equation holds.
- Non-degenerate: There exists $p \in G_1$, if $e(g, p) = 1$, then $g = O$.
- Computable: For g, $p \in G_1$, there is an efficient algorithm to compute $e(g, p)$.

Typically, the map e will be derived from either the Weil or Tate pairing on an elliptic curve over a finite field. Pairings and other parameters should be selected in proactive for efficiency and security [11].

3.2 Complexity Assumptions

- Computational Diffie-Hellman Assumption

Given g^a and g^b for some $a,b \in \mathbb{Z}_q^*$, compute $g^{ab} \in G_1$. A (τ, ε)-CDH attacker in G_1 is a probabilistic machine Ω running in time τ such that

$$Succ_{G_1}^{cdh}(\Omega) = \Pr[\Omega(g, g^a, g^b) = g^{ab}] \geq \varepsilon$$

where the probability is taken over the random values a and b. The CDH problem is (τ, ε)-intractable if there is no (τ, ε)-attacker in G_1. The CDH assumption states that it is the case for all polynomial τ and any non-negligible ε.

- Decisional Diffie-Hellman Assumption [14]

We say that an algorithm π that outputs $b \in \{0,1\}$ has advantage ε in solving the **Decisional Diffie-Hellman** (**DDH**) problem in G_1 if

$$|\Pr[\pi(g, g^a, g^b, g^c, e(g,g)^{abc}) = 0] - \Pr[\pi(g, g^a, g^b, g^c, T) = 0]| \geq \varepsilon$$

where the probability is over the random bit of π, the random choice of $a, b, c \in \mathbb{Z}_q^*$, and the random choice of $T \in G_2$. The **DDH** problem is intractable if there is no attacker in G_1 can solve the **DDH** with non-negligible ε.

- V-Decisional Diffie-Hellman Assumption

An algorithm π that outputs $b \in \{0,1\}$ has advantage ε in solving the **V-Decisional Diffie-Hellman** (**V-DDH**) problem in G_1 if

$$|\Pr[\pi(g, g^a, g^{ab}, g^{ac}, g^{bc}) = 0] - \Pr[\pi(g, g^a, g^{ab}, g^{ac}, T) = 0]| \geq \varepsilon$$

where the probability is over the random bit of π, the random choice of $a, b, c \in \mathbb{Z}_q^*$, and the random choice of $T \in G_1$. The **V-DDH** problem is intractable if there is no attacker in G_1 can solve the **V-DDH** with non-negligible ε.

3.3   Security Notions

The proposed re-encryption scheme consists of five algorithms, namely **KeyGen**, **ReKeyGen**, **Enc**, **ReEnc** and **Dec**.

- **KeyGen** (1^λ). On input the security parameter, outputs the public key P_{pub} of each group and the corresponding private key s_{ki} for each member.
- **ReKeyGen** (s_{k1}, s_{k2}). On input two private key s_{k1} and s_{k2}, outputs a bidirectional re-encryption key $r_{k(1 \leftrightarrow 2)}$.
- **Enc** (P_{pub}, m). On input message $m \in \{0,1\}^*$ and a public key P_{pub}, outputs a ciphertext C.
- **ReEnc** $(r_{k(1 \leftrightarrow 2)}, C_1)$. On input ciphertext C_1 and the re-encryption key $r_{k(1 \leftrightarrow 2)}$, outputs a ciphertext C_2 or an error symbol $\perp$.
- **Dec** (s_k, C). On input ciphertext C and a private key s_k, outputs the corresponding message m.

The indistinguishable chosen ciphertext attack (IND-CCA) [15] presented by Goldwasser and Micali has been widely used to analyze the security of an encryption scheme. In this model, several queries are available to the attacker to model his capability. Subsequently, Rackhoff and Simon [16] enhanced it and proposed adaptively

chosen ciphertext attack (IND-CCA2). Since this notion is stronger, it is becoming a prevalent model in analyzing encryption scheme. Green and Ateniese [8] enhanced the model and used it to discuss the security of proxy re-encryption scheme, then followed by Canetti and Hohenberger [6].

In this part, we define adaptively chosen ciphertext security of the group-based proxy re-encryption scheme. Compared to the model mentioned in [6], we don't consider the case of group A or B's corruption due to the properties of our key generation. Security is defined using the following game between an Attacker and Challenger.

Setup. The Challenger initializes the system and gives the Attacker the resulting system parameters and the public key P_{pub}. It keeps private key to itself.

Query phase 1

a) **Decrypt queries.** The Attacker issues a query (c_{i1}, c_{i2}, c_{i3}). The Challenger outputs **Decrypt** (c_{i1}, c_{i2}, c_{i3}), otherwise outputs error symbol $\perp$.

b) **Re-encrypt queries.** The Attacker issues a query (c_{i1}, c_{i2}, c_{i3}) encrypted using the public key of group A. The Challenger outputs **Re-encrypt** $(r_{k(A \leftrightarrow B)}, c_{i1}, c_{i2}, c_{i3})$. Obviously, the output is a ciphertext encrypted using the public key of group B.

The Attacker is allowed to perform the **Query phase 1** several times.

Challenge. Once the Attacker decides that Query phase 1 is over, the Attacker outputs two equal length messages $\{M_0, M_1\}$ to the Challenger. Upon receiving the messages, the Challenger chooses a random bit $e \in \{0,1\}$, invokes Encrypt (P_A, M_e) and outputs (c_1^*, c_2^*, c_3^*) as the answer.

Query phase 2. The Attacker continues to adaptively issue Decrypt queries and Re-encrypt queries. The Challenger responds as in the phase 1. These queries may be asked adaptively as in Query phase 1, but the query on (c_1, c_2, c_3) is not permitted.

Guess. Finally, the Attacker outputs a guess $e' \in \{0,1\}$ for e and wins the game if $e' = e$.

The encryption scheme is secure against chosen ciphertext attack, if the Attacker has a negligible advantage $\varepsilon = \left| \Pr(e = e') - \dfrac{1}{2} \right|$ to win the game.

4 The Proposed Bidirectional Proxy Re-encryption Scheme

We assume that there exist two groups in our scheme, namely A and B. The function of the Proxy is to transform ciphertext corresponding to the public key of group A into ciphertext for the public key of group B without revealing any information about the secret decryption keys or the clear text, and vice versa. It means that our proxy re-encryption is a bidirectional scheme. The proposed scheme consists of following steps.

4.1 Initialize

Let G_1 be a cyclic multiplicative group generated by g, whose order is a prime q and G_2 be a cyclic multiplicative group of the same order q. A bilinear pairing is a map: $e : G_2 \times G_1 \rightarrow G_2$ that can be efficiently computed.

PKG chooses $a, b \in \mathbb{Z}_q^*$ and $h \in G_1$ uniformly at random, and then computes $g_1 = g^a$ and $g_2 = g^b$. The master private keys are a and b, and the master public keys are g_1, g_2 and h.

4.2 Key Generation

PKG chooses $k \in \mathbb{Z}_q^*$ uniformly at random as the tag of the group A. Using $P_A = g^k$ as group A's public key. The private key of the member $p_i \in A$ can be generated as follows:

PKG chooses $r_i \in \mathbb{Z}_q^*$ uniformly at random.

compute and output $d_{i1} = h^{r_i} g^{r_i}$, $d_{i2} = h^{(r_i - ak) \cdot b^{-1}} g^{r_i b^{-1}}$, and $d_{i3} = g^{ak} h^{r_i}$.

The member p_i's private key is $d_i = \{d_{i1}, d_{i2}, d_{i3}\}$.

PKG chooses $l \in \mathbb{Z}_q^*$ uniformly at random as the tag of the group B. Using $P_B = g^l$ as group B's public key. The member's private key can be generated as $p_i \in A$.

4.3 Encrypt

In order to encrypt a message $M \in \{0,1\}^l$ for the group A, the sender (S_{Enc}) first chooses $s \in \mathbb{Z}_q^*$ uniformly at random, and computes the ciphertext

$$c_1 = e(g_1, P_A)^s \cdot M \quad c_2 = (hg)^s \quad c_3 = g_2^s.$$

The ciphertext for message M is $c = (c_1, c_2, c_3)$. The sender S_{Enc} sends the ciphertext to all the members in the group A by broadcast over Internet.

4.4 Re-encrypt

In order to transform the ciphertext to group B whose public key is $P_B = g^l$, PKG generates a Re-encrypt key $r_{k(A \leftrightarrow B)} = (l - k) \cdot b^{-1} a$ and sends it to proxy. Then using the Re-encrypt key, the proxy can perform

$$\tilde{c}_1 = e(g_1, P_A)^s \cdot M \cdot e(c_3, g^{r_{k(A \leftrightarrow B)}})$$

$$= e(g, g)^{ask + sb(l-k)b^{-1}a} \cdot M$$

$$= e(g, g)^{asl} \cdot M$$

$$\tilde{c}_2 = c_2, \quad \tilde{c}_3 = c_3.$$

The Re-encrypted ciphertext is $(\tilde{c}_1, \tilde{c}_2, \tilde{c}_3)$.

4.5 Decrypt

After receiving the re-encrypted message $c = (\tilde{c}_1, \tilde{c}_2, \tilde{c}_3)$, the member $p_i \in B$ can decrypt the ciphertext as follows:

1. compute $T = e(\tilde{c}_2, d_{i3})e(\tilde{c}_3, d_{i2}) / e(\tilde{c}_2, d_{i1})$.
2. compute $M = \tilde{c}_1 / T$.

Any member $p_i \in B$ can compute T correctly, since

$$
\begin{aligned}
T &= \frac{e(\tilde{c}_2, d_{i3})e(\tilde{c}_3, d_{i2})}{e(\tilde{c}_2, d_{i1})} \\
&= \frac{e(g^s h^s, h^{r_i} g^{al})e(g_2^s, h^{-ab^{-1}l} h^{r_i b^{-1}} g^{r_i b^{-1}})}{e(g^s h^s, g^{r_i} g^{r_i})} \\
&= \frac{e(h^s, h^{r_i})e(h^s, g^{al})e(g^s, h^{r_i})e(g^s, g^{al})}{e(h^s, h^{r_i})e(h^s, g^{r_i})} \cdots \frac{e(g_2^s, h^{-ab^{-1}l})e(g_2^s, h^{r_i b^{-1}})e(g_2^s, g^{r_i b^{-1}})}{e(g^s, h^{r_i})e(g^s, g^{r_i})} \\
&= e(g^s, g^{al}) = e(g, g)^{als}
\end{aligned}
$$

So the member p_i can get the plaintext

$$
M = \tilde{c}_1 / T
$$

To the user in group A, he can get the plaintext M from (c_1, c_2, c_3) similarly to the user in group B.

5 Security

In this section, we will discuss the security of the proposed proxy re-encryption scheme in standard model. The measure used to prove our scheme comes from the paper [6].

Lemma 1. Suppose the **CDH** assumption holds. Then given $g^a, g^{ab}, g^{ac} \in G_1$, computing g^{bc} is intractable.

Proof. Assume that given $g^a, g^{ab}, g^{ac} \in G_1$, the attack Alice has ability to compute another g^{bc} . Then we can design an algorithm to solve CDH problem. In other words, given $g^m, g^n \in G_1$, the challenger Bob can compute $g^{m \cdot n}$ by running Alice as a subroutine.

To the given $g^m, g^n \in G_1$, Bob chooses a random number $t \in \mathbb{Z}_q^*$, computes g^{mt} and g^{nt} , and then sends g^t , g^{mt} and g^{nt} to Alice. With the assumption, Alice can output $g^{m \cdot n}$, then Bob can solve CDH problem.   □

Theorem 1. Suppose that the **V-DDH** is intractable. Then our proxy re-encryption scheme is secure against adaptively chosen ciphertext attack.

Proof. Assume that if the attacker Alice has ability to break the proposed encryption scheme via chosen ciphertext attack with non-negligible probability ε, then we can prove that there exists challenger Bob that can solve **V-DDH** problems with the same probability. In other words, given g^{a^*}, $g^{a^*s^*}$, $g^{a^*k^*} \in G_1$ and $T \in G_1$, Bob can decide if T is equal to $g^{s^*k^*}$ with non-negligible probability by running Alice as a subroutine. The challenger Bob interacts with Alice by simulating **Decrypt**, **Re-encrypt** oracles.

Bob initializes the system, chooses random numbers $w, v \in \mathbb{Z}_q^*$. Let

$$g_1 = g^{a^*} \qquad g_2 = g^{a^* \cdot k^* \cdot w} \qquad P_A = g^{a^* k^*} \qquad h = g^{a^* \cdot k^* \cdot v - 1}.$$

Then Bob chooses a random number $\alpha \in \mathbb{Z}_q^*$ and publishes $P_A = g^{a^* k^*}$ and $P_B = g^{a^* k^* \alpha}$.

Query phase 1

- **Decrypt queries.** To every new query (c_1, c_2, c_3), Bob computes and outputs $M = c_1 / e(g_1, c_3^{1/w})$ as the answer.
- **Re-encrypt queries.** To every new query (c_1, c_2, c_3), Bob computes

$$\tilde{c}_1 = e(g_1, P_A)^s \cdot M \cdot e(c_3^{1/w}, g^{a^* \alpha - a^*})$$
$$= e(g, g)^{(a^*)^2 k^* s + s(a^*)^2 k^* (\alpha - 1)} \cdot M$$
$$= e(g, g)^{(a^*)^2 k^* s \alpha} \cdot M$$

and sets $\tilde{c}_2 = c_2$ and $\tilde{c}_3 = c_3$, and then outputs $(\tilde{c}_1, \tilde{c}_2, \tilde{c}_3)$ as the answer.

Since $w, \alpha \in \mathbb{Z}_q^*$ are two random number, Alice can't distinguish the simulated answers from the actual results. Thereby, we say above simulation is perfect. Alice is allowed to perform **Decrypt** and **Re-encrypt** queries several times.

Challenge phase. When Alice decides Query phase 1 is over, she chooses two equal length messages M_1, M_0, and sends them to Bob. Bob chooses a random bit $e \in \{0,1\}$, computes and outputs

$$c_1^* = e(g_1, T) \cdot M_e = e(g^{a^*}, g^{a^* \cdot k^*})^{s^*/a^*} \cdot M_e$$
$$c_2^* = (T)^v = (g^{k^* s^*})^v = (g \cdot g^{a^* \cdot k^* \cdot v - 1})^{s^*/a^*}$$
$$c_3^* = (T)^w = (g^{k^* s^*})^w = (g^{a^* k^* w})^{s^*/a^*}$$

as the answer. The **Challenge phase** can be performed only once.

Query phase 2. Alice continues to adaptively issue **Decrypt** and **Re-encrypt** queries. Bob responds as in the phase 1. However, the query on (c_1^*, c_2^*, c_3^*) is not permitted.

Guess. Finally, Alice outputs a guess $e^{'} \in \{0,1\}$ for e. If $e^{'} = e$, then Bob decides $T = g^{s^{*}k^{*}}$, otherwise Bob decides $T \neq g^{s^{*}k^{*}}$.

Obviously, above simulation is perfect. We say that Alice can break the proxy re-encryption scheme with non-negligible probability ε. It means that Alice can output correct $e^{'}$ with probability ε. Then Bob can solve the **V-DDH** with same probability ε by running Alice as a subroutine.   □

6 Conclusions

Recently, most researchers focused their attention on how to convert ciphertext for one user into ciphertext for another without revealing underlying plaintext. According to the proxy function, we can divide these schemes into two categories: bidirectional and unidirectional. In this paper, we extend this notion and present bidirectional proxy re-encryption scheme used for group communications. In our scheme, the proxy diverts the ciphertext for group A into ciphertext for group B, and vice versa. To the member in group A/B, he can independently decrypt the ciphertext for the group. Obviously, the performance of encryption in our proposed scheme is similarly to that of paper [6], and it is crucial to the group communication since lots of members are involved in. Decryption operation is independently completed by each group member.

References

1. Mambo, M., Okamoto, E.: Proxy Cryptosystems: Delegation of the Power to Decrypt Ciphertexts. IEICE Trans. Fund. Electronics Communications and Computer Science E80-A/1, 54–63 (1997)
2. Ivan, A., Dodis, Y.: Proxy Cryptography Revisited. In: Proceedings of the Tenth Network and Distributed System Security Symposium (February 2003)
3. Blaze, M., Bleumer, G., Strauss, M.J.: Divertible protocols and atomic proxy cryptography. In: Nyberg, K. (ed.) EUROCRYPT 1998. LNCS, vol. 1403, pp. 127–144. Springer, Heidelberg (1998); Ateniese, G., Fu, K., Green, M., Honhenberger, S.: Improved Proxy Re-encryption Schemes with Applications to Secure Distributed Storage. In: Proceedings of NDSS, pp. 29-43 (2005)
4. Ma, C., Mei, Q., Li, J.: Broadcast Group-oriented Encryption for Group Communication. Journal of Computational Information Systems 3(1), 63–71 (2007)
5. Canetti, R., Hohenberger, S.: Chosen-Ciphertext Secure Proxy Re-Encryption (2007), http://eprint.iacr.org/2007/171
6. Jakobsson, M.: On Quorum Controlled Asymmetric Proxy Re-encryption. In: Proceedings of Public Key Cryptography, pp. 112–121 (1999)

7. Green, M., Ateniese, G.: Identity-based proxy re-encryption. In: Katz, J., Yung, M. (eds.) ACNS 2007. LNCS, vol. 4521, pp. 288–306. Springer, Heidelberg (2007)
8. Kim, H., Baek, J., Lee, B., Kim, K.: Computing with Secrets for Mobile Agent Using One-time Proxy Signature. In: Proceedings of SCIS 2001, vol. 2(2), pp. 845–850 (2001)
9. MacKenzie, P.D., Reiter, M.K.: Two-party generation of DSA signatures. In: Kilian, J. (ed.) CRYPTO 2001. LNCS, vol. 2139, pp. 137–154. Springer, Heidelberg (2001)
10. Boneh, D., Lynn, B., Shacham, H.: Short signatures from the weil pairing. In: Boyd, C. (ed.) ASIACRYPT 2001. LNCS, vol. 2248, pp. 514–532. Springer, Heidelberg (2001)
11. Blaze. M.: A Cryptographic File System for Unix. In: First ACM Conference on Communications and Computing Security, Fairfax, VA (November 1993)
12. Freeman, W., Miller, E.: Design for a Decentralized Security System for Network-attached Storage. In: Proceedings of the 17th IEEE Symposium on Mass Storage Systems and Technologies, College Park, MD, March 2000, pp. 361–373 (2000)
13. Boneh, D., Boyen, X.: Efficient selective-ID secure identity-based encryption without random oracles. In: Cachin, C., Camenisch, J.L. (eds.) EUROCRYPT 2004. LNCS, vol. 3027, pp. 223–238. Springer, Heidelberg (2004)
14. Goldwasser, S., Micali, S.: Probabilistic Encryption. Journal of Computer and System Sciences 28, 270–299 (1984)
15. Rackoff, C., Simon, D.R.: Non-interactive zero-knowledge proof of knowledge and chosen ciphertext attack. In: Feigenbaum, J. (ed.) CRYPTO 1991. LNCS, vol. 576, pp. 433–444. Springer, Heidelberg (1992)
16. Phan, T., Huan, L., Dulan, C.: Challenge: Integrating Mobile Wireless Devices into the Computational Grid. In: Proceedings of MobiCom, pp. 271–278 (2002)
17. Libert, B., Vergnaud, D.: Unidirectional chosen-ciphertext secure proxy re-encryption. In: Cramer, R. (ed.) PKC 2008. LNCS, vol. 4939, pp. 360–379. Springer, Heidelberg (2008)

A Method for Modeling Gene Regulatory Network with Personal Computer Cluster

Jinlian Wang, Jian Zhang, and Lin Li[*,**]

School of Biomedical Engineering Capital Medical University, Beijing, China, 100069
lil@ccmu.edu.cn

Abstract. It is one of the serious challenges in bioinformatics that the algorithms are used to study cell's molecular pathways and to simulate cell's molecular network so as to reveal tumor's molecular characteristics at a molecular level. In this paper we aim at disclosing gene nonlinear interactions under parallel computing environment. First based on graph-coloring scheme, determine the types of the higher order logical relationship among multi-genes to get their expression pattern. Secondly gets the sample supporting degree for the logical expression patterns. Thirdly take the supporting degree for the weight of the regulatory network model to show the probability with which the logical relation happens among these samples, and further build a weighted and directed regulatory network of gene expression. Finally apply this method to the colon cancer mRNA micro-array dataset to build a higher order logical regulatory network and to visualize the tumors' molecular signal pathways. Results show that with this way we can not only extract multi-gene's nonlinear logical relations hidden in the gene expression profile but also analyze effectively tumor cell's signal pathways and build a regulatory network of gene expression, which provides a tool for the study on tumor-gene's molecular bioinformatics.

Keywords: gene regulatory network, gene expression profile, computer cluster, graph coloring, colon cancer.

1 Introduction

Microarray technology has been providing an effective platform for a more comprehensive and systematic study on tumor characteristics at gene expression level, which brought a lot of documents about the analysis to tumor's gene expression profile [1,2]. In the past the analysis to tumor's gene expression profile was focused on the discovery of tumor's gene relation pattern, oncogene screening, drug sensibility, classification of tumor molecule, the molecular characteristics in early diagnoses[3], etc, and there was less in a large-scale study on the regulatory network of gene expression with analog-digital computation. In recent years, the research institutes both China

---

and aboard have achieved something in this field and proposed some representative methods, such as, in the study on the network models of gene's regulatory mechanism there were weighted matrix[4], Boolean Network[5], linear combination[6], model of difference equations[7], etc. and in the models studying gene expression based on gene expression profile, there were Logistic Regression, mutual information, Petri Net, model of probability graph, etc. There was also the research on the regulatory network of gene expression made by Young Laboratory at Massachusetts Institute of Technology. A lot of work done with these methods focused on the study on the modeling of regulatory or correlative network. In 2004, Bowers et al[8] initially made a way to analyze the high-order logical relation of ternary protein with LAPP (Logic Analysis of Phylogenetic Profile), and in 2005 they applied LAPP again to the analysis to the data of neurglioma gene expression profile and got the logical pathways regulating neurglioma's oncogene[9]. At same year Zhangxin et al [10] promoted LAPP to the logical relation of quaternary protein. What they did provided an idea to study multi-factor's regulatory pattern, that is, the regulatory pattern of logical expression could reflect synergistic genome's main functional pattern.

Parallel computing has been using widely in bioinformatics research area recently. Lots of bioinformatics tools have been developed running on the parallel computing platform, such as BLASTA[11], FASTA[12], HMMER[13] and so on, in particular, genome sequencing rely on parallel computing techniques. In our previous modeling higher order logical regulatory network [14], we found that the complexity of computing increased with the number of genes. To decrease the time complexity of the higher order logical genes relationships, we used parallel computing platform under the local network contained 15 personal computers run on Linux operating system. The result shows that the computers cluster greatly improve the computing efficient. We attempt to supply domestic researchers a parallel computing platform with personal computer to solve the computing complexity.

2 Methods

2.1 Terms

Duo to the complicacy of the molecular network inside the cells, the genes act synergistically through the regulatory relation modes, such as parallel, intersection, branching, etc, to complete the specific cell function. According to Boolean definition for the q_{th}-order logical relation, if $q=1$, it is called the first-order logical relation (low order) and shows the logical relevance between two genes. Here it is also called genes' binary logical relation; if $q \geq 2$, it is called high-order logical relation and shows the logical relevance of the genes of three and upwards. Here it is also called gene's ternary or quaternary logical relation. Xin et al [10] defined the prediction function to to measure the predicting or regulating ability of the predicting gene *(m, n, k)* based on the target gene *j*. Y is a probability of the expression of the target gene *j*.

The indeterminate coefficient of the third-order logic is defined as follows:

$$U[d \mid f(a,b,c)] = \frac{\{H(d) + H[f(a,b,c)] - H[d, f(a,b,c)]\}}{H(d)} \tag{1}$$

where $U[d \mid f(a,b,c)]$ ($0 \le U[d \mid f(a,b,c)] \le 1$) is the third-order logical predicting function Y and shows the probability that gene a, b and c predict the expression of gene d through the logical function f(a, b, c). See the appendix for the detailed expressions of the logical function f(a, b, c). H(c) is the independent entropy of gene c, H(f(a, b, c)) is the joint entropy of gene a, b and c and H(d, f(a, b,)) is the joint entropy of gene a, b, c and d.

The third-order logical relation must satisfy:

$$
\begin{cases}
U(d \mid f(a,b,c)) >> U(d \mid f(a,b)) \& U(d \mid f(a,c)) \& U(d \mid f(b,c)) \\
T_3 > T_2 > T_1 \\
U(d \mid f(a,b)) < T_2 \\
U(d \mid f(a,c)) < T_2 \\
U(d \mid f(c,b)) < T_2 \\
U(d \mid f(a,b,c)) < T_3 \\
U(d \mid a) < T_1 \\
U(d \mid b) < T_1 \\
U(d \mid c) < T_1 \\
S > 0.5
\end{cases}
\tag{2}
$$

Where, T_1 , T_2 and T_3 are the threshold of the first-order, the second-order and the third-order logical relation respectively and S is the supporting degree of expression pattern.

2.2 The Higher Order Logical Relationship Identification

The types of the third-order logical relations are determined with the coloring scheme on Venn diagram as shown in Fig. 1. Each coloring scheme corresponds to a logical type. The three circles in the figure represent respectively three genes, A, B and C. The shadow shows the regulatory relation of three genes' logical function combination to the forth gene D. The figures represent the divided areas of the Venn diagram to be colored.
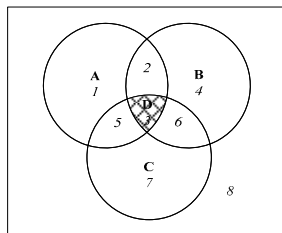
Fig. 1. Principle of Graph Coloring

If the eight areas in Figure 1 are colored, there are 256 different coloring schemes including 218 pure third-order logical relations, 68 structurally equivalent logics and 68 logical relation types.

2.3   Sample's Support Level f and Logical Expression Patterns

The more and more researches show that tumor is a complicated disease resulted in the accumulation caused by multi-gene's interactions and multi-step mutations. The complexity shows the randomness or indeterminacy in the phenotype and genotype. Together with the individual difference of each patient, on gene's expression profile it is shown that gene's expression amount is greatly different. Take the second-order logical relation for an example. When the expression amount of gene a, b changes, gene c expresses among some samples and doesn't express among others. Multi-gene's logical relations can reflect, to a certain extent, the indeterminacy and randomness in gene's expression pattern. By counting the probability with which the expression pattern of gene's logical relation happens among samples, we can find the expression pattern occurring among the majority of samples to deduce the changes in gene network and pattern related to the disease and further to understand the relationship between genome and disease.

Due to sample's individual difference, the combination of multi-gene's logical relation may be different in its expression pattern. On the question whether the expression pattern coming from data analysis truly reflected its relation with tumor's occurrence, we put forward the sample supporting degree of the expression pattern. First analyze the expression pattern with higher order logical relation, and then count the probability with which each expression pattern happens among the samples and take this probability for the sample supporting degree, and finally through disturbing test, regard the expression pattern with significance as gene's logical relation to find out tumor's gene network structure and expression pattern and to reveal the relation between genome and disease.

2.4   The Parallel Computing Modules

The framework of parallel computing contained parallel hardware architecture, software, and parallel algorithm strategies. 1) The distributed parallel architecture is adopted for its advantages of large storage capacity, fast program processing ability in the distributed located computers, small communication cost between nodes and high cache hit ratio. 2) The soft ware environment is composed of Linux operating system, C program language and Message Passing Interface (MPI) which have several version including MPICH,LAM, IBM MPL. Here we use MPICH downloaded free from ftp://ftp.mcs.anl.gov/pub/mpi/mpich.tar.gz. 3) We adopt strategy of divide and rule to design parallel algorithm, which used master-slave parallel programming mode. First the higher order logic computing task is divided into several same sub-tasks executed on all nodes including slave and master nodes, and this division task is accomplished by master node and sub-tasks computing are finished by the slave nodes. After slave nodes finishing computing, the master node collects the results from every slave node. Finally the network model is completed in the master node. We adopted remote shell service to avoid the communication between slave nodes during their executing tasks.

The process of the method contains several modules including system initiated, tasks dispatching, logical relationship computing, result collection, threshold choosing and network construction modules.

3   Application

In order to reveal the mechanism of the colon cancer in the molecular level, we construct higher order logical regulatory gene network of colon cancer on the cluster computing platform employed 15 personal computers configured dual PIII 350MH CPU, 2G main memory, 10M/100M adapted network card.

3.1   Dateset

The sample came from the cancer tissues of 37 colon caner patients in Finland collected by Laiho P and others [15]; In June 2006 Laiho and others openly published their data in GEO Data Bank with the sequence number GSE4045. They used 22283 probes of HG133A chip to check 37 samples' mRNA abundance and got the original laser scanning chip data after they were processed by software Dchip6.0. Here we use 1 to represent Precence_call value in the expression profile, which means the gene expresses in samples, and 0 to represent Absence_call value, which means the gene doesn't express in samples. Then we use this binary code for the inputting value to analyze genes' logical relations[14].

We got 51 colon cancer genes and cancer-suppressing genes recalled by PubMed, which was related to the Dukes period division, the metastasis of colon cancer, vasculariztion, point mutation, DNA mispairing repair and cell's periodically regulatory function module. Each gene's ID, gene name and probe ID in NCBI Data Bank show in Table 1.

Table 1. A list of 51 oncogenes and cancer-suppressing genes

Gene Symbol	Gene Symbol	Gene Symbol	Gene Symbol	Gene Symbol	Gene Symbol
ERK	EGR1	k-ras	MSH3	EPHB2	COX-2
TCF4	CD44	Raf	MSH2	IL-10	VEGF C
C-erbB2	CASP9	AKT	PMS1	IL-12	VEGF A
RalGDS	CASP3	BAD	Bcl2	u-PA	TGFB
Survivin	DCC	P13K	BAX	SMAD2	TGFB-RII
cyclin-D1	EGF	Axin	PMS2	SMAD3	MDM2
B-Catenin	MMP3	Fos	P53	SMAD4	MLH1
APC	MMp9	LEF	P16	EGFR(RTK)	MSH6
E-cadherin	MCC	Apaf			

4   Result

To test the determinacy and the statistical significance for the second-order and third-order logical relations, suppose that the genes with the second-order and the third-order logical relation have no difference both before and after the test. Here we use statistical simulation to produce the random samples and make disturbance test to samples' attribute to count the probabilities P with which the gene combination of the second-order logical relation occurred during the test. Here the significant level. The main test steps are following: a. Produce an independent and homo-distributional random matrix that keeps the 0-1 distributions invariant in the original matrix; b. Calculate the value, the number of the combinations of the second-order logical genes in the random matrix and count the combination number larger than the one of the second-order logical genes in the original matrix U; c. Repeat the above steps 100 times; d. Calculate P value related to the logical combination in the random matrix. P is the probability that makes the value of logical gene combination in the disturbance test larger than or equal to the U value in the original matrix.

$$P = \frac{(|U^*| \geq |U|)}{C_{51}^4 \times 100} \tag{3}$$

Where U is counted from the original matrix and   is value U of the random matrix produced each time.

With the way mentioned above we determine the third order logical relation. The threshold of the first-order logical relation is T1=0.25 and the one of the second-order logical relation is T2=0.55. we choose T3=0.955 to be the threshold of the third-order logic U, whose supporting degree is over 85% and which deals with 168 relations of logical combination of 22 genes. Through the disturbance test for the logical combination we choose the logical gene combination whose P value is smaller than 0.05 and count the bar chats of the logical relation corresponding to the gene combination. As shown in Figure 2.
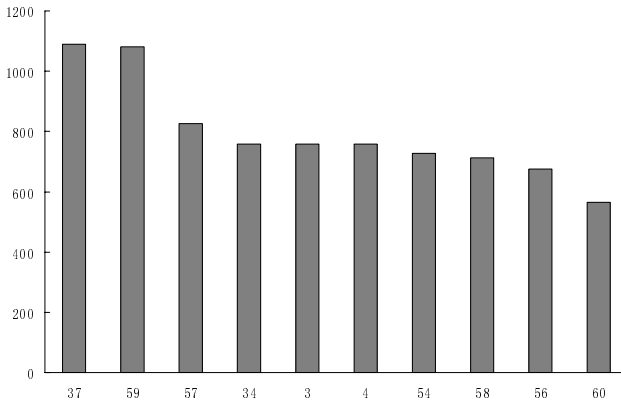
Fig. 2. The histogram showing the number of identified quartets, the X-axis is the type of the logic and Y axis is the number of the logic types

The first 10 types of quaternary gene's logical relation we got after the random test are shown in Figure 2, in which the type of logical relation 37 is the highest.

In the network of the third-order logical relations shown in Figure 3, the logical relation that three genes act on the forth gene simultaneously is in majority (it includes 9 logical relation types). The type that two genes jointly act on the third gene then these three genes act on the forth gene is much less. In the analysis to the quaternary gene logical relation, we found many relations "weeded out" by lower order logical relations. E.g., between p53 and gene Tcf4, MDM2 and Casp9, the logical relation 89 is satisfied, which means when Casp9 expresses and Tcf4 and MDM2 are suppressed simultaneously, p53 expresses. P53, a tumor-suppressing factor, is deficient or inactive in 50% human cancers. The activation of p53 is related to the interactions among MDM2s. MDM2 suppresses the activation of p53 through two ways: a. As E3 ubiquitn ligase, MDM2 can make p53 degradation so as to influence its stability; b. MDM2 can directly act on the N-end to suppress the activity of p53's anti-activation. On the other hand, p53 induces MDM2 to transcription and the result of this process is MDM2 suppresses the activity of p53. The activation of p53 can increase the number of MDM2 to form a negative feedback between p53 and MDM2, i.e., the outcome
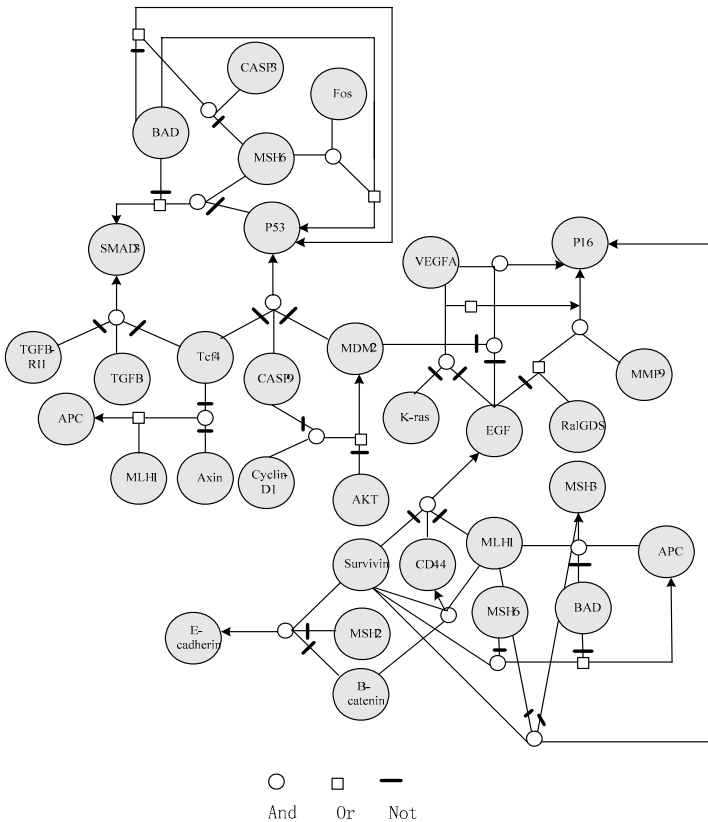
Fig. 3. The third order logical gene expression network of colon cancer

of both can be mutually suppressed. As a molecular switch Tcf4 plays a key function in Wnt pathway and Casp9 partakes the withering and program death pathway of the mitochondria dielectric. The abnormal expression of Casp9's withering and dying gene is related to mini-cell lung cancer, colon cancer, prostate cancer, pancreas cancer, etc. Our analysis shows the expression of cancer-suppressing gene p53 is related to the mutual function between Wnt and cell's program death pathway. The interaction among gene AKT, Cyclin-D1, Casp9 and MDM2 satisfies the 26th logical relation, i.e., when AKT doesn't express or Cyclin-D1 expresses and Casp9 is suppressed, MDM2 expresses. AKT is a kind of serine (threonine) protein kinase.

The activated AKT plays an important role in the growth and proliferation of the dielectric cells, cell's motion and invasion, cell's withering and dying as well as the resistance to chemotherapy and radiotherapy. The transducing pathway of AKT signal is the important signal-transduction pathway inside cells and gives full play to an important biological function in the activities like cell's withering and dying, cell's surviving, proliferation of cell, the change of cellular skeleton, etc., among which the even more important is its regulatory function for cell's withering, dying and surviving. Akt can catalyze Ser166 and Ser188's phosphorylation in MDM2, while the phosphorated MDM2 can more effectively be transposed into the nucleus, functioning as E3 pan-protein ligase to catabolize (degrade) p53 so as to suppress its withering and dying. Feng and others fund that in the fibrocyte of mouse embryo lacking Akt, the expression amount of MDM2 reduces and the expression amount of p53 increases, which greatly increases cell' withering and dying under ultraviolet induction. The regulatory point between cell's G1 stage and S stage is the key regulatory point of cell' proliferation period. YclinD1 is the key protein of cell's proliferation period from G1 stage to S stage, and through the combination with CDK4 it regulates the phosphorylation status of Rb protein so as to regulate cell's proliferation status. CyclinD1 plays a key function of forward regulation during the change from G1 stage to S stage and cell's getting out of control periodically is an important reason for cell's proliferation and canceration. During the process of tumor occurrence, CyclinD1 has various metamorphic forms such as gene cloning, chromosome inversion or transposition, etc., among which the gene cloning is most common. CyclinD1 gene cloning was recorded in the study on esophagus cancer, head and neck neoplasm, breast cancer, etc. According to their logical relations we guess when AKT doesn't express, which makes the regulation for cell's withering, dying and surviving inactive and when the expression cells of cell cycle element CyclineD1 proliferates ceaselessly and the withering and dying gene Casp9 is inactivates, MDM2 gene expresses highly, while the rising of MDM2 makes p53 degrade.

5 Discussion

We used the simple personal computer method for deciphering third order logical relationships among genes. The acceleration ratio and parallel efficiency are two traditional benchmarks to assess the performance of the parallel computing. Figure 4 shows that the computing speed is improved with the nodes increasing. The speed ratio is stable when the number of nodes near 30. From figure 5 the efficiency of the parallel is also increased with the number of nodes added.
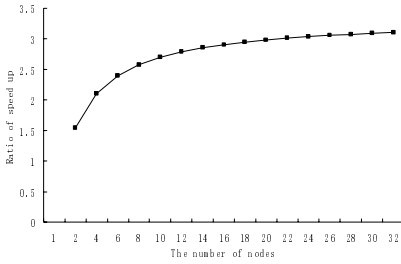
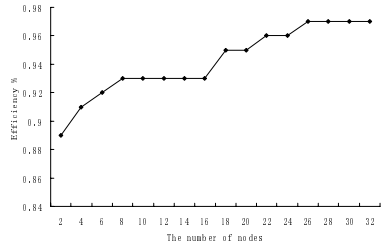**Fig. 4.** Plot of acceleration ratio of parallel computing

Fig. 5. Plot of parallel efficiency

Duo to the little knowledge about the normal and pathologic molecular mechanism insider cells, together with the fact that the space-time property makes the network relations of the cell molecule perplexing, at present it is impossible to totally explain the molecular mechanism for tumor's occurrence. It is also impossible to find the complicated relations among genes by biological experiments. However this complicated relation objectively existing among proteins or genes is of utmost importance for studying and revealing cell's complicated gene network and the inactions among molecules, especially in the study on the molecular mechanism resulting in tumor's occurrence. It has become a main method to build the regulatory network model by the large-scale gene expression through computer simulation together with the technology of artificial intelligence, statistics method, etc. The logical relations studied here reflect, to a certain extent, tumor's complexity; it is not enough to result in tumor's occurrence by one gene's mutation, expression or non-expression. Just as what we mentioned above, since under given conditions, two genes whose mutual function is relatively small in higher order logic may act on the 3rd gene with a certain pattern of logical combination, and these three genes may have relation to the 4th gene through a certain logical relation. Thus we analyzed and studied, one by one, the binary, ternary and quaternary logical relations among genes, excavated all kinds of relations hidden in gene expression profiles and constructed the logical network model for tumor gene expression.

References

1. Segal, E., Friedman, N., Kaminski, N.: From Signatures to Models: Understanding Cancer using Microarrays. Nat. Genet. 37(Suppl.), S38–S45 (2005)
2. Van't, V.L.J., Dai, H., Vande, V.M.J., et al.: Gene Expression Profiling Predicts Clinical Outcome of Breast Cancer. Nature 415(6871), 530–536 (2002)
3. Golub, T.R., Slonim, D.K., Tamayo, P., et al.: Molecular Classification of Cancer: Class Discovery and Class Prediction by Gene Expression Monitoring. Science 286(5439), 531–537 (1999)
4. Weaver, D.C., Workman, C.T., Stormo, G.D.: Modeling Regulatory Networks with Weight Matrices. Pacific. Symp. Biocomp. 4, 112–123 (1999)
5. Akustsu, T., Miyano, S., Kuhara, S.: Identification of Genetic Networks from a Small Number of Gene Expression Patterns under the Boolean Network Model. Paific. Syrup. Biocomp. 4, 17–28 (1999)

6. De, H.M., Imoto, S., Kobayashi, K., Ogasawara, N., Miyano, S.: Inferrring Gene Regulatory Networks from Time-ordered Gene Expression Data of Bacillus Subtilis Using Differential Equations. In: Proceeding of the pacific symposium on biocomputing, vol. 8, pp. 17–28 (2003)

7. Luscombe, N.M., Babu, M.M., Yu, H., Snyder, M.T.A.S., Gerstein, M.: Genomic Analysis of Regulatory Network Dynamics Reveals Large Topological Changes. Nature 431, 308–312 (2004)

8. Bowers, P., Cokus, S., Eisenberg, D., et al.: Use of Logic Relationship to Decipher Protein Network Organization. Science 306, 2246–2249 (2004)

9. Freije, W.A., Castro-Vargas, F.A., et al.: Gene Expression Profiling of Gliomas Strongly Predicts Survival. Cancer Research 64, 6503–6510 (2005)

10. Zhang, X., Kim, S., Wang, T., et al.: Joint Learning of Logic Relationships for Studying Protein Function using Phylogenetic Profiles and the Rosetta Stone Method. IEEE Transaction on signal processiong 54(6), 2427–2435 (2005)

11. Zhong, S., Li, C., Wong, W.H.: ChipInfo: Software for Extracting Gene Annotation and Gene Ontology Information for Microarray Analysis. Nucleic Acids Research 31, 3483–3486 (2003)

12. Kinzler, E.R., Vogelstein, B.: A Genetic Model for Colorectal Tumorigenesis. Cell 66, 589–600 (1990)

13. Ruan, X.G., Wang, J.L., et al.: The Use of Logic Relationships to Model Colon Cancer Gene Expression Networks with mRNA Microarray Data. Journal of Biomedical Informatics (2007), doi:10.1016/j.jbi.11.006

14. Symyrk, T., Lynch, H.T.: Colorectal Cancer: Molecular and Cellular Abnomalities. In: Bertino, J.R. (ed.) Encyclopedia of cancer, San Deigo, pp. 1451–1463. Academic Press, London (1997); Laiho, P., Kokko, A., Vanharanta, S., et al.: Serrated Carcinomas Form a Subclass of Colorectal Cancer with Distinct Molecular Basis. Oncogene 3 (2006), doi:10.1038/sj.onc.1209778

Study on the Agricultural Knowledge Representation Model Based on Fuzzy Production Rules

Chun-Jiang Zhao[1,2] and Hua-Rui Wu[1,2]

[1] National Engineering Research Center for Information Technology in Agriculture,
Beijing, China, 100097
[2] Key Laboratory for Information Technologies in Agriculture, the Ministry of Agriculture,
Beijing, China, 100097
zhaocj@nercita.org.cn

Abstract. The broad knowledge source in the agricultural field causes many problems such as poor knowledge structure, fuzzy and uncertain representation of objective phenomena, which requires that, in the agricultural intelligent system, the knowledge representation and processing pattern could reflect this kind of uncertainty or fuzziness. The representation and reasoning capability of traditional production rules, however, is somewhat insufficient in the representation of knowledge uncertainty or fuzziness. In order to overcome the foregoing insufficiency, the weighed fuzzy logic production rule was put forward to characterize the uncertainty or fuzzy knowledge; the descriptive method of fuzzy production rules was proposed based on BNF, finally, the feasibility and validity of fuzzy production rules on the representation of the uncertain and fuzzy agricultural knowledge was tested with the implemented instance of wheat expert system.

Keywords: Agricultural intelligent system, Agricultural knowledge, Fuzzy production rules, Knowledge representation model.

1 Introduction

A lot of issues in agricultural production are so imprecise that most of the information and knowledge related to human's understanding of the agricultural issues are inaccurate which primarily embodied on fuzziness and uncertainty[1-5]. To satisfy these two aspects, the knowledge representation and processing pattern in the agricultural intelligent system are needed. Uncertainty is becoming more important in the agricultural intelligent system. Studying uncertainty not only needs self-contained and accurate theoretical knowledge, but also requires theory of how construct a model which can reflect expert's thinking and understanding [6,7]. There are a lot of methods for uncertainty study including Bayesian, uncertain theory and proof theory. All these methods based on the probability theory and general probability theory and they are not very feasible to resolve the uncertainty issues in the agricultural issues, in particular, the regular representation of agricultural knowledge. It is hard to the experts elaborating their experienced agricultural knowledge as well as the logical correlation. Usually experts employ some fuzzy words to describe their experienced knowledge as

the reasoning basis [8-14]. Also it is hard for the traditional production rules to be represented and applied with the ambiguous agricultural knowledge rules [15-17].

In order to precisely represent the agricultural knowledge, we aimed at providing a reference model for agriculture intelligent development and proposed a serial agriculture knowledge representation with the agricultural intelligent system which named fuzzy production rules-based agricultural knowledge representation model (FPRAKRM) by analyzing and contrasting to the frequently used knowledge representation ways. On the one hand, we defined the Backus-Naur Form (BNF) of fuzzy production knowledge rule. The flexible knowledge representation modes conform the actual requirement of agricultural, meanwhile the complex agricultural knowledge were organized and enclosed according to the properties and types of question, types of crop and source of knowledge. In addition uncertainty reasoning, certainty reasoning and weighed fuzzy reasoning were well supported.

2   Production Rule of Weighed Fuzzy Logic

In practice, the production rule is different from each other in their importance and information in each production rule. For example, in the rule "if there is cumulonimbus accompanying with wind, thunder and lighting in the sky, it will rain most probably", what most important is obvious "there is cumulonimbus in the sky", whereas windy is not so important. The above rule is characterized by uncertainty. Thus representation is impossible to generate a premise, exclusive and certain rules. However such knowledge can be easily represented by the fuzzy logic rules. In the above rule, the premise, windy, is not important for the results of the rule, whereas, both premises, cumulonimbus in the sky and thundering and lighting, are very important for the results determination. Therefore, we can use the weights method for adding to each premise to represent the knowledge more accurately.

In this paper we use weighted fuzzy logic rule for representing the agricultural knowledge.

Let X demote the weighted propositional rules; T(X) represent the true value of X, whereas W(X) represents the weights of X. W(X) characterizes the dominant degree of the affirmative part of X relative to the existence of X-related event.

Definition 1. Assume P is an atomic proposition, $W_P \in [0,1]$ is the weights of P, and $<P, W_P>$ is a weighed atomic proposition. The true value of $<P, W_P>$ is the same as that of P, i.e. $T(<P, W_P>) = T(P)$ and $W(<P, W_P>) = W_P$.

Definition 2. Every weighed atomic proposition $<P, W_P>$ is a weighed propositional formula, and the true value and weights are $T(P) \in [0,1]$ and $W(P) = W_P$, respectively. Assume X and $\neg X$ are weighed propositional rules respectively, called the false proposition of X. The true value of $\neg X$ is $1 - T(X)$ and its weight is $W(X)$.

Assume $X_1, X_2, \cdots, X_n$ is n pieces of weighed propositional rules and "$\oplus$" represents weighted conjunctive operation, then $X = X_1 \oplus X_2 \oplus \cdots \oplus X_n$ is also a weighed propositional rule, called the weighed conjunctive formula of $X_i (1 \leq i \leq n)$. The

existence of a certain propositional rule means that the true value (or called degree of truth and credibility) of the weighted propositional formula is bigger and equal the stated threshold. Assume all the weighted propositional rules in the collection $\{X_1, X_2, \cdots, X_n\}$ are existent and there are h pieces of minimal elements, $\left\{ T(X_{mj})/[W(X_{mj})/\sum_{i=1}^{n} W(X_i)]^{\alpha} \mid j=1, 2, \cdots, h \right\}$, in the collection $\left\{ T(X_k)/[W(X_k)/\sum_{i=1}^{n} W(X_i)]^{\alpha} \mid k=1, 2, \cdots, n \right\}$. Where α is a constant functioned to regulate the effect of weights on the weighted conjunctive operation.

If $T(X_{mi}) = \min\{T(X_{mj}) \mid j=1, 2, \cdots, h\}$, **then** *the true value and weights of X are*

$T(X) = T(X_{mi}), W(X_{mi})$.

Assume $X_1, X_2, \cdots, X_n$ is n pieces of weighted propositional rule and "$\oplus$" represents the weighted disjunctive operation, then $X = X_1 \otimes X_2 \otimes \cdots \otimes X_n$ is also a weighed propositional rule called the weighted disjunctive rule of $X_i (1 \le i \le n)$. Assume the collection $\{X_{r1}, X_{r2}, \cdots, X_{rm}\}$ is all the existent weighed propositional rules in $\{X_1, X_2, \cdots, X_n\}$, and there are h pieces of maximal elements, $\left\{ T(X_{mt}) \times [W(X_{mt})/\sum_{i=1}^{n} W(X_i)]^{\alpha} \mid t=1, 2,, \cdots, h \right\}$, in the set $\left\{ T(X_{rk}) \times [W(X_{rk})/\sum_{i=1}^{n} W(X_i)]^{\alpha} \mid k=1, 2,, \cdots, m, \quad 1 \le m \le n \right\}$. where α is a constant functioned to regulate the effect of weights on the weighed disjunctive operation.

If $T(X_{ms}) = max \{T(X_{mt}) \mid t=1, 2, \cdots, h\}$,

Then $T(X) = T(X_{ms}), W(X) = W(X_{ms})$ *are the true value and weights of X.*

Assume $X_1, X_2, \cdots, X_n$ is n pieces of weighed propositional rules, then the weighted integrated logistic rules of $X_1, X_2, \cdots, X_n$ is $X = X_{s1} \nabla X_{s2} \nabla \cdots \nabla X_{sm}$. where ∇ is the weighed integrated logistic operator; in $k \le m \le n$, k is bigger than 1; $X_{s1}, X_{s2}, \cdots, X_{sm}$ are all the existent weighed propositional rules in $X_1, X_2, \cdots, X_n$. The true value of X is given

$$T(X) = \frac{\sum_{s=1\ldots m} T(X_s) \times W^{\alpha}(X_s)}{\sum_{j=1}^{n} W^{\alpha}(X_j)} \quad (k \le m \le n), \quad \text{and the weight of X is}$$

$W(X) = \sum_{sj=1}^{m} W(X_{sj}) / \sum_{j=1}^{n} W(X_j)$.

Assume $X = X_{s1} \nabla X_{s2} \nabla \cdots \nabla X_{sm}$ is a weighed propositional formula, then (X) is also a weighed propositional rules equal to X. The true value and weight of (X) are $T((X)) = T(X)$, $W((X)) = W(X)$. All of these weighed propositional rules are formed by using the rules mentioned above with limit times. For any weighed propositional rules X in the weighed fuzzy logic, $T(X) \in [0, 1]$ is true.

In the expert system that discriminates the different importance of various assertion (or called proof and knowledge element) in the premise (or precondition) of the rules, the production rules can be characterized with weighed fuzzy logic operation as follows:

- **IF** *<assertion 1, weight 1>⊗ <assertion 2, weight 2<⊗ … ⊗ <assertion k, weight k> THEN <conclusion, weight k+1><rule's strength>*

The starting premise of the rules is that assertions 1, 2, … and *K* must all be existent (the existence of a assertion means the true value (or value of uncertainty or credibility) of which shall be ≥ the stated threshold, and the true value (or value of uncertainty) of the premise of the rules shall be≥the starting threshold of the rule. ⊗ is the weighed logic and correlation between assertions.

- **IF** *<assertion 1,weight 1>⊕<assertion 2,weight 2>⊕ … ⊕<assertion n,weight n>THEN <conclusion, weight n +1><rule's strength>*

The starting premise of the rules is that at least one of the assertions *1, 2, …, and K* must be existent, and the true value (or value of uncertainty) of the premies of the rule shall be ≥ the starting threshold of the rule. ⊕ is the weighed logic or the correlation between assertions.

- **IF** *<assertion,weight 1> ∇<assertion 2,weight 2>∇ ∇ <assertion m,weight m>THEN <conclusion,weight m +1><rule's strength>*

The starting premise of the rules is that assertions $X_{s1}, X_{s2}...X_{sk}$ are all existent (*t* ≤ *k* ≤ *m*, *t* is usually > 1) in the premise of the rule. True value, $T(X) = \left. \sum_{s=1...m} T(X_s) \times W^{\alpha}(X_s) \middle/ \sum_{j=1}^{n} W^{\alpha}(X_j) \right.$ of the premise of the rules shall be ≥the starting threshold of the rules. ∇ is the weighed integrated logistic correlation between assertions.

3   BNF Representation of Weighed Fuzzy Logic Production Rule

BNF normal form (Backus-Naur Form) is a language to represent the context-free grammar [18,19]. BNF is an ideal format of representing rules, which realize the grammatical irrelevance and the translational conversion of multiple data description format easily, e.g. from BNF to relational and XML modes by programming. We proposed a fuzzy production rules-orientated representation mode of BNF which supplies a method for the standardized representation of agricultural knowledge as well as the reasoning and application agricultural knowledge. The rules are described as follows:

<knowledge bank> ∷ = <name space of knowledge bank>,<name of knowledge bank>,<rule group> | <knowledge bank>,<rule group>

<rule group> ∷ = <name of rule group>,<name of fact table>,<rule> | <rule group>,<rule>

<rule> ∷ = <rule name>[<priority of rules>] ∶ IF <rule premise> THEN <rule conclusion> <rule strength> <condition threshold> <rule interpretation> <rule notes>

< rule premise > ∷= <mode set> [,fuzzy logic relation of mode set]

<mode set> ∷= < mode set >AND<mode>|< mode >|(< mode >)|NULL

<mode> ∷=<comparative unit sets>,<mode weight >,<mode's note>

<comparative unit sets> ∷ =   <comparative unit sets > OR <comparative unit> | < comparative unit >

<comparative unit> ∷ = <fact name> <comparative operator> <comparative value> <unit> | <fact name> | (<comparative unit >)

<fact name> ∷= <character string>

<comparative operator > ::= <> | = | <= | >= | < | >

<comparative value> ::= <value of constant> | <fact name>

< value of constant > ::= <integer> | <real number> | <character string>

<unit> ::= (<character string>) | NULL

<mode's note> ::= <super link> | <multimedia info> | NULL

<fuzzy logic relation of mode sets> ::= <1:min & max fuzzy logic> | <2:weighed average fuzzy logic > | <3:min & max weighed average fuzzy logic> | <4:weight-changing fuzzy logic> | <5:weak logic relation> | …

<rule conclusion>::= <action sets>

< action sets > ::=<action sets >,<action> | <action>

<action> ::= <ASSERT action> | <CALL action>

<ASSERT   action> :: =ASSERT(<name   of   decision-making   data>,<character string>,<unit>,<truth of fuzziness>) | ASSERT(<name of decision-making data>,<expression formula>,<unit>,<fuzzy truth >)

<expression   formula> :: =(<expression   formula>)|<expression   formula><operator> <expression formula> | <Fun1>(<expression formula>) | <Fun2>(<expression   formula string>) |< expression formula>^<number> |<atomic representation formula>

<operator>::= + | ˉ | * | / | %

<Fun1> ::= ABS | INT | REAL | SQR | LOG | EXP | SIN | COS | TAN | ATAN |VAL | DATE |…| (note：Fun1 is a function with only one parameter)

<Fun2> ::= SUM | MIN | MAX | ORDER | … | (note：Fun2 is a function with multiple parameters)

<expression   formula   string> :: = <expression   formula>,<expression   formula> | <expression formula string>,<expression formula>

<atomic   representation   formula> :: = <face name> | <integer> | <real   number> | <character string>

<CALL action > ::= CALL(<model>)

<model> ::= <model name>(<parameter table input> ; <parameter table output>)

::=

::=

< model name > ::= <character string>

::=< parameter table >,< parameter> | <parameter> | NULL

:: =   <name of data-decision item> | <integer> | | <character string>

<name of data-decision item> ::= <true name> | <character string>

<mode weight> ::= real number in [0,1] | NULL

<rule name>::= <character string>

<rule's priority>::= <positive integer>

<name of rule sets>::= <character string>

<name of fact table>::= <character string>

<name of knowledge bank>::= <character string>

<name space of knowledge bank>::=<character string>

<rule strength> ::= [0,1] | NULL

<conditional threshold> ::= real number in [0,1] | NULL

<fuzzy truth> ::= real number in [0,1] | NULL

<rule interpretation> ::=// <character string>| NULL

<rule note> ::=// time of rule acquisition：<character string> ; rule gainer：<character string> ; error scope of rule strength: <real number>%

4   Fuzzy Production Rules-Based Reasoning

Based on the BNF above, the fuzzy set of operator-based fuzzy reasoning model was implemented. The normal pattern of agricultural fuzzy production rule is :P→Q,CF,τ. where P is the various conclusion or action, the premise Q and the conclusion P can be fuzzy, CF is the confidence of rule($0<CF\leq1$), τ is the threshold that satisfies the rule. The meaning of the rule is "if the premise Q is satisfied to a certain degree, then the conclusion P (or executed action P) can be derived with a certain degree of truth, and confidence of the rule is CF.

When we infer the production rules, the matching process of the rule is different with the products. We defined the rules as a fuzzy matching which is similar to measuring the difference between fuzzy data with semantic distance. We proposed a method for weighting the coupling matching degree between any templates. When the matching degree is close to a certain level, it will be assumed successful matching and allows the application of a certain production rule [7]. To fuzz up the premise or condition of the production rule, a fuzzy logistic rule is used to represent the premise or condition in the agricultural intelligent system. And the matching principle is defined as follows: when the sub logistic rule matches to the known knowledge with a certain fuzzy mode, and the matching degree can be converted into degree of truth. Then it can be operated according to the logistic rule. The truth degree of rules are calculated and served as the matching degree of the premise. This processing way is not significantly different from the implicative formula of the fuzzy logic representation and can be complemented easily [9].

Assume the production rule is P→Q,CF,τ, if Q is a fuzzy logic rule, it will be matched hierarchically as follows:

1. Check whether the entire Q matches to the known knowledge or not. If so, then the degree of truth that matches the known knowledge will serve as the matching degree of the premise Q; otherwise, it will be resolved further into sub-formulae of Q.
2. Check whether the sub-rule match to the known knowledge or not. If so, then the degree of truth that matches the known knowledge will serve as the sub-rule. And the degree of truth of the entire logistic rule will be calculated according to the logistic rules and served as the matching degree of the premise Q; otherwise, it will be resolved into sub-rule further, and matches item by item until it is resolved into irresolvable atomic rule. But because every item of the known technology has degree of truth, a compound degree of truth can be calculated and served as the matching degree of premise Q.

Assuming that the matching degree m has been obtained according to the method mentioned above, and if $t = min(m,CF)$,or $t=m*CF$, then the rule can be activated only when $t \geq \tau$. Or action P is executed with strong t, or conclusion P with the degree of truth t is derived. If P ever not appeared in the known knowledge set, it will be added. If P has ever appeared, then P in the known knowledge set will be compared with that of the newly derived one. If it is smaller than the newly derived degree of truth, the old one will be replaced. Otherwise, keep the original degree of truth.

In practice, the importance or information in the sub-premise of reasoning rule is not probably the same [14]. A common fuzzy rule is as follows:

If soil fertility is high (weight=0.1); the leaves are dark green (weight=0.1); the maximal number of stems of the population is more than 1.20 million/mu (weight=0.7) and the length of spring tri-foliage is longer than 24cm (weight=0.1).

then the time for fertilization and watering of jointing will be the emerging of six-foliage in spring, and confidence of the rule shall be 0.95 and the condition threshold be 0.85.

In the above rule, the premise (p) such as high soil fertility and dark green leaves is a fuzzy concept. Their degree of truth is real number ranged in [0, 1]. 0 means a null premise, and 1 mean complete tenable. The maximal number of total stems (> 1.20 million stems/mu) in the population is more important than other items. Therefore, every prerequisite in the prerequisite has a weight (w), and when the degree of truth of the condition is higher than the condition threshold (τ), the conclusion (Q) will be activated. The confidence degree of this rule (CF) is 0.95, and the degree of truth of conclusion (Q) is $t \wedge CF$. where $\wedge$ is cross operation. In order to express this kind of practical knowledge, it is suitable to adopt the weighed fuzzy logic rules. Here, the reasoning rule is characterized by $w_1 * P_1, w_2 * P_2, \ldots, w_n * P_n \rightarrow Q, CF, \tau$, where Q and P are ($j=1,2,\ldots,n$) fuzzy logistic predications; the true value ranged between [0,1]; $j=1,2,\ldots,n$, is the weight coefficient of sub-premise P_j; $CF : 0 < CF \le 1$ is the confidence degree of the rule and $0 < \tau \le 1$ is the applicable threshold of this rule.

The meaning of the above rule is that when the degree of truth, $Dt = j * T(P_j), \ge \tau$, this rule is applicable, where P_j, is the degree of truth ($j = 1,2,\ldots,n$). Application of it will result in the conclusion Q. The degree truth of which is $Dt(Q) = Dt \wedge CF$. where $\wedge$ means the cross operation such as minimal value and multiplication. Therefore, the degree of truth of the conclusion is always $\le$ that of the premise.

In summary, the fuzzy production rules-based reasoning is one kind of subjective confidence –insufficient reasoning. So the flow of reasoning algorithm is as follows:

Assume there is a decision-making data item for question reasoning, the input is the name of the data item for decision-making (ObjectItem) and the output include the value of the decision-making data item, degree of truth (i.e. credibility), conclusion of the decision (interpretation) and remarks of the conclusion.

1. Put the known raw fact (e.g. fact represented with fuzzy proposition) as well as its degree of truth into the middle database called blackboard. The raw known fact is collected from fact bank and user inputting.
2. Check whether the name of the data item is decided (ObjectItem) in the middle database or not. If so, calculate the true value of this data item and search its credibility and rebound it.
3. Search the rule table and find the ID number of the knowledge to be used for deciding this data item. The returned value is a set of knowledge that satisfies this data item. Then this set of knowledge is processed in turn, and if this set of knowledge is used up, exit and output "this item to be decided can not be decided because of scant proof or knowledge."
4. Search the knowledge table according to the ID, and get this knowledge. It is a piece of production rule. This rule is then be processed and if this production rule is used up, back to step 3.

5. Get amount, name and weight value (weightj) of the data item of the precondition from this production rule, and then process every pre-conditional data item of this production rule.
6. If the name of the pre-conditional data item *Pi* of this production rule exit in the blackboard (middle data sheet), then return to the value (degree of truth) of sub-precondition data item (*Pi*).
7. If not, starting from step 2 to resolve the degree of truth and value of this pre-conditional data item *(Pi)* with this *Pi* being a sub-question.
8. If the representation rules of this precondition is empty or a '*', then the degree of truth of this precondition is = ∑(Weight i)*(Confidence i).
9. Otherwise, search and calculate the degree of truth *(T)* of this precondition by substituting the calculated value *(Value j)* of this data item, degree of truth *(Confidence j)* and weight value *(Weight j)* in the representation formula of this precondition.
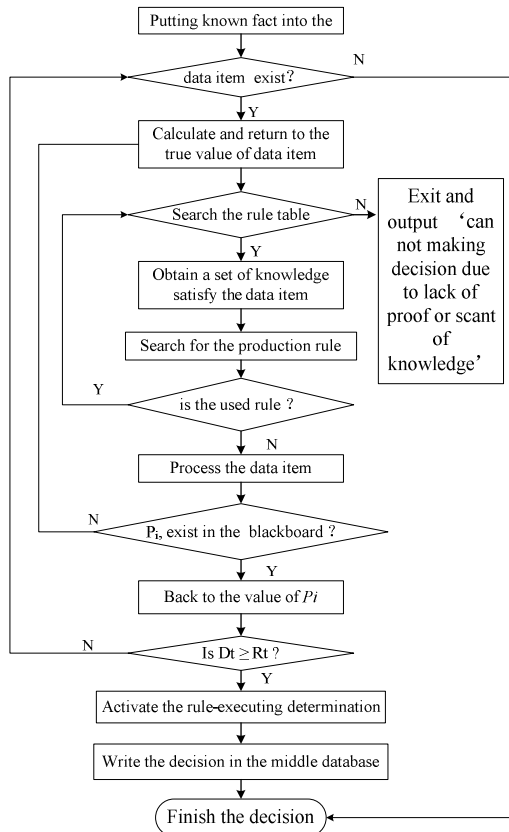
Fig. 1. Reasoning flow based on fuzzy production rules, Dt is degree of truth, Rt is the threshold of the rule

10. If the degree of truth *(T)* of the precondition of this production rule is ≥ the threshold of this production rule, then this rule is activated. The returned value of the decision results can be calculated by substituting the value of pre-conditional data item of this production rule in the result-representation formula of this rule. If *T<CF*, the confidence degree of this rule is *T*. Conclusion of the rule (rule interpretation) and remarks of the conclusion shall be endowed to the returned corresponding item.
11. Write the conclusion in the blackboard of the middle data sheet, and skip out of this recursive function.
12. Otherwise, if the truth degree of this pre-condition of the production rule is smaller than the threshold of this production rule, the next rule shall be chosen and skip to step 4.

5   Experimental Application

AS mentioned above, fuzzy production rule-based agricultural knowledge representation model is applied to the decision-making system of wheat production management. The meteorological information, experimental data, parameter of model and measures will be established firstly for the decision-making system of wheat production and management, and upon which the wheat management information and knowledge-support system are constructed; the second is the establishment of model system. The model bank of wheat management includes the self-constructed model based on engineer's experience and existing mature model, for example, the models regarding to wheat growth and development, yield response, climate prediction, soil moisture, soil and plant nutrients and disease, pests and weeds forecast; then the decision-making subsystems of wheat expert system shall be established, which helps to establish the decision-making service mechanism including meteorological forecast, confirmation of system recognition, comprehensive object and material condition, pre-sowing prediction, over wintering management, pre-stage management in spring, middle and late-stage management in spring, harvesting-stage management and diseases, weeds and pests management.

5.1   The Fuzzy Determination Rule of Target Yield: A Case Study

In the computer expert system of wheat growing and management, the production aim (target yield) conclude the location, sowing time, soil fertility, and irrigation level and cyclic of rotational irrigation. We introduced the following parameters in proposed model:

1. Production potential of light, temperature and rain fall (Y). and it can be calculated through the fact or read from the database according to the different production potential of light, temperature and rainfall in different location;
2. Determine the correction coefficient *(C1)* of sowing time according to expert's experience such as if wheat was sown during 09/18~10/05, *C1=1.0*; if wheat was sown during 10/05~10/08, *C1=0.98;* if wheat was sown after 10/08, *Cl=0.98-sowing time-10/08) ×0.01* (in case of one day later, *C1* will be reduced by 0.01);
3. The correction coefficient of soil fertility *(C2)*: if soil nutrient data are available: *C2=Y1 * Y2 * Y3*. The soil fertility is classified into five grade including high fertility, medial-high fertility, medial fertility, medial-low fertility and low fertility.

Y1 is calculated as alkaline hydrolysable nitrogen, if there is not alkaline hydrolysable nitrogen, *Y1* is calculated as total nitrogen. If there is not total nitrogen, *Y1* is calculated as organic matter, if three of them are not existent, *Y1* will be processed as nitrogen-free data; if *Y1* is exi*t, C2* is *y2*y2*y3;* if there is not *y2, C2*is *y1*y1*y3* ; if there is *y3, C2* is y*1*y1*y2* ; if both *y1* and *y2* are absent, *C2* will be *y3*y3*y3*; when both *y1* and *y3* are absent, *C2* is *y2*y2*y2;* if *y2* and *y3* are absent, *C2* is y1*y1*y1; if any of *y1, y2* or *y3* is absent, *C2* is computed following: *C2=1.0* (high fertility); *C2=0.95~1.0* (medial high fertility); *C2=0.90~0.95* (medial fertility); *C2=0.85~0.90* (medial low fertility) and *C2=0.80~0.85* (low fertility).

Table 1. The correction coefficient of soil fertility

Y1						Y2		Y3	
Alkaline hydrolysable nitrogen		Total nitrogen		Organic matter		Available phosphorus		Available potassium	
ppm	C	%	C	%	C	Ppm	C	Ppm	C
>110	1.0	>0.12	1.0	>2.0	1.0	>40	1.0	>150	1.0
90~110	1.0	0.1~0.12	1.0	1.5~2.0	1.0	30~40	1.0	120~150	1.0
70~90	0.98	0.08~0.1	0.98	1.3~1.5	0.98	20~30	0.98	100~120	0.98
50~70	0.96	0.06~0.08	0.96	1.0~1.3	0.96	10~20	0.96	80~100	0.96
<50	0.94	<0.6	0.94	<1.0	0.94	<10	0.94	<80	0.94

* *C is the coefficient.*

4. Irrigation level *(C3)* is determined according to amount of water used for irrigation including 4-water, 3-water, 2-water and 1-water.
5. The parameter of rotational irrigation cyclic *(C4)* is determined according to the cyclic of rotational irrigation water, which corresponds to a rotational cyclic being shorter than 7 days, 7-10 days, 10-15 days and longer than 15 days, respectively. The theoretical target yield $Yg=Y*C1*C2*C3*C4$ (customer's expected target yield).

According to IF-THEN production knowledge representation rule, the best rule is as follows: If customer's target yield *(Yt)* is bigger than *Yg*0.95* and the determined target yield *(Yd) is Yg*0.95*, then the target yield is slightly higher and *Yd=Yg*0.95*. If customer's expected target yield *(Ye) is Yg*0.80*, and *Yd=Yg*0.80*, then *Yt* is slightly lower, *Yt=Yg*0.80. If Yg*0.8 < Yg*0.95* (*Ye*, and *Yd = Yt*, then the target yield is proper *(***/2 kg).*

5.2 Extraction Knowledge Rule

The models proposed in this paper are complemented with Vs.net. Fig 2 is the interface of pre-sowing decision-making rule. The rule named BPJC, described as sowing determination and classified as the final result. There are two kinds of rule types: the final result and media result. The final result is the customer wants to extract from the rules, while the medial result is the medial variable of the final result. Fig 3 is the reasoning result derived from the definition-based fuzzy rule.
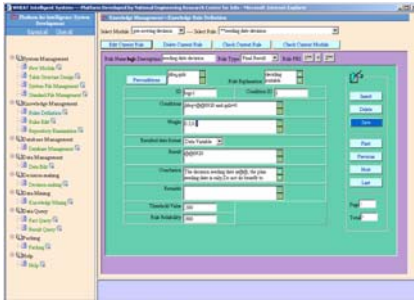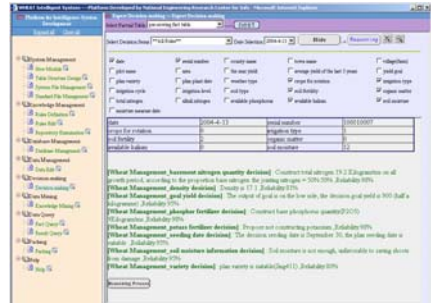
Fig. 2. Editing the knowledge rule

Fig. 3. Results of decision-making

6 Conclusion

In summary the standardized representation of agricultural knowledge is complemented based on the extended model of the weighed fuzzy logic production knowledge rule. The implementation of standardized representation of fuzzy agricultural knowledge avoids the problem that uses the same weight in the precondition bearing different importance in the knowledge rule. Meanwhile, the proposition of BNF-based descriptive way of knowledge rule realizes the formalized description of weighted fuzzy logic production knowledge rule, extends the usable knowledge source during the reasoning agricultural intelligent decision-making system and supplies methodological basis for the application of agricultural indigenous expert's experimental knowledge. Furthermore, the precision of reasoning is improved via the extension of knowledge source. The case study demonstrated that this model and the knowledge description method can use accomplish the agricultural intelligent decision-making using various fuzzy knowledge effectively. As for the construction of weighed fuzzy logic production knowledge rule, weight is an important factor that affects the precision of the rule reasoning. Therefore, emphasis will be imposed on how to define precisely the weight of knowledge rule in the future.

Acknowledgements

References

1. Li, L.L., Zhao, H.Y.: Agricultural Insect Pests Expert System based on Case and Fuzzy Technology. Computer Engineering and Design 28(22), 5570–5572 (2007)
2. Zhao, L.F., Xiong, X.Y.: Development and application of agricultural expert system. Agriculture Network Information 6, 43–45 (2007)

3. Gao, L.W., Chen, J.G., Yu, X.W., et al.: Research and Development of the Expert System Platform for Forecast and Prediction of Agricultural Pests. Transactions of the Chinese Society of Agricultural Engineering 22(10), 154–158 (2006)
4. Zhu, R.X., Qiao, J.Y., Wang, F.L.: Research Status and Problems of Expert Systems in Agricultural Mechanization Field. Journal of Northeast Agricultural University (6), 852–855 (2007)
5. Zhang, Q.X., Li, M., Zhang, J.: Research and Realization of Control Strategy of Agriculture Intelligence System Network Reasons. Computer Engineering and Applications 42(35), 216–218 (2006)
6. Silvano, M.: Diagnostic Expert Systems: A Method for Engineering Knowledge Used in Sequential Diagnosis Expert Systems 17(4), 55–59 (2000)
7. Han, J.W., Kamber, M.: Data Mining: Concepts and Techniques, pp. 112–117. Morgan Kaufmann, San Mateo (2000)
8. An, S.H.: On the Weak Norms of the Compound Fuzzy Proposition. Chinese Journal of Computers 24(20), 1071–1075 (2001)
9. Liu, D.Y., Yang, K., Tang, H.Y., et al.: A Convex Evidence Theory Model. Journal of Computer Research and Development 37(2), 175–18 (2001)
10. Zhifeng, L.: Research on Fuzzy Logic. Computer Engineering and Applications 8, 10–13 (1999)
11. Mookerjee, V.S., Mannino, M.V.: Sequential Decision Models for Expert System Optimization. IEEE trans. on Knowledge and Data Engineering. 9(5), 675–687 (1997)
12. Guoqiang, Y., Shunxi, W., Wei, D.: Development of Research on in Agriculture Expert System in China. Journal of Shanxi Agricultural University 24(3), 303–305, 312 (2004)
13. Peizhuang, W., Hongxing, L.: Fuzzy System Theory and Fuzzy Computer, pp. 13–18. Science Press, Beijing (1996)
14. Van, A.P., Ragab, R.A.: Improving Irrigation Management by Modeling the Irrigation Schedule. Agricultural Water Management 13, 113–125 (1998)
15. McCown, R.L.: Model Testing and Simulation in Agricultural Systems Research. Agricultural Systems 50, 255–271 (1996)
16. Wang, X.Z., Liu, J.S., Wei, Y.Q.: Consistence of the Judgement Matrix and Ordering of Weights in Fuzzy AHP. Systems Engineering Theory & Practice 1, 28–35 (1995)
17. Wei, L.Y.: The Consistency of Interval Number Judgement Matrix And Its Discriminating Methods. Journal of Guangxi Normal University (Natural Science) 19(4), 40–16 (2001)
18. Ming, G.: BNF Description And Definition of Extended Constraints-Based Modular Petri Nets of Use Case Diagram. Computer Applications and Software 22(11), 26–27, 114 (2005)
19. Yang, M.H., Guo, Z.D.: A General Syntax Analysis Algorithm Based on Extended BNF Grammar. Journal of Software 3(3), 24–32 (1992)

A Plausible Model for Cellular Self-defense Mechanisms in Response to Continuous Ion Radiation(IR) under Radiotherapy

Jin-Peng Qi, Shi-Huang Shao, and Yi-Zhen Shen

College of Information Sciences and Technology, Donghua University,
Shanghai 201620, P.R. China
qipengkai@dhu.edu.cn

Abstract. Radiotherapy can induce DNA damage into cells and trigger the cell cycle arrest and cell apoptosis by regulating the vital genes and their signal pathways. To illustrate the cellular self-defense mechanisms in fighting against genome stresses under radiotherapy, a model of P53 stress response networks is proposed by using the methods with the system biology and cyber-biology at single cell level. The kinetics of Double Strand Breaks (DSBs) generation and repair, ARF and ATM activation, P53-MDM2 feedback regulation, as well as the toxins degradation are presented versus continuous radiation time. The model provides a theoretical framework to illustrate the complicated kinetics in cellular response to acute IR under radiotherapy.

Keywords: IR; DNA damage, Cellular network, Radiotherapy, P53.

1 Introduction

Radiotherapy is one of the major tools in fighting against cancer. Under acute IR, cells can trigger their self-defense mechanisms in response to genome stresses [1],[2]-[4]. As one of the pivotal anticancer genes, P53 controls the transcription and translation of series genes through their complicated signal pathways [3], and then trigger cell cycle arrest and abnormal cell apoptosis further [4]. Abnormalities in the P53 tumor suppressor have been identified in over 60% of human cancers [4], the inactivation of P53 gene may nullify the activity of certain antitumor agents [2],[5]. Conversely, the P53 suppressor gene is the most commonly known specific target of mutation in tumorigenesis The status of P53 within tumor cells has been proposed to be one of the determinant response to anticancer therapies [3],[4]. The outcome of radiotherapy may depend on the presence of functional P53 protein to induce tumor regression through apoptotic pathways [3],[5]. Therefore, P53 status in tumor cell can be considered as a predictor for long-term biochemical control during and after radiotherapy [6]-[8].

Recently, several models have been proposed to explain the damped oscillations of P53 in cell populations [9]-[13]. However, the complicated cellular self-defensive mechanisms in response to DNA damage need to be further addressed [14]-[16].

Based on the existing models [9]-[13] and the latest biomedical studies [2]-[4],[14]-[18], a model of P53 stress response networks under radiotherapy is proposed at single cell level. The module of DSBs generating and their repair, ATM and ARF activation, P53-MDM2 feedback loop, as well as the toxins degradation including depression of the oncogenes, elimination of the mis-repair part of DSBs (Fw) and degradation of the mutation of P53 (mP53), are implemented by using differential equations and graphic methods.

The combined approaches of system analysis, control theory, and computer science can stimulate new approaches to simulate the investigation of the complicated mechanisms of cellular responding genome stresses [19]. These methods provide a good link among the diverging areas of biomedicine, mathematics and bioinformatics [20],[21]. In addition, using differential equations and graphic approaches to study various dynamical and kinetic processes of biological systems can provide useful insights. It is indicated by many previous studies on a series of important biological topics, such as enzyme-catalyzed reactions etc. [10],[19]. In this study, we use the differential equations and graphic approaches in order to study the cellular response mechanisms under radiotherapy.

2   Model Implementation

2.1   Model Overview

As shown in Fig.1, the integrated model scheme of P53 stress response networks under radiotherapy is composed of four parts, including DSBs generation and repair, ARF and ATM activation, P53-MDM2 feedback regulation, as well as toxins degradation. According to the latest biomedical studies [1]-[5], the more vital components are involved in this model, such as oncogenes, ARF and mP53, as well as their related regulation pathways.
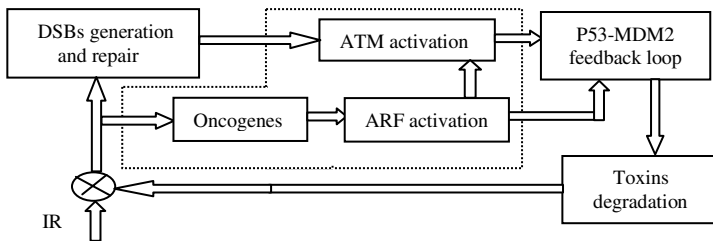
Fig. 1. The model scheme of P53 stress response networks under radiotherapy. It is composed of four parts, including DSBs generation and repair, ATM and ARF activation, P53-MDM2 feedback loop, as well as toxins degradation.

As acute IR is applied into a cell, the resulting DSBs occur and form the DSB–protein complexes (DSBCs) at damage sites after interacting with the DNA repair proteins (RP) [2],[3]. As a sensor of DNA damage, ATM is activated by DSBCs transferring from damage sites [4]. Meanwhile, oncogenes are prompted by acute IR, and further trigger the activation of ARF [5],[14]. Under the cooperation of active

ATM (ATM*) and active ARF (ARF*), P53-MDM2 feedback loop switches on or off the cellular defense mechanisms [2],[7],[14], and further controls the toxins degradation process by regulating the downstream genes and their complicated signal pathways [5],[8]. Here, we use the superscript * to represent the activate state.

2.2   DSBs Generation and Repair

The typical DSBs repair process exhibits biphasic dynamics, both of which include first-order and second-order processes [2],[3]. As shown in Fig.2, DSBs repair process contains both a fast and a slow kinetics, each of which is composed of a reversible binding of repair proteins and DSB lesions into DSBCs, and an irreversible process from the DSBCs to the fixed DSBs [2],[3],[10]. DSBCs are synthesized by binding the resulting DSBs with repair proteins (RP), which is the main signal source to transfer the DNA damage to P53-MDM2 feedback loop by ATM activation [10].
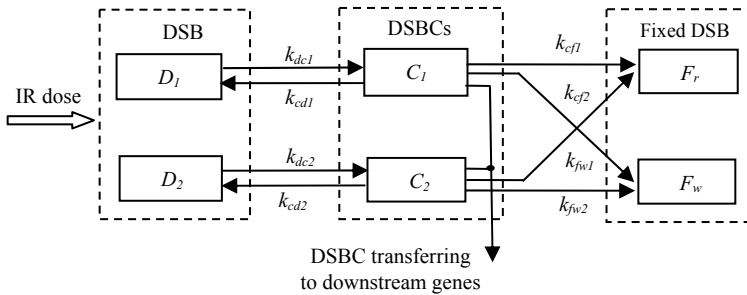
Fig. 2. The module scheme of DNA damage repair process. It includes both a fast repair pathway and a slow one. DSB can be in one of four states: intact DSB (DSB), DBSC, Fr and Fw. Subscripts '1' and '2' refer to the fast kinetics and the slow one.

As continuous IR is applied into a cell, we assume that the quantity of the resulting DSBs per time scale is proportional to the number generated by a Poisson random function under constant IR [2],[3],[10]-[13]. The DSBs generation process is formulated as follows:

$$\frac{d[DT]}{dt} = k_t \times Poissrnd \ (a_{ir} \times IR) \, , \qquad (1)$$

where [DT] is the concentration of total resulting DSBs induced by IR in both fast and slow repair processes. kt is the parameter to set the number of DSBs per time scale, and air is the parameter to set the number of DSBs per IR dose.

Due to the misrepair part of DSBs (Fw) have serious consequences on the subsequent cellular viability and the cellular response in fighting against genome stresses [2],[3], we obviously distinguish between correct repair part of DSBs (Fr) and Fw [11]-[13]. Moreover, the total Fw in both repair processes is dealt as a part of toxins remaining within the cell [2],[13]. In addition, we deal that the limited repair proteins are available around the damage sites, and DSB can be in one of four states, including intact DSB, DSBC, Fr and Fw in each damage locus [11]-[13]. The main formulations used in this module are listed as follows:

$$\frac{d[D_1]}{dt} = a_1[D_t] + k_{cd1}[C_1] - [RP](k_{dc1}[D_1] + k_{cross}([D_1] + [D_2])), \tag{2}$$

$$\frac{d[D_2]}{dt} = a_2[D_t] + k_{cd2}[C_2] - [RP](k_{dc2}[D_2] + k_{cross}([D_1] + [D_2])), \tag{3}$$

$$\frac{d[C_1]}{dt} = k_{dc1}[D_1] - k_{cd1}[C_1] - k_{cf1}[C_1], \tag{4}$$

$$\frac{d[C_2]}{dt} = k_{dc2}[D_2] - k_{cd2}[C_2] - k_{cf2}[C_2], \tag{5}$$

$$\frac{d[RP]}{dt} = S_{rp} + k_{cd1}[C_1] + k_{cd2}[C_2] - [RP](k_{dc1}[D_1] + k_{dc2}[D_2] + k_{cross}([D_1] + [D_2])), \tag{6}$$

$$\frac{d[F_w]}{dt} = k_{fw1}[C_1] + k_{fw2}[C_2], \tag{7}$$

where [D], [C], [RP], and [Fw] represent the concentrations of DSBs, DSBCs, and Fw in both repair kinetics respectively. kdc, kcd, kcf, and kfw are the transition rates between four states of DSBs. kdc, and kcross represent the first-order and second-order rate in both the fast and the slow repair kinetics respectively [11],[13]. Srp is the basal induction rate of repair mRNA, and subscripts '1' and '2' refer to the fast and the slow kinetics.

2.3 ATM and ARF Activation

Shown in Fig.3 is the module scheme of ATM and ARF activation. As acute IR is applied, DSBCs which are synthesized by DSBs and RP are transferred from damage sites, and then trigger the phosphorylation of inactive ATM monomers [2],[5]. Meanwhile, ARF, another tumor suppressor, is activated by hyperproliferative signals emanating from oncogenes, such as Ras, c-myc etc., and further prompts the ATM activation [2],[7],[14]. Therefore, the pathways of ATM and ARF activation are needed to efficiently enhance the regulatory functions of P53 to eliminate the malignant cells indirectly [2],[5],[7],[8].

Furthermore, according to the latest biomedical studies [2],[3],[5], the rate of ATM activation is treated as a function of the amount of DSBCs, ARF* and the self-feedback of ATM* [10],[14], and the total concentration of ATM is dealt as a constant, including ATM dimer, ATM monomer and ATM* [11]-[13]. ATM can be activated rapidly with the cooperating effects of DSBCs, ARF* and the positive feedback from ATM* for intermolecular autophosphorylation [1],[2],[10],[14]. The main formulations used in this module are as follows:

$$\frac{d[ATM_d]}{dt} = \frac{1}{2}k_{dim}[ATM_m]^2 - k_{undim}[ATM_d], \tag{8}$$

$$\frac{d[ATM_m]}{dt} = 2k_{undim}[ATM_d] - k_{dim}[ATM_m]^2 - k_{af}f[ATM_m] + k_{ar}[ATM*], \tag{9}$$

$$\frac{d[ATM*]}{dt} = k_{af}f[ATM_m] - k_{ar}[ATM*], \tag{10}$$

$$\frac{d[ARF]}{dt} = S_{arf} - k_{ad}[ARF] - k_{onf}[Onco][ARF], \tag{11}$$

$$\frac{d[ARF^*]}{dt} = k_{onf}[Onco][ARF] - k_{pad}[ARF^*] \ , \tag{12}$$

$$f(C,[ARF^*],[ATM^*]) = a_1 C + a_2[ATM^*] + a_3 C[ATM^*] + a_4[ARF^*], \tag{13}$$

where [ATMd], [ATM] and [ATM*] represent the concentrations of ATM dimer, ATM monomer, and active ATM respectively; [Onco], [ARF] and [ARF*] represent the concentrations of oncogenes, ARF, and active ARF respectively; kundim, and kdim are the ATM undimerization and dimerization rates, respectively; kar, and kaf are the ATM monomer inactivation and activation rates, respectively. In addition, f is the function of ATM activation, the term a1C implies the fact that DSBs somehow activate ATM molecules at a distance, a2[ATM*] indicates the mechanism of auto-phosphorylation of ATM, a3C[ATM*] represents the interaction between the DSBCs and ATM* [11]-[13], and a4[ARF*] represents the interaction between the ARF* and ATM* [1],[3],[7],[14].
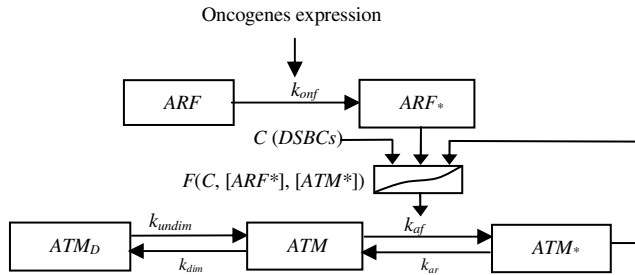
Fig. 3. The module scheme of ATM and ARF activation under IR. ARF is activated by the over-expression of oncogenes, ATM is activated from ATM monomers under the cooperating effects of DSBCs, ARF*, and self-feedback of ATM*.

2.4 P53-MDM2 Feedback Loop

P53-MDM2 feedback loop is the core part in the integrated networks [10]-[13]. Recent studies show that due to the synergistic functions of ARF and ATM kinases in P53 activation, both ARF* and ATM* are required for the efficient induction of apoptosis in response to genome and oncogenes stresses [2],[7],[14].

In this module, as shown in Fig.4, we deal that DSBCs are the main signal source from damage sites to the downstream P53-MDM2 feedback loop through ATM activation, which is important to transfer DNA damage and activate the cellular self-defensive mechanisms in fighting against genome stresses further [3],[6],[10]-[13]. Meanwhile, ARF* can increase the regulating function of P53 by inhibiting the expression of MDM2 and preventing P53 degradation further [2],[7],[14]. With the cooperation of both ATM and ARF, this module can produce oscillations in response to sufficiently strong IR dose [2],[7],[9]-[13]. The main formulations in this part are as follows:

$$\frac{d[P53_R]}{dt} = S_{P53} - d_{rp}[P53_R] - k_{rp}[P53_R], \tag{14}$$

$$\frac{d[P53_P]}{dt} = k_{rp}[P53_R] + k_{p*p}[P53*] - d_{pp}[P53_P]$$

$$- k_{app*}[ATM*]\frac{[P53_P]}{[P53_P]+k_p} - k_{mp}[MDM2_P]\frac{[P53_P]}{[P53_P]+k_d} \quad , \tag{15}$$

$$\frac{d[P53*]}{dt} = k_{app*}[ATM*]\frac{[P53_P]}{[P53_P]+k_p} - k_{p*p}[P53*] - d_{pp*}[P53*]$$

$$- k_{p*p}[P53*] - d_{pp*}[P53*] - k_{mp*}[MDM2_P]\frac{[P53*]}{[P53*]+k_{d*}} , \tag{16}$$

$$\frac{d[MDM2_R]}{dt} = S_{mdm2} + k_{p*m}\frac{[P53*]^n}{[P53*]^n + k^n} - k_{mrp}[MDM2_R] - d_{mr}[MDM2_R] , \tag{17}$$

$$\frac{d[MDM2_P]}{dt} = k_{mrp}[MDM2_R] - d_{mp}[MDM2_P]$$

$$- (k_{mat}\frac{[ATM*]}{[ATM*]+k_{at}} + k_{mar}\frac{[ARF*]}{[ARF*]+k_{ar}})[MDM2_P] , \tag{18}$$

where [P53R], [P53P], [P53*], [MDM2R], and [MDM2P] represent the concentrations of P53 mRNA, P53 protein, active P53, MDM2 mRNA, and MDM2 protein, respectively. SP53, and SMDM2 is the basal induction rate of P53 mRNA, and MDM2 mRNA. k represent the regulation rates among genes and proteins, and d represent the degradation rates of mRNAs and proteins.

Fig. 4. The diagram of P53-MDM2 feedback loop under radiotherapy. P53 is translated from p53mRNA and phosphorylated by ATM* and ARF*. Mdm2 protein promotes a fast degradation of P53 protein and a slow degradation of P53*. In addition, ATM* and ARF* stimulate the degradation of Mdm2, and then indirectly increase the regulatory activation of P53* further.

2.5 Toxins Degradation

P53 acts the important role in triggering the cellular response to fight against genome stresses. The module of toxins degradation is shown in Fig.5, in which the mutation of P53 (mP53) from P53R triggered by oncogenes is added in this module [3],[5]. mP53 is further dealt as another outcome indictor in cellular response to genome stresses under radiotherapy [8],[18]. Especially, we deal that only P53* can induce target genes to depress the over-expression of oncogenes, and degrade the Fw and mP53 within the cell [3],[4],[11]-[13]. The main formulations used in this module are as follows:

$$\frac{d[Onco]}{dt} = k_{onIR}[Onco][IR] - k_{onp}[Onco][P53^*] \,, \tag{19}$$

$$\frac{d[F_w]}{dt} = k_{tfw}[F_w] - k_{pt}[P53^*][F_w] \,, \tag{20}$$

$$\frac{d[mP53]}{dt} = k_{mp}[P53_R][Onco] - k_{pmd}[P53_{P*}][mP53] \,, \tag{21}$$

where [Onco], [Fw] and [mP53] represent the concentrations of oncogenes, Fw and mP53 respectively. [IR] represents the strength of IR dose. k represents the regulation rate among genes and proteins.
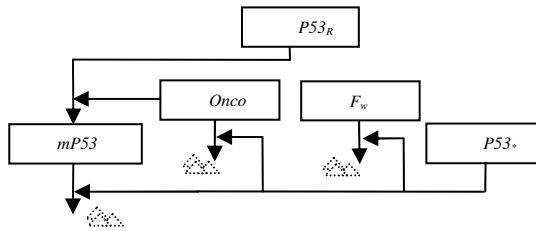
Fig. 5. The diagram of toxins degradation under effect of P53*. Oncogenes, Fw and mP53 are degraded directly by the degradation functions of P53*.

3   Simulation Results

Based on the studies in previous models [9]-[13], we implement the kinetics of cellular response under continuous IR by using simulation platform in MATLAB7.0.

3.1   Kinetics of DSBs Generation and Repair

In our simulations, the continuous 9Gy IR is applied into a cell. Shown in Fig.6a is the stochastic trace of DSBs generation versus the continuous radiation time.

As the resulting DSBs increase and the available RP are limited around increasing damage sites, the simulation in Fig.6b shows that RP available around damage sites keep decreasing, and the DSBCs synthesis increase slowly and trend to a dynamic equilibrium. Subsequently, the intact DSBs increase, and Fw is accumulated within the cell fast without the regulating functions of P53*.

3.2   Kinetics of ARF and ATM Activation

The kinetics of ARF activation is shown in Fig.7a. ARF is activated fast due to the over-expression of oncogenes without depressing functions of P53*.

Meanwhile, Fig.7b shows that ATM is activated rapidly and switches to "on" state with the cooperating functions of DSBCs, ARF* and the positive self-feedback from ATM*. These simulations above suggest that under the cooperation of both ATM and ARF activation, DNA damage signals can be further transferred to the P53-MDM2 feedback loop more efficiently [2],[7],[14].
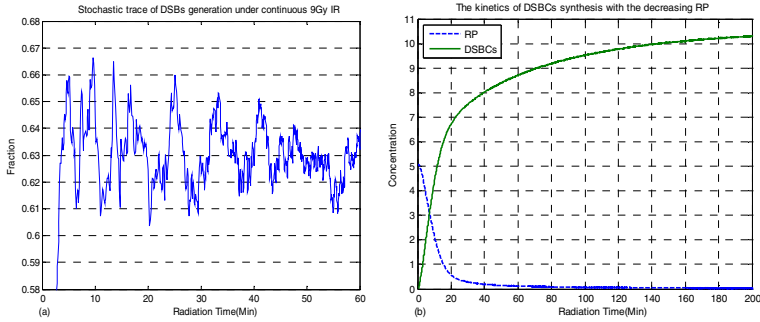
Fig. 6. The kinetics of DSBs generation and repair process under continuous effect of 9Gy IR. (a) the stochastic trace of DSBs generation versus radiation time. (b) the kinetics of DSBCs synthesis with the available RP decreasing.
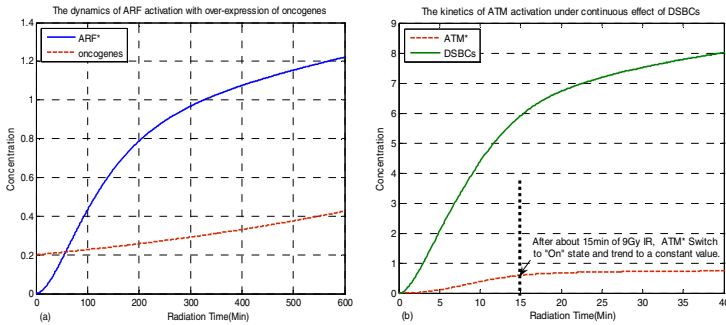
Fig. 7. The kinetics of ARF and ATM activation under 9Gy IR. (a) the kinetics of ARF activation in response to over-expression of oncogenes induced by acute IR. (b) the switch-like kinetics of ATM activation, ATM* switch to "on" state in response to DSBCs transduction.

3.3   Kinetics of Toxins Degradation

With the cooperation of ATM* and ARF*, as shown in Fig.8a, P53 and MDM2 generate oscillations in response to continuous application of 9Gy IR from time 0. The period of these oscillations is about 400min, and their phase difference is about 100min. Moreover, the first pulse is slightly higher than the second, and the oscillating kinetics trend to a new equilibrium versus radiation time. All of these results are quite similar with the experimental observations in [2],[7],[14] and the simulation results in [11]-[13].

Meanwhile, we deal that toxins within the cell, including Fw, oncogenes, and mP53, can be directly degraded by P53*. As shown in Fig.8b, Fw keeps increasing dramatically without P53*. Contrarily, Fw keeps a low level with slightly increase under the degrading functions of P53*. In addition, the kinetics of oncogenes and mP53 degradation are plotted in Fig.8c and Fig.8d respectively. We can see that the expression levels of oncogenes and mP53 keep decreasing and similarly trend to dynamic equilibrium state under P53*. Whereas, without the degradation of P53*, the expression level of oncogenes and mP53 increase dramatically versus radiation time. All of these simulations above suggest that P53* indeed act an important role in fighting against genome damage under radiotherapy.
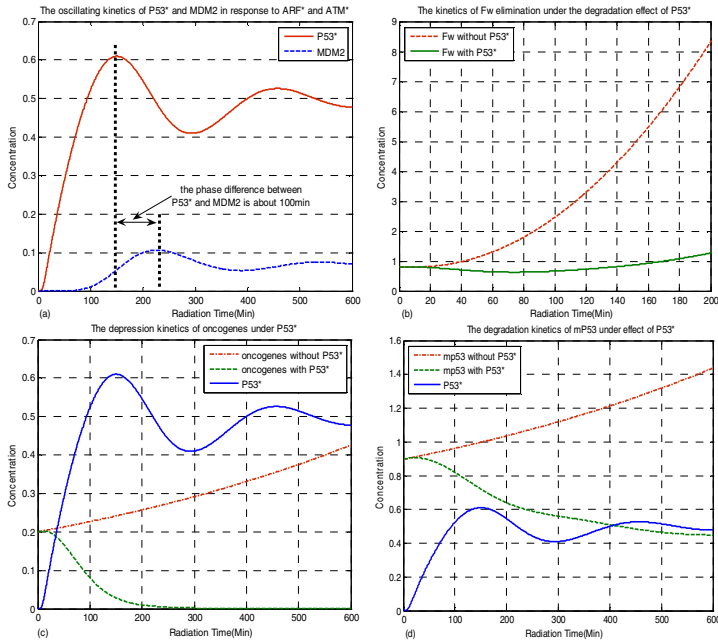
Fig. 8. The kinetics of cellular response under 9Gy IR of radiotherapy. (a) the oscillating kinetics of P53* and MDM2 under the cooperative effect of ATM* and ARF*. (b) the kinetics of Fw elimination triggered by the functions of P53*. (c) the depressing dynamics of oncogenes over-expression with the regulations of P53*. (d) the kinetics of mP53 elimination triggered by the effect of P53*.

4   Discussion and Conclusion

By using a set of differential equations, as well as graphic methods, a plausible model is proposed for illustrating the cellular self-defense mechanisms in response to genome stresses under radiotherapy. In our simulations, DSBCs are synthesized after DSBs combined with RP as acute IR is applied into a cell. ATM and ARF exhibit a strong sensitive behavior in response to DNA damage and over-expression of oncogenes. Especially, P53-MDM2 feedback loop acts vital roles in triggering the cellular self-defense mechanisms and eliminating toxins within the cell, including Fw, oncogenes, and mP53.

References

1. Wang, X.F., Chen, G.: Complex Networks: Small World, Scale World, Scale Free and Beyond. IEEE Circuits and System Magazine 3(1), 6–21 (2003)
2. Li, L., Story, M., Legerski, R.: Cellular Responses to Ionizing Radiation Damage. Int. J. Radiat. Oncol. Biol. Phys. 49, 1157–1162 (2001)

3. Nicolas, M., Toillon, R.A., et al.: NF-κB Modulation and Ionizing Radiation: Mechanisms and Future Directions for Cancer Treatment. Cancer Letters 2(18), 158–168 (2006)

4. Oren, M.: Decision making by P53: Life, Death and Cancer. Cell Death and Differentiation 10, 431–442 (2003)

5. Kohn, K.W., Pommier, Y.: Molecular Interaction Map of the P53 and MDM2 Logic Elements, Which Control the Off-On Switch of P53 in Response to DNA Damage. Biochemical and Biophysical Research Communications 331, 816–827 (2005)

6. Vogelstein, B., Lane, D., Levine, A.J.: Surfing the P53 Network. Nature 8(6810), 307–310 (2000)

7. Perez, C., Brady, L.: Principles and Practice of Radiation Oncology, pp. 784–785. Lippincott-Raven, Philadelphia (1998)

8. Tjebbes, G.W.A., Kreijveldb, P.A., et al.: P53 Tumor Suppressor Gene Mutations in Laryngeal Cancer and in Recurrent Disease Following Radiation Therapy. Oral Oncology 38, 296–300 (2002)

9. Bar-Or, R.L., Maya, R., et al.: Generation of Oscillations by the P53-MDM2 Feedback Loop. Proc. Natl. Acad. Sci. USA 97, 11250–11255 (2000)

10. Ma, L., Wagner, J., et al.: A Plausible Model for the Digital Response of P53 to DNA Damage. PNAS 2(40), 14266–14271 (2005)

11. Qi, J.P., Shao, S.H., Shen, Y.Z.: Cellular Responding DNA Damage: An Improved Modeling of P53 Gene Regulatory Networks under Ion Radiation (IR). Applied Mathematics and Computation 205, 73–83 (2008)

12. Qi, J.P., Shao, S.H., et al.: A Dynamic Model for the P53 Stress Response Networks under Ion Radiation. Amino Acids 33(1), 75–83 (2007)

13. Qi, J.P., Shao, S.H., et al.: A Mathematical Model of P53 Gene Regulatory Networks under Radiotherapy. Biosystems 90(3), 698–706 (2007)

14. Lindstrom, M.S., Wiman, K.G.: Myc and E2F1 Induce P53 Through p14ARF-Independent Mechanisms in Human Fibroblasts. Oncogene 22, 4993–5005 (2003)

15. Cuddihy, A.R., Bristow, R.G.: The P53 Protein Family and Radiation Sensitivity: Yes or no? Cancer and Metastasis Reviews 23, 237–257 (2004)

16. Ritter, M.A., Gilchrist, K.W., et al.: The Role of P53 in Radiation Therapy Outcomes for Favorable-to-Intermediate-Risk Prostate Cancer T. Radiation Oncology Biol. Phys. 53, 574–580 (2002)

17. Weller, M.: Predicting Response to Cancer Chemotherapy: The Role of P53. Cell Tissue Res. 292, 435–445 (1998)

18. Stark, J., Callard, R., Hubank, M.: From the Top Down: Towards a Predictive Biology of Signalling Networks. Trends Biotechnol. 21(7), 290–293 (2003)

19. Chou, K.C.: Review: Structural Bioinformatics and its Impact to Biomedical Science. Current Medicinal Chemistry 11, 2105–2134 (2004)

20. Dfichting, W., et al.: Cancer: A Challenge for Control Theory and Computer Modeling. European Journal of Cancar 32, 1283–1295 (1996)

21. Diichting, W., Ginsberg, T.: Computer Simulation Applied to Radiation Therapy in Cancer Research. Applied mathematics and computation 74, 191–207 (1996)

Plausible Model of Feedback-Control for Cellular Response Based on Gene Regulatory Networks under Radiotherapy

Jin-Peng Qi, Shi-Huang Shao, and Yi-Zhen Shen

College of Information Sciences and Technology, Donghua University,
Shanghai 201620, P.R. China
qipengkai@dhu.edu.cn

Abstract. In response to genome stresses, cell can trigger its self-defensive mechanism by regulating the vital genes and their complicated signal pathways. To illustrate the cellular response DNA damage under radiotherapy, a plausible feedback-control model of P53 stress response networks is proposed at single cell level. The kinetics of double strand breaks (DSBs) generation and repair, ARF and ATM activation, P53-MDM2 regulation, toxins degradation, as well as ion radiation (IR) dose feedback-control are presented.

Keywords: DNA damage, Cellular Network, Radiotherapy, P53.

1   Introduction

As one of the major tools in fighting against cancer, radiotherapy can induce DNA damage into the abnormal cell and trigger its apoptosis by regulating the vital genes and their complicated signal pathways. P53, one of the pivotal anticancer genes, controls the transcription and translation of series genes by regulating downstream genes, and further induces cell cycle arrest and cell apoptosis [1],[2]. Therefore, the status of P53 within tumor cells has been proposed to be one of the determinant in response to tumor therapies [3],[4]. The outcomes of radiotherapy may depend on the presence of functional P53 protein to induce tumor regression through apoptotic pathways [3]. Thus, P53 status in tumor cell can be considered as a predictor for long-term biochemical control during and after radiotherapy[5],[6]-[10].

Recently, several models have been proposed to explain the damped oscillations of P53 in cell populations [11]-[15]. Based on the existing models and the latest biomedical studies on P53 and radiotherapy [2]-[4],[16], a plausible feedback-control model of P53 stress response networks under radiotherapy is proposed at single cell level. The modules of DSBs generating and their repair, ATM and ARF activation, P53-MDM2 feedback loop, the degradation of toxins including the oncogenes, the mis-repair part of DSBs (F_w) and the mutation of P53 (mP53), are implemented. Especially the part of feedback-control for self-adjusting IR dose is added into our model, due to the plausible reason that this feedback-control mechanism can plausibly bring much fewer F_w and much better outcomes under radiotherapy.

2   Model Implementation

2.1   Model Overview

Shown in Fig.1 is the integrated model scheme of P53 stress response networks under radiotherapy; it is composed of four parts, including DSBs generation and repair, ARF and ATM activation, P53-MDM2 feedback regulation, toxins degradation and feedback-control to IR dose.
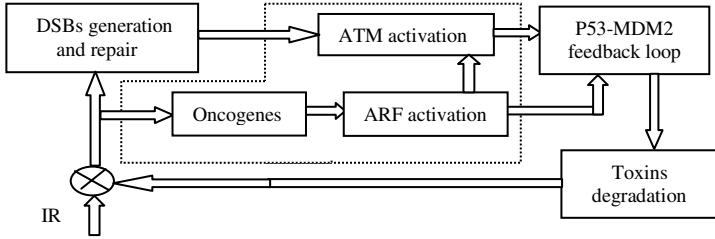

Fig. 1. The integrated model scheme of P53 stress response networks under radiotherapy. It is composed of four parts, including DSBs generation and repair, ATM and ARF activation, P53-MDM2 feedback loop, as well as toxins degradation and feedback-control to IR dose.

As acute IR is applied into a cell, the resulting DSBs occur and form the DSB–protein complexes (DSBCs) at damage sites after interacting with the DNA repair proteins [1]. ATM, a sensor of DNA damage, is activated by DSBCs transferring from DSBs [2],[3]. Meanwhile, oncogenes are prompted and further trigger the activation of ARF. Under the cooperation of active ATM (ATM*) and active ARF (ARF*), P53-MDM2 feedback loop switches on or off the cellular response mechanisms [2],[7],[11], and further control the toxins degradation [2],[3]. Especially, the numbers of normal P53 and toxins including mP53, F_w, oncogenes are dealt as feedback-control signals to regulate IR dose. Here, we use the superscript * to represent the active state.

2.2   DSBs Generation and Repair

In accordance with the experimental results in [1],[2],[11], we deal that the quantity of the resulting DSBs per time scale is proportional to the number generated by a Poisson random function under continuous IR [11]-[15]. The DSBs generation process is formulated as follows:

$$\frac{d[DT]}{dt} = k_t \times Poissrnd \quad (a_{ir} \times IR) \text{ ,} \tag{1}$$

where [DT] is the concentration of total resulting DSBs induced by IR in both fast and slow repair processes. k_t is the parameter to set the number of DSBs per time scale, and a_{ir} is the parameter to set the number of DSBs per IR dose.

DSBs repair module contains both a fast and a slow kinetics, as shown in Fig.2, each of which is composed of a reversible binding of repair proteins (RP) and DSBs lesions into DSBCs, and an irreversible process from the DSBCs to the fixed DSBs [1]-[3]. In this part, DSBCs are the main signal source to transfer the DNA damage to P53-MDM2 feedback loop by ATM activation [2].
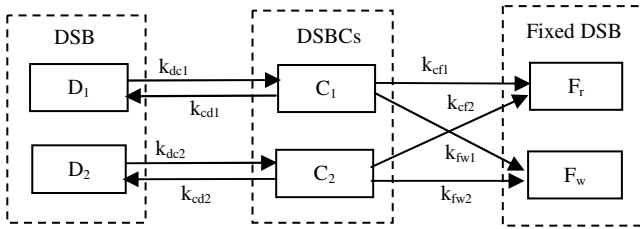
Fig. 2. The module scheme of DNA damage repair process. It includes both a fast repair pathway and a slow one. DSB can be in one of four states: intact DSB (DSB), DBSC, Fr and Fw. Subscripts '1' and '2' refer to the fast kinetics and the slow one.

Due to the fact that F_w has the severe consequences on the subsequent cellular viability and the response capability in fighting against genome stresses [1],[3], we obviously distinguish between correct repair part of DSBs (Fr) and F_w [12]-[14]. Moreover, the total F_w in both repair processes is dealt as a part of toxins remaining within the cell [2],[4],[15]. These toxins can be eliminated by the degradation functions of P53 during and after radiotherapy [2]-[5]. In addition, we deal that DSBCs and RPs are the dynamic variables [11]-[15], which mean that the limited repair proteins are available around the damage sites. The main formulations used in this module are listed as follows:

$$\frac{d[D_1]}{dt} = a_1[D_t] + k_{cd1}[C_1] - [RP](k_{dc1}[D_1] + k_{cross}([D_1] + [D_2])) , \qquad (2)$$

$$\frac{d[D_2]}{dt} = a_2[D_t] + k_{cd2}[C_2] - [RP](k_{dc2}[D_2] + k_{cross}([D_1] + [D_2])) , \qquad (3)$$

$$\frac{d[C_1]}{dt} = k_{dc1}[D_1] - k_{cd1}[C_1] - k_{cf1}[C_1] , \qquad (4)$$

$$\frac{d[C_2]}{dt} = k_{dc2}[D_2] - k_{cd2}[C_2] - k_{cf2}[C_2] , \qquad (5)$$

$$\frac{d[RP]}{dt} = S_{rp} + k_{cd1}[C_1] + k_{cd2}[C_2] - [RP](k_{dc1}[D_1] + k_{dc2}[D_2] + k_{cross}([D_1] + [D_2])), \qquad (6)$$

$$\frac{d[F_w]}{dt} = k_{fw1}[C_1] + k_{fw2}[C_2] , \qquad (7)$$

where $[D]$, $[C]$, $[RP]$, and $[F_w]$ represent the concentrations of DSBs, DSBCs, and F_w in the fast and the slow repair kinetics respectively, k_{dc}, k_{cd}, k_{cf}, and k_{fw} are the transition rates between four states of DSBs. k_{dc}, and k_{cross} are the first-order and second-order rate in both repair kinetics respectively [11],[13]. S_{rp} is the basal induction rate of repair mRNA, and subscripts '1' and '2' refer to the fast kinetics and the slow one.

2.3   ATM and ARF Activation

Shown in Fig.3 is the module scheme of ATM and ARF activation. Compared with the previous studies in [11]-[15], ARF, oncogenes and related regulation pathways are involved in this module [2],[7].
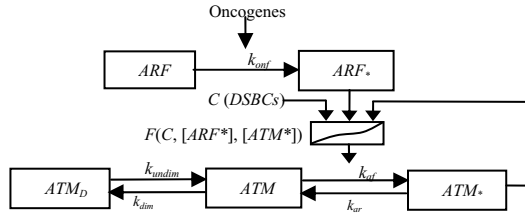
Fig. 3. The module scheme of ATM and ARF activation under continuous IR. ARF is activated by the over-expression of oncogenes, and ATM is activated from ATM monomers under the cooperating effects of DSBCs, ARF*, and self-feedback of ATM*.

Under continuous IR, DSBCs prompt the phosphorylation of inactive ATM monomers [1],[2],[15]. Meanwhile, ARF, another tumor suppressor, is activated by hyperproliferate signals emanating from oncogenes, and further prompts the ATM activation [2],[7],[14]. ATM is activated rapidly with the cooperating effects of DSBCs, ARF* and the positive feedback from ATM* account for intermolecular autophosphorylation [1],[2],[11]. The main formulations used in this module are as follows:

$$\frac{d[ATM_d]}{dt} = \frac{1}{2}k_{dim}[ATM_m]^2 - k_{undim}[ATM_d] \,, \tag{8}$$

$$\frac{d[ATM_m]}{dt} = 2k_{undim}[ATM_d] - k_{dim}[ATM_m]^2 - k_{af}f[ATM_m] + k_{ar}[ATM*] \,, \tag{9}$$

$$\frac{d[ATM*]}{dt} = k_{af}f[ATM_m] - k_{ar}[ATM*] \,, \tag{10}$$

$$\frac{d[ARF]}{dt} = S_{arf} - k_{ad}[ARF] - k_{onf}[Onco][ARF] \,, \tag{11}$$

$$\frac{d[ARF*]}{dt} = k_{onf}[Onco][ARF] - k_{pad}[ARF*] \,, \tag{12}$$

$$f(C,[ARF*],[ATM*]) = a_1C + a_2[ATM*] + a_3C[ATM*] + a_4[ARF*] \,, \tag{13}$$

where $[ATM_d]$, $[ATM]$ and $[ATM*]$ represent the concentrations of ATM dimer, ATM monomer, and active ATM respectively; $[Onco]$, $[ARF]$ and $[ARF*]$ represent the concentrations of oncogenes, ARF, and active ARF respectively; k_{undim}, and k_{dim} are the ATM undimerization and dimerization rates; k_{ar}, and k_{af} are the ATM monomer inactivation and activation rates, respectively. In addition, f is the function of ATM activation, the term a_1C implies the fact that DSBs somehow activate ATM molecules at a distance, $a_2[ATM*]$ indicates the mechanism of autophosphorylation of ATM, $a_3C[ATM*]$ represents the interaction between the DSBCs and ATM* [11]-[15], and $a_4[ARF*]$ represents the interaction between the ARF* and ATM* [1],[3],[7].

2.4   P53-MDM2 Feedback Loop

P53-MDM2 feedback loop is the core module in the integrated networks [11]-[15]. Recent studies show that both ARF* and ATM* are required for the efficient induction of apoptosis in response to genome and oncogenes stresses [2],[7].
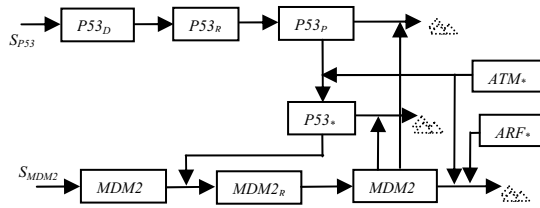
Fig. 4. The diagram of P53-MDM2 feedback loop under radiotherapy. P53 is translated from p53mRNA and phosphorylated by ATM* and ARF*. Mdm2 protein promotes a fast degradation of P53 protein and a slow degradation of P53*. In addition, ATM* and ARF* stimulate the degradation of Mdm2, and then indirectly increase the regulatory activation of P53* further.

As shown in Fig.4, we deal that DSBCs are the main signal to the downstream P53-MDM2 feedback loop through ATM activation. Meanwhile, ARF* can further increase the regulating function of P53 by inhibiting the expression of MDM2 and preventing P53 degradation [2],[7],[14]. With the cooperation of both ATM* and ARF*, this module can produce oscillations in response to sufficiently strong IR dose [2],[7],[11]-[15]. The main formulations are as follows:

$$\frac{d[P53_R]}{dt} = S_{P53} - d_{rp}[P53_R] - k_{rp}[P53_R] \tag{14}$$

$$\frac{d[P53_P]}{dt} = k_{rp}[P53_R] + k_{p*p}[P53*] - d_{pp}[P53_P]$$

$$- k_{app*}[ATM*]\frac{[P53_P]}{[P53_P]+k_p} - k_{mp}[MDM2_P]\frac{[P53_P]}{[P53_P]+k_d} \tag{15}$$

$$\frac{d[P53*]}{dt} = k_{app*}[ATM*]\frac{[P53_P]}{[P53_P]+k_p} - k_{p*p}[P53*] - d_{pp*}[P53*]$$

$$- k_{p*p}[P53*] - d_{pp*}[P53*] - k_{mp*}[MDM2_P]\frac{[P53*]}{[P53*]+k_{d*}} \tag{16}$$

$$\frac{d[MDM2_R]}{dt} = S_{mdm2} + k_{p*m}\frac{[P53*]^n}{[P53*]^n + k^n}$$

$$- k_{mrp}[MDM2_R] - d_{mr}[MDM2_R] \tag{17}$$

$$\frac{d[MDM2_P]}{dt} = k_{mrp}[MDM2_R] - d_{mp}[MDM2_P]$$

$$- (k_{mat}\frac{[ATM*]}{[ATM*]+k_{at}} + k_{mar}\frac{[ARF*]}{[ARF*]+k_{ar}})[MDM2_P] \tag{18}$$

where $[P53_R]$, $[P53_P]$, $[P53*]$, $[MDM2_R]$, and $[MDM2_P]$ represent the concentrations of P53 mRNA, P53 protein, active P53, MDM2 mRNA, and MDM2 protein, respectively. S_{P53}, and S_{MDM2} is the basal induction rate of P53 mRNA, and MDM2 mRNA. k represent the regulation rates among genes and proteins, and d represent the degradation rates of mRNAs and proteins.

2.5   Toxins Degradation and IR Dose Feedback-Control

Shown in Fig.5 is the module scheme of toxins degradation and IR dose feedback-control, the mutation of P53 (mP53) from P53R is triggered by oncogenes, and F_w, mP53, and oncogenes are degraded directly by the regulating functions of P53* [3],[4]. In addition, mP53 is dealt as another outcomes indictor of radiotherapy, besides F_w, and P53* [5],[7],[11]-[15]. Especially, the remaining number of P53*, oncogenes, mP53, and F_w are dealt as feedback-control factors to regulate the strength of IR dose. This mechanism can plausibly take much fewer F_w and much better outcomes under radiotherapy. The main formulations in this module are as follows:

$$\frac{d[Onco]}{dt} = k_{oniR}[Onco][IR] - k_{onp}[Onco][P53*] \, , \tag{19}$$

$$\frac{d[F_w]}{dt} = k_{tfw}[F_w] - k_{pt}[P53*][F_w] \, , \tag{20}$$

$$\frac{d[mP53]}{dt} = k_{mp}[P53_R][Onco] - k_{pmd}[P53_{P*}][mP53] \, , \tag{21}$$

$$\frac{d[IR]}{dt} = k_{oncoIR}[IR][Onco] + k_{mp53IR}[IR][mP53] - k_{FwIR}[IR][F_w] - k_{p53IR}[P53*][IR] \, , \tag{22}$$

where [$Onco$], [F_w] and [$mP53$] represent the concentrations of oncogenes, F_w and mP53, respectively. [IR] represent the strength of IR dose. k represents the regulation rates among genes and proteins.
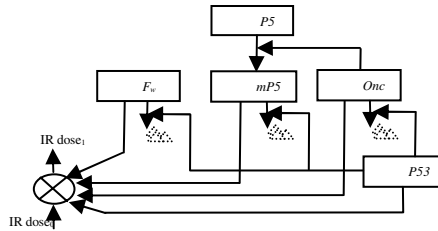
Fig. 5. The diagram of toxins degradation and feedback-control for IR dose. Oncogenes, F_w and mP53 are degraded directly by the regulatory functions of P53*, and IR dose is further regulated by F_w, oncogenes, mP53, and F_w .

3   Simulating Results

By using simulation platform in MATLAB7.0, we implement the dynamic kinetics of P53 stress response networks under continuous IR, illustrate the cellular self-defense mechanisms in fighting against genome stresses, and accomplish the feedback-control to IR dose for taking much better outcomes of radiotherapy.

3.1   Kinetics of DSBs Generation and Their Repair

In our simulations, 5Gy IR is applied into a cell initially. Under the combined feedback-control, as shown in Fig6a, IR dose increases from 5Gy, and then begins to

decrease after reaching the first climax, and the trace of resulting DSBs generation shown in Fig6b has the similar dynamic trend with Fig6a. Meanwhile, as shown in Fig.6c, RP available around damage sites keep decreasing and the rate of DSBCs synthesis begins to consequently decrease. In addition, due to the limited RP available, as shown in Fig.6d, the intact DSBs dramatically increase, and then begin to decrease after the first climax following the decreasing IR dose. Meanwhile, F_w is accumulated within the cell fast without the eliminating functions of P53*.
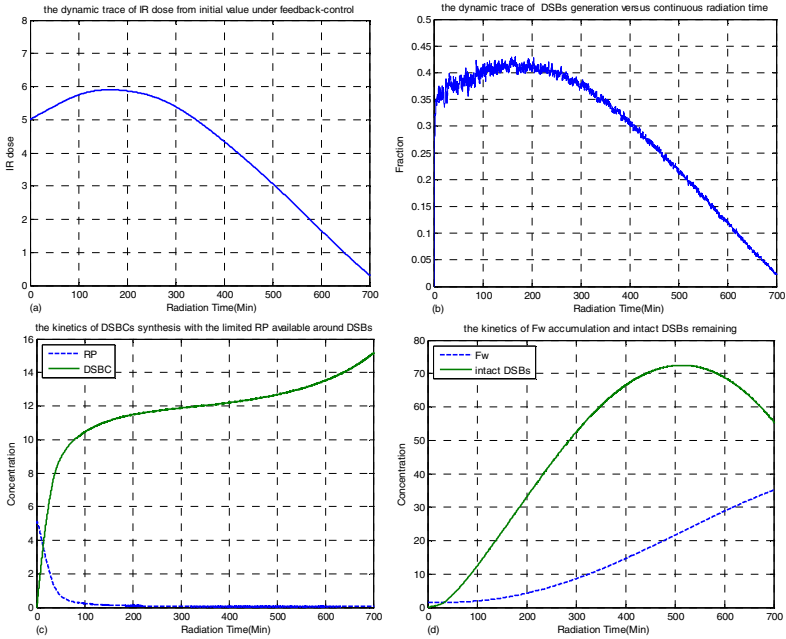
Fig. 6. The kinetics of DSBs generation and repair process under decreasing IR from initial 5Gy. (a) the decreasing trace of IR dose with feedback-control. (b) the dynamic trace of resulting DSBs generation. (c) the kinetics of DSBCs synthesis with the decreasing RP. (d) the kinetics of F_w accumulation against intact DSBs.

3.2 Kinetics of ARF and ATM Activation

Shown in Fig.7a is the kinetics of ATM activation, with the cooperation of DSBCs, ARF* and the positive self-feedback of ATM*. ATM is activated rapidly and switches to "on" state. The step-like trace suggests that the ATM module can produce an on-off switching signal, and transfer the damage signal to the P53-MDM2 feedback loop further [3],[11]. Meanwhile, the kinetics of ARF activation is plotted in Fig.7a, without depressing functions of P53*. ARF is activated fast by over-expression of oncogenes. Under the effects of both ATM and ARF activation, DNA damage signals can be transferred to the P53-MDM2 feedback loop efficiently [2],[7].
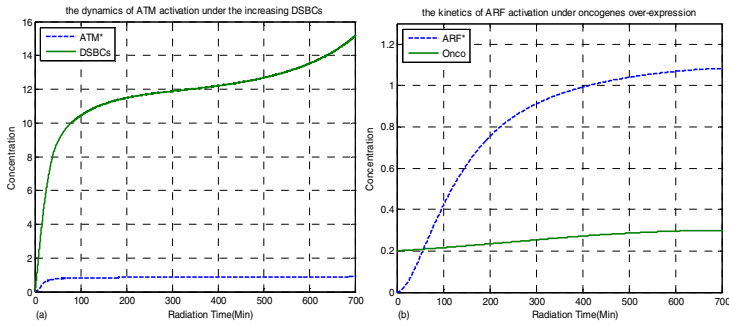
Fig. 7. The kinetics of ARF and ATM activation under decreasing IR dose. (a) the switch-like kinetics of ATM activation, ATM* switch to "on" state versus radiation time. (b) the kinetics of ARF activation in response to over-expression of oncogenes.
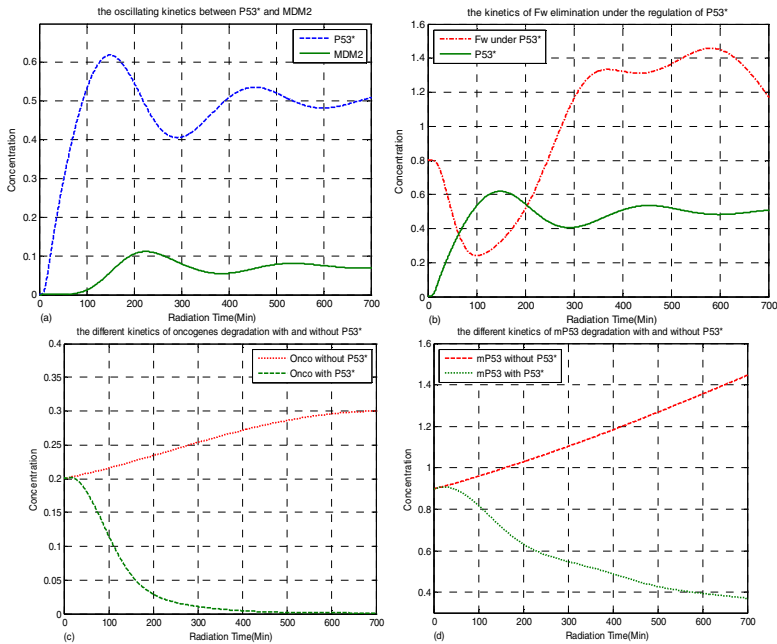

Fig. 8. The kinetics of cellular response versus continuous radiation time. (a) the oscillating kinetics of P53* and MDM2 under the cooperative effects of ATM* and ARF*. (b) the kinetics of F_w elimination with the degradation of P53*. (c) the depressing dynamics of oncogenes with the regulations of P53*. (d) the kinetics of mP53 elimination under the degradation function of P53*.

3.3   Kinetics of Txins Dgradation

In our simulations, we deal that toxins, including F_w, oncogenes, and mP53, can be directly degraded by regulation functions of P53*. Especially the remaining quantities of these toxins within the cell are dealt as the feedback-control factors to regulate the

strength of IR dose. This mechanism can plausibly decrease the number of Fw efficiently, and take much better outcomes of radiotherapy. With the cooperation of ATM* and ARF*, as shown in Fig.8a, P53 and MDM2 generate oscillations versus radiation time. The period of these oscillations is about 350min, and their phase difference is about 100min. Moreover, the first pulse is slightly higher than the second, and the oscillating kinetics trends to a new equilibrium, all of which are quite similar with the experimental observations [2],[7] and the simulation results [11]-[15]. Shown in Fig.8b is the kinetics of F_w eliminating with the functions of P53*. Compared with Fig6d, the remaining F_w shown in Fig8b keeps decreasing from initial value, and then increases with some oscillations. It suggests that the capabilities of cellular damage repair and toxins degradation are limited within some threshold.

Meanwhile, the kinetics of oncogenes degradations with and without P53* is plotted in Fig.8c. The expression levels of oncogenes keep decreasing and trend to Zero versus radiation time under the negative regulation of P53*. In addition, shown in Fig.8d is the kinetics of mP53 elimination with and without P53*. Similarly with Fig.8c, mP53 keeps decreasing with slight oscillations after the first climax. These simulations above suggest that P53* indeed act important roles in fighting against genome damage, and much better outcomes of radiotherapy might be brought with fewer byproduct under the mechanisms of IR dose feedback-control.

4   Conclusion

By using a set of differential equations, as well as graphic methods, we proposed a feedback-control model for cellular response DNA damage under radiotherapy. In our simulations, the resulting DSBs are generated and repaired, ATM and ARF exhibit a strong sensitivity behavior in response to genome stresses, and P53-MDM2 feedback loop exhibit oscillating kinetics in fighting against continuous IR. Under degradation functions of P53*, the toxins including F_w, oncogenes, and mP53 are eliminated, these toxins are further dealt as the feedback-control factors to regulate IR dose in order to get much better outcomes of radiotherapy with much fewer byproducts. Our mode, although simple, provides a theoretical framework to more effectively illustrate and analyze the cellular defensive mechanisms in response to genome stresses under radiotherapy.

Acknowledgment

References

1. Li, L., Story, M., Legerski, R.: Cellular Responses to Ionizing Radiation Damage. Int. J. Radiat. Oncol. Biol. Phys. 49, 1157–1162 (2001)
2. Kohn, K.W., Pommier, Y.: Molecular Interaction Map of the P53 and MDM2 Logic Elements, Which Control the Off-On Switch of P53 in Response to DNA Damage. Biochemical and Biophysical Research Communications 331, 816–827 (2005)

3. Perez, C., Brady, L.: Principles and Practice of Radiation Oncology, pp. 784–785. Lippincott-Raven, Philadelphia (1998)
4. Tjebbes, G.W.A., Kreijveldb, P.A., et al.: P53 Tumor Suppressor Gene Mutations in Laryngeal Cancer and in Recurrent Disease Following Radiation Therapy. Oral Oncology 38, 296–300 (2002)
5. Dfichting, W., UlmeP, W., Ginsberg, T.: Cancer: A Challenge for Control Theory and Computer Modeling. European Journal of Cancer 32, 1283–1295 (1996)
6. Ritter, M.A., Gilchrist, K.W., et al.: The Role of P53 in Radiation Therapy Outcomes for Favorable-to-Intermediate-Risk Prostate Cancer T. Radiation Oncology Biol. Phys. 53, 574–580 (2002)
7. Lindstrom, M.S., Wiman, K.G.: Myc and E2F1 Induce P53 Through p14ARF-Independent Mechanisms in Human Fibroblasts. Oncogene 22, 4993–5005 (2003)
8. Weller, M.: Predicting Response to Cancer Chemotherapy: The Role of P53. Cell Tissue Res. 292, 435–445 (1998)
9. Werner, D., Thomas, G.: Computer Simulation Applied to Radiation Therapy in Cancer Research. Applied Mathematics and Computation 74, 191–207 (1996)
10. Chou, K.C.: Review: Structural Bioinformatics and its Impact to Biomedical Science. Current Medicinal Chemistry 11, 2105–2134 (2004)
11. Qi, J.P., Shao, S.H., et al.: Modeling and Simulation of P53 Gene Regulatory Networks under DNA Damage. Journal of System Simulation 19, 3321–3323 (2007)
12. Qi, J.P., Shao, S.H., Zhu, Y.: Mathematical Modeling of P53 Gene Regulatory Network. Dynamics of Continuous Discrete and Impulsive Systems-Series B-Applications & Algorithms 13, 11–15 (2006)
13. Qi, J.P., Shao, S.H., et al.: A Dynamic Model for the P53 Stress Response Networks under Ion Radiation. Amino Acids 33, 75–83 (2007)
14. Qi, J.P., Shao, S.H., et al.: A Mathematical Model of P53 Gene Regulatory Networks under Radiotherapy. Biosystems 90(3), 698–706 (2007)
15. Lan, M., John, W., et al.: A Plausible Model for the Digital Response of P53 to DNA Damage. PNAS 2(40), 14266–14271 (2005)
16. Siim, P., Arnold, K., Toivo, M., et al.: ARF and ATM/ATR Cooperate in P53-Mediated Apoptosis upon Oncogenic Stress. Biochemical and Biophysical Research Communications 334, 386–394 (2005)
17. Chou, K.C., Zhou, G.P.: Role of the Protein Outside Active Site on the Diffusion-Controlled Reaction of Enzyme. Journal of American Chemical Society 104, 1409–1413 (1982)

A Numerical Simulation Study of the Dependence of Insulin Sensitivity Index on Parameters of Insulin Kinetics

Lin Li and Wenxin Zheng

School of Biomedical Engineering, Capital Medical University, Beijing, 10069
`lil@ccmu.edu.cn`

Abstract. Objective: To investigate the dependence of insulin sensitivity index (SI) and glucose effectiveness (SG) on the parameters of insulin kinetics in Bergman's minimal model. Methods: The concentration values of glucose and insulin at series of time were computed, where the parameters in the model were generated randomly in a certain range. Based on the optimization method, insulin sensitivity index and glucose effectiveness were estimated. Results: The largest relative error of SI was less than 1.66‰ and the largest relative error of SG was less than 0.96‰, if a data set of 240 points of concentration value was used in computation. However, if there are only 27 points (according to intra venous glucose tolerance test) in the data set, the relative error of SI ranged from 0.45% to 48.1%, and the relative error of SG was less than 5.6%. Conclusion: The dependence of SI and SG upon parameters of insulin kinetics is not significant.

1 Introduction

The ability to maintain a relatively constant blood glucose concentration is an essential feature for the life of higher vertebrates. Although some cells in higher animals can utilize alternative energy sources, such as amino acids or fatty acids, the brain relies upon a continuous supply of glucose delivered by the blood circulation. The hormone mainly responsible for the maintenance of blood glucose levels is insulin, secreted by the endocrine pancreas. The main function of insulin is to stimulate anabolic reactions for carbohydrates, proteins and fats in tissues, as well as inhibiting glycogenolysis and gluconeogenesis in the liver, all of which lead to a reduction in blood glucose levels. If one's glucose concentration level is constantly out of the range (70–110 mg/dl), this person is considered to have blood glucose problems known as hyperglycemia or hypoglycemia. Diabetes mellitus is a disease of the glucose-insulin regulatory system [1,2], which is referred to as hyperglycemia. The dynamic relationship between glucose and insulin in endocrine and metabolic systems has been mathematically modeled and studied by many researchers ([2-9], and the references cited in [1]).

The most widely used model in physiological research on the metabolism of glucose is the so-called "minimal model", which describes intra venous glucose tolerance

test (IVGTT) experimental data well with the smallest set of identifiable and meaningful parameters. The minimal model enables estimation of two key indices of glucose/insulin dynamics: insulin sensitivity (S_I) and glucose effectiveness (S_G). The minimal model takes the form of[6]

$$\frac{dG(t)}{dt} = -[P_1 + X(t)]G(t) + P_1 G_b, G(0) = P_0, \tag{1}$$

$$\frac{dX(t)}{dt} = -P_2 X(t) + P_3 (I(t) - I_b), X(0) = 0, \tag{2}$$

$$\frac{dI(t)}{dt} = P_4 R(t)t - P_6 (I(t) - I_b), I(0) = P_7 + I_b, \tag{3}$$

where[6] $R(t)$ is $G(t) - P_5$, if $G(t) > P_5$ and 0 otherwise. The $G(t)$ [mg/dl],$I(t)$ [μUI/ml] is the plasma glucose, insulin concentration at time t [min], respectively. $X(t)$ is an auxiliary function representing insulin excitable tissue glucose uptake activity, roughly proportional to insulin concentration in a "distant" compartment. G_b [mg/dl], I_b [μUI/ml] is the subject's baseline glycemia, insulinemia, respectively. The[6] P_1 is the glucose "mass action" rate constant, i.e. the insulin-independent rate constant of tissue glucose uptake; P_2 is the rate constant expressing the spontaneous decrease of tissue glucose uptake ability; P_3 is the insulin-dependent increase in tissue glucose uptake ability, per unit of insulin concentration excess over baseline insulin; P_4 is the rate of pancreatic release of insulin after the bolus, per minute and per mg/dl of glucose concentration above the "target" glycemia; P_5 is the pancreatic "target glycemia" is the first order decay rate constant for insulin in plasma; Refer to De Gaetano and Arino [6] for the meanings of the other parameters. Eq. (3) describes the insulin kinetics, so the parameters (P_4, P_5, and P_6) in it are called parameters of insulin kinetics, in this paper.

The parameters of the minimal model are determined using data from either a standard or modified IVGTT. In the standard IVGTT a bolus of glucose (300 mg/kg) is intravenously injected, and blood samples are collected over the 3 h following the glucose injection. The insulin sensitivity (S_I) and glucose effectiveness (S_G) are determined by

$$S_I = \frac{P_3}{P_2}, \qquad S_G = P_1. \tag{4}$$

The minimal model has minimal number of parameters and has been widely used in physiological research works. In the applied use of the minimal model, the most important piece of information is the S_I index. However, [6] argues that it has the following drawbacks. In this model, the fitting process of the parameters is divided into two separate stages: first, the recorded insulin concentration is used as given input data in order to derive the parameters in Eq. (1)-(2); second, the recorded glucose concentration is used as given input to derive the parameters of the third equation. However, the glucose-insulin system is an integrated physiologic dynamical system and it should be described as a whole with a single-step parameter fitting process.

In fact, if S_I and S_G vary slightly within a range of variations of insulin kinetics parameters (P_4, P_5, P_6), it indicated the two-separate-stage method is reliable. On the contrary, if S_I and S_G have great changes in a range of changes of insulin kinetics parameters (P_4, P_5, P_6), then the estimation above should be unbelievable. The aim of this paper is to investigate the dependence of insulin sensitivity index (S_I) on parameters of insulin kinetics (P_4, P_5, P_6) by numerical simulation, in which a solution of Eq.(1)-(3) was regarded as "the recorded data" in a IVGTT. By randomly taking values of P_4, P_5, P_6, we shown that the dependence of S_I and S_G upon parameters of insulin kinetics is not significant.

2 Methods

In accordance with the range of parameter P_4, P_5, P_6 in [4,5], we let

$$P_4 \in (0.01, 0.07), P_5 \in (90, 140), P_6 \in (0.2, 0.6), \tag{5}$$

and fix the other parameters in Eq.(1)-(3),

$$G_b = 84, I_b = 10, P_0 = 200, P_7 = 82, P_2 = 0.03, P_3 = 2.0 \times 10^{-5}, P_1 \in (0,1), \tag{6}$$

We got the numerical solutions of Eq.(1)-(3) by applying the Runge-Kutta's method, $G_j = G(t_j), I_j = I(t_j), (j = 1, 2, \cdots, n)$, which can be regarded as "experiment data". A linear interpolation can be used to get the function $G(t)$ and $I(t)$. Substituting them into Eq.(1)-(2), we estimated the parameters P_1, P_2, and P_3. Therefore, it is possible to study the dependence of insulin sensitivity index and glucose effectiveness on parameters of insulin kinetics, P_4, P_5, P_6.

The following method was utilized to get the estimations of parameter P_1, P_2, P_3. By Eq. (2), we get

$$X(t) = P_3 \int_0^t e^{-P_2(t-s)}(I(s) - I_b) ds \tag{7}$$

Substitute it into Eq. (1), which gives

$$\frac{dG(t)}{dt} = P_1(G_b - G(t)) - P_3 G(t) \int_0^t e^{-P_2(t-s)}(I(s) - I_b) ds \tag{8}$$

Let $t = t_j$ and $G(t_j) = G_j$, we have

$$\begin{aligned}
\frac{dG(t)}{dt}\Big|_{t=t_j} &= P_1(G_b - G_j) - P_3 G_j \sum_{k=1}^{j-1} \int_{t_k}^{t_{k+1}} e^{-P_2(t_j-s)}(I(s) - I_b) ds \\
&\approx P_1(G_b - G_j) - P_3 G_j \frac{e^{-P_2 t_j}}{P_2} \sum_{k=1}^{j-1} [I(\frac{t_{k+1}+t_k}{2}) - I_b](e^{P_2 t_{k+1}} - e^{P_2 t_k})
\end{aligned} \tag{9}$$

and rewrite it as the following form

$$y_j = P_1 A_j + P_3 B_j \tag{10}$$

where

$$y_j = \frac{dG(t)}{dt}\Big|_{t=t_j}, \qquad A_j = (G_b - G_j),$$

$$B_j = -e^{-P_2 t_j}\frac{G_j}{P_2}\sum_{k=1}^{j-1}[I(\frac{t_{k+1}+t_k}{2}) - I_b](e^{P_2 t_{k+1}} - e^{P_2 t_k}) \tag{11}$$

Consequently, If P_2 is given, then each (A_j, B_j, y_j) can be known by using interpolation function $G(t)$ and $I(t)$. We applied linear fitting to obtain the estimation values $\hat{P}_1, \hat{P}_3$ of P_1, P_3. In a range of P_2 the simple iterate algorithm was implemented to locate the optimal value $\hat{P}_2$, which was implemented two times. For the first time, the step size was set to 0.005 and an optimal value ($\bar{P}_2$) was found. For the second time, restricted in the area around $\bar{P}_2$, the one-dimensional search was done with step size of 0.0001, and a more optimal value $\hat{P}_2$ was found. Then $\hat{P}_1$ and $\hat{P}_3$ were gotten.

3 Results

Randomly given 7 different P_1 (Table 1), with 100 sets of P_4, P_5, P_6 generated randomly by computer, we obtained "experimental data" $G_j = G(t_j), I_j = I(t_j), \ (j=1,2,\cdots,n)$ (n=240, and n=27 respectively). The relative errors of S_G, P_2, P_3, S_I were calculated, where the relative error of P_2 is defined as $|\hat{P}_2 - P_2|/P_2$. The means and standard deviations of 100 times computations were listed in Table 1 (n=240). We can see that the overall relative error is relatively small. The largest mean value of relative errors of S_G, P_2, P_3, S_I are $9.63\times10^{-4}, 3.33\times10^{-3}, 1.72\times10^{-3}, 1.66\times10^{-3}$, respectively in case of n=240.

Table 1. The relative errors of each parameter and index in cases of n=240 (mean ± s.d)

P_1	S_G (10^{-6})	P_2 (10^{-3})	P_3 (10^{-3})	S_I (10^{-3})
0.910602	13.589 ± 0.337	3.333 ± 0.000	1.678 ± 0.081	1.661 ± 0.081
0.843094	9.355 ± 0.431	3.333 ± 0.000	1.694 ± 0.079	1.645 ± 0.079
0.709755	2.512±0.603	3.333 ± 0.000	1.694 ± 0.064	1.645 ± 0.064
0.628754	1.206 ± 0.824	3.333 ± 0.000	1.709 ± 0.059	1.630 ± 0.060
0.470172	8.492 ± 1.656	3.333 ± 0.000	1.724 ± 0.053	1.615 ± 0.053
0.229834	32.586 ± 5.771	3.333 ± 0.000	1.789 ± 0.014	1.550 ± 0.014
0.024098	963.718±31.523	3.333 ± 0.000	2.126 ± 0.341	1.212 ± 0.343

According to the standard IVGTT, we took $G_j, I_j \ (j=1,\cdots,27)$ corresponding to getting blood samples at $0', 1', 2', 3', 4', 5', 6', 7', 8', 9', 10', 11', 12', 15', 20', 25', 30', 35', 40', 50', 60', 80', 100', 120', 140', 160', 180'$. The means and standard deviations of 100 times computations were listed in Table 2. The largest mean value of relative errors of S_G, P_2, P_3, S_I are $5.6\times10^{-2}, 5.1, 2.6, 0.48$.

Table 2. The relative errors of each parameter and index in cases of n=27 (mean ± s.d)

P_1	S_G (10^{-3})	P_2 (10^{-1})	P_3 (10^{-1})	S_I (10^{-1})
0.910602	56.040 ± 0.435	50.867 ± 0.000	26.253 ± 7.067	4.044±1.161
0.843094	46.727 ± 0.455	50.867 ± 0.000	23.306 ± 6.519	4.528±1.071
0.709755	30.526 ± 0.839	42.931 ± 17.036	15.479 ± 6.842	4.813 ± 1.298
0.628754	22.317 ± 1.605	34.615 ± 21.815	11.237 ± 6.787	0.436 ± 0.185
0.470172	9.407 ± 0.213	2.660 ± 1.480	1.734 ± 0.819	0.683 ± 0.421
0.229834	1.088 ± 0.083	0.580 ± 2.436	0.218 ± 0.031	0.855 ± 0.314
0.024098	1.717 ± 0.770	0.064 ± 0.065	0.0188 ± 0.019	0.045 ± 0.055

4   Discussion and Conclusion

In order to investigate the dependence of S_I and S_G upon parameters of insulin kinetics P_4, P_5, P_6, the numerical stimulation was implemented, in which the solutions of Eq.(1)-(3) are used as the "experimental data".

In Table 1, we can see that for different P_1, the relative errors of P_2, P_3, S_I change slightly within a small order of magnitude. Although the amplitude of changes of S_G is bigger than those of the others, the values of the relative errors are small enough. It is noted that the results in Table 1 were obtained by taking values of P_4, P_5, P_6 generated randomly by computer for 100 times, therefore the results indicated that the dependence of S_I and S_G upon parameters P_4, P_5, P_6 was not significant if a sufficiently large number of experimental data (here we took n=240) was used to estimate the parameters P_1, P_2, P_3. In fact, $G(t)$ and $I(t)$ are determined by Eq.(1)-(3), which means that they must be different for different values of parameters P_4, P_5, P_6. However this in the normal (physiological) ranges of these parameters S_I and S_G are almost unchanging. Therefore, S_I and S_G mainly depended upon the other parameters, and it is reliable that to estimate S_I and S_G by Eq.(1)-(2).

In the case of n=27, corresponding to the number of blood sample in the standard IVGTT, the results were shown in Table 2. The relative errors of each argument varied from small to large along with the changes of P_1. But the relative errors here are too large to be acceptable. It may be explained as the location of t_j is inappropriate, which means that the time points and the frequency to get blood samples in IVGTT may be reconsidered.

There are several limitations of current investigation. One assumption is that the minimal model (Eq.(1)-(3)) is the robust model to govern the glucose-insulin system. However, some of the mathematical results produced by this model are not realistic[4]. The main reason is that Eq.(3) is a non-automatic, i.e., the right hand of it involves t explicitly, which results in unboundedness of the solution $X(t)$. However, in the view of computation, Eq.(3) is the best one chosen by Toffolo [4] from a series of proposed models, according to their features, such as, the identifiability, plausibility of parameters and statistical characteristics of error. Therefore the assumption is reasonable in this study.

Another limitation of current study is that the "experimental data" did not come from physiological test. Some real experimental data need to get for further verifying the results. Additionally, from statistical point of view, the number of the computation times is not big enough. However It can be shown that there is no significant difference in the numbers of the computation times, for example, P_1=0.470172 (Table 3).

Table 3. The relative errors in different computation times as P_1=0.470172 (m.± s.d)

Times	S_G (10^{-6})	P_2 (10^{-3})	P_3 (10^{-3})	S_I (10^{-3})
40	8.44 ± 1.46	3.33 ± 0.00	1.72 ± 0.05	1.62±0.05
54	8.41 ± 1.71	3.33 ± 0.00	1.72 ±0.05	1.62±0.05
73	8.52 ± 1.67	3.33 ± 0.00	1.72 ±0.05	1.61 ± 0.05
100	8.49 ± 1.66	3.33 ± 0.00	1.72 ± 0.05	1.62 ± 0.05

In conclusion, the dependence of S_I and S_G upon parameters of insulin kinetics is not significant, and which implies that the two-separate-stage method to estimate S_I and S_G is reliable. The frequency and time points of blood samples collection in IVGTT need to be studied further.

Acknowledgement

References

1. Makroglou, A., Li, J., Kuang, Y.: Mathematical Models and Soft-ware Tools for the Glucose-insulin Regulatory System and Diabetes: an Overview. Applied Numerical Mathematics 56, 559–573 (2006)
2. Bergman, R.N., Finegood, D.T., Kahn, S.E.: The Evolution of Beta-cell Dysfunction and Insulin Resistance in Type 2 Diabetes. Eur. J. Clin. Invest. 32(Suppl. 3), 35–45 (2002)
3. Bergman, R.N., Ider, Y.Z., Bowden, C.R., et al.: Quantitative Estimation of Insulin Sensitivity. Am. J. Physiology 236, E667–E677 (1979)
4. Toffolo, G., Bergman, R.N., Finegood, D.T., et al.: Quantitative Estimation of Beta Cell Sensitivity to Glucose in the Intact Organism. Diabetes 29, 979–990 (1980)
5. Boston, R.C., Stefanovski, D., Moate, P.J., et al.: MINMOD Millennium: a Computer Program to Calculate Glucose Effectiveness and Insulin Sensitivity from the Frequently Sampled Intravenous Glucose Tolerance Test. Diabetes Technol. Ther. 5, 1003–1015 (2003)
6. De Gaetano, A., Arino, O.: Mathematical Modelling of the Intravenous Glucose Tolerance Test. J. Math. Biol. 40, 136–168 (2000)
7. Li, J., Kuang, Y., Li, B.: Analysis of IVGTT Glucose Insulin Interaction Models with Time Delay. Discrete Contin. Dyn. Syst. Ser. B 1, 103–124 (2001)
8. Palumbo, P., Panunzi, S., De Gaetano, A.: Qualitative Behavior of a Family of Delay-Differential Models of the Glucose-insulin System. Discrete Contin. Dyn. Syst. Ser. B 7, 399–424 (2007)
9. Vu Giang, D., Lenbury, Y., De Gaetano, A., et al.: Delay Model of Glucose-insulin Systems: Global Stability and Oscillated Solutions Conditional on Delays. J. Math. Anal. Appl. 343, 996–1006 (2008)

Design of the Performance Evaluation Library for Speech Recognition Systems Based on SystemC

Jin-wei Liu[*], Si-jia Huo[*], Zhang-qin Huang[*], Yi-bin Hou[*], and Jin-jia Wang[*]

College of Computer Science, Beijing University of Technology, Beijing 100124, China
`liujw5618@163.com, huosijia@emails.bjut.edu.cn`

Abstract. In the process of embedded systems modeling and performance evaluation, the performance indicators, such as average delay and CPU occupancy rate, are often need to be analyzed and evaluated. In this paper, the method for performance evaluation has been proposed, and the library function for performance evaluation based on SystemC language has been designed. The library function is designed for evaluating performance parameters, calculating expectations and confidence interval, and terminating the simulation when the performance indicators meet the accuracy requirements. In this paper, the performance evaluation of embedded speech recognition systems by using the library function and the method proposed was introduced in detail. For validating the correctness and practicality of the library function further, the library function has been also used in the performance evaluation of wireless AP systems, as well as in performance evaluation of complex embedded systems based on SystemC. It is of great value for extending the library of SystemC.

Keywords: SystemC; System-level design; Embedded system; Performance evaluation; Library function.

1 Introduction

System-level design has become important increasingly because of the increase of the design complexity and the curtail of the design cycle of embedded systems. By modeling the system abstractly before the hardware and software of the system have been achieved, the qualitative and quantitative attributes of the system have been analyzed. Based on the performance evaluation results, the scheme which designed for the system would be estimated when the performance meets the requirement, and the repeating of the design has been avoided, because the design scheme has been improved in the initial stages of design[1] [2].

---

[*] Jin-wei Liu, (1972－), male, study on the software/hardware co-designing, system-level modeling methods, and network. Si-jia Huo, (1985-), male, study on embedded systems, system-level modeling methods, and network. Zhang-qin Huang, (1965－), male, study on the software/hardware co-designing, system-level modeling methods, and the network. Yi-bin Hou, (1952－), male, study on the software/hardware co-designing, system-level modeling methods, and network. Jin-jia Wang, (1983-), male, study on embedded systems.

System-level modeling and performance evaluation is the core of system-level design, and it is important to evaluate effectively with a abstract model which could reflect the attributes of the system accurately[3]. In this paper, the library function for performance evaluation is designed for speech recognition systems, and it has been designed successfully based on SystemC. It could evaluate the performance effectively for embedded systems, such as the average delay and CPU occupancy rate.

The performance evaluation results are very important in the process of performance evaluation for embedded speech recognition systems. In this paper, the performance evaluation results has been gotten by studying only a limited number of performance parameters based on the simulation methods. In the process of performance evaluation, there are two important problems must be solved. One of them is how to get and deal with the samples during the process of simulation, and the other is how to determine the process of simulation automatically when the performance evaluation results meet the need of the requirement[4] [5].

The library function for performance evaluation has been designed, and it solved the problems successfully. It is designed for counting the samples of performance evaluation results and calculating the expectations and the confidence interval. When the performance evaluation results meet the need of the requirement, it could terminate the simulation. In addition, a function of the library which could make the normal distribution random samples has been designed.

2   The Method of Performance Evaluation

For evaluating the performance of embedded speech recognition systems, a system-level framework based on SystemC for embedded speech recognition systems has been designed. As figure 1 shows.
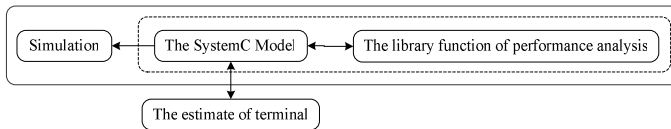
Fig. 1. The framework of performance evaluation based on SystemC

First of all, modeling speech recognition systems abstractly, and partition the algorithm into some tasks based on SystemC at system-level with the need of requirement.

Second, design the library for performance evaluation, and the abstract model for performance evaluation has been done. As the broken line frame in figure 1 shows.

Third, the performance evaluation results of speech recognition systems, such as the average delay and CPU occupancy rate, could be gotten. As the real line frame in figure 1 shows.

Finally, the model would terminal the simulation when the performance evaluation results meet the need of the requirement.

3   The Design of the Performance Evaluation Library

In this paper, the performance evaluation library has been designed with the flow like this:

First of all, the random number of normal distribution class has been designed for making the normal distribution random samples.

Second, the point-estimate value of the average performance evaluation results of the samples and its credibility which is called the confidence interval for embedded speech recognition systems has been calculated with the confidence class.

Third, the number of simulation times were controlled by the convergence class, and it would terminate the process of simulation when the performance evaluation results meet the need of the requirement.

Finally, the statistics class has been designed successfully with using the classes which just be designed expediently, and it could count the number of the samples which meet the need of the requirement.

The UML of the performance evaluation library is shown in figure 2.
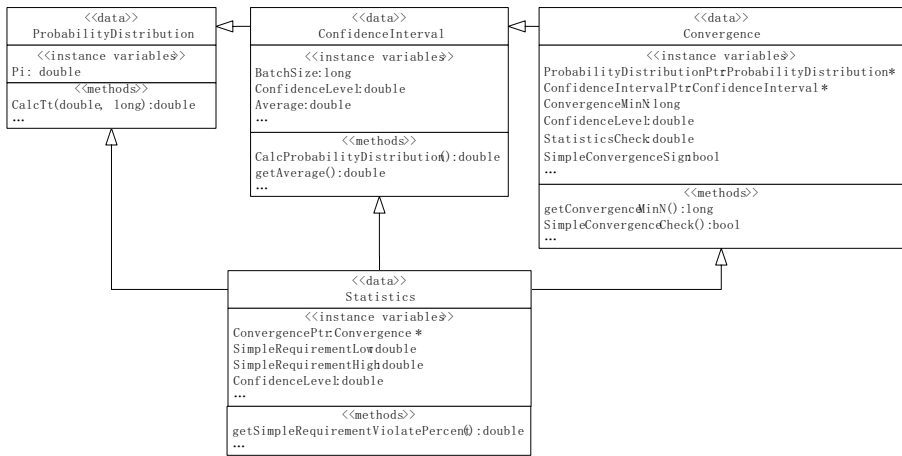
Fig. 2. The UML of the library function for performance evaluation

3.1   The Random Number of Normal Distribution Class

In order to make the random normal distribution samples, the random number of normal distribution class was designed, and its mathematical methods are as follows.

According to the function of the probability density of the normal distribution, expression 1 could be gotten.

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \tag{1}$$

The expression shows that the probability density which is near the average value which called μ is bigger, and which is far from the average value is smaller. So, the random samples would be made could meet the kind of distribution if its probability

is bigger when is close to the average value, but is smaller when is far from the average value. Therefore, the samples could be gotten like this: in one hand, the value which is below the curve of the probability density would be kept; and in the other hand, the value which is up the curve of the probability density would be removed.

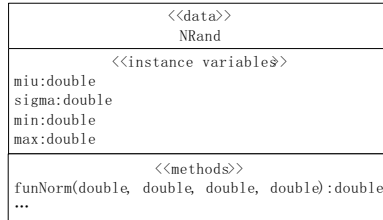In this way, the random number of normal distribution class has designed, and the UML of it is shown in figure 3.

Fig. 3. Random number of normal distribution class

The class has verified in the environment of the embedded speech recognition system which is based on the system-level of SystemC language, and it could make random samples which could meet the normal distribution. Because of the positive correlation between the probability density and the number of the sample, the probability density would be bigger if the number of the sample is larger, and the length of voice frame samples is 72. As figure 4 shows.
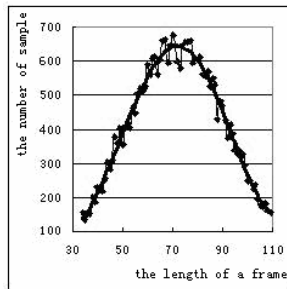
Fig. 4. Normal distribution of the length of voice frame samples

3.2 The Confidence Class

The confidence category includes two sub-classes: one if them is the probability distribution class, and the other is the confidence interval class. The probability distribution class is used to calculate the median value of the samples which confidence level is certain, and the confidence interval class is designed for calculating the average value and the confidence interval of the random samples.

3.2.1 The Probability Distribution Class
The median value of the T distribution which confidence level is certain could be calculated by using the median value of the beta distribution which is lebeled β_p, and

the median value of the beta distribution could be gotten with the iterative method. The mathematics methods are as follows[6]:

The relationship between the T distribution and the expression $I_x(a,b)$ is shown in expression 2.

$$T(t \mid n) = \begin{cases} 1 - \dfrac{1}{2} I_x(\dfrac{n}{2}, \dfrac{1}{2}), & t > 0 \\ \dfrac{1}{2} I_x(\dfrac{n}{2}, \dfrac{1}{2}), & t \le 0 \end{cases} \tag{2}$$

When the quantile $p \in (\frac{1}{2}, 1)$ and $t_p > 0$ are true, expression 3 would be true.

$$\frac{n}{n + t_p^2} = \beta_{p*}(\frac{n}{2}, \frac{1}{2}) \tag{3}$$

So, expression 4 could be gotten.

$$t_p(n) = \sqrt{\frac{n}{\beta_{p*}(\frac{n}{2}, \frac{1}{2})} - n} \tag{4}$$

If the quantile $p \in (0, \frac{1}{2})$ and $t_p < 0$ are true, the expression $t_p(n) = -t_{1-p}(n)$ would be true too.

Therefore, the expression $t_p(n)$ could be gotton by using expression 5.

$$t_p(n) = sign \ (p - \frac{1}{2}) \sqrt{\frac{n}{\beta_{p*}(\frac{n}{2}, \frac{1}{2})} - n} \tag{5}$$

And the quantile p^* in expression 5 could be gotten by using expression 6.

$$p^* = \begin{cases} 2 * p, & p \in (0, \dfrac{1}{2}) \\ 2 * (1 - p), & p \in [\dfrac{1}{2}, 1) \end{cases} \tag{6}$$

The expression $\beta_{p*}(\frac{n}{2}, \frac{1}{2})$ which is labeled 6 could be gotten with the iterative method, and the iterative method is suitable for calculating the median value.

Because of the truth of the expression $\beta_p(a,b) \in (0,1)$, the resultful interval of the equation $f(x) = I_x(a,b) - p = 0$ is $(0,1)$. With the iterative method, the approximative results of the median value which could meet the requirement, and the expression $I_x(a,b)$ could be gotten with expression 7.

$$\begin{cases} I_x(a+1, \ b) = I_x(a, \ b) - \dfrac{1}{a} U_x(a, \ b) \\ I_x(a, \ b+1) = I_x(a, \ b) + \dfrac{1}{b} U_x(a, \ b) \\ U_x(a+1, \ b) = \dfrac{a+b}{a} x U_x(a, \ b) \\ U_x(a, \ b+1) = \dfrac{a+b}{b} (1-x) U_x(a, \ b) \end{cases} \tag{7}$$

According to mathematical methods, the probability distribution class has designed, and the UML of it is shown in figure 5.

Fig. 5. Probability distribution class

The class has verified in the environment of the embedded speech recognition system which is based on the system-level of SystemC language, and it could get the T distribution value. When the number of the random samples is 2147483647, it is considered that the number of the random samples is infinitude, and the T distribution value which has calculated is 1.95996. It is very accord with the value which is refer to the sheet of the T distribution value.

3.2.2   The Confidence Interval Class

The confidence interval technical is used for calculating the definition of the point-estimate value of the average value of the samples in the confidence interval class. Its mathematics methods is as follows[7]:

The performance evaluation results of the embedded speech recognition system which is based on SystemC language could be gotten through the average value of the random samples which is labeled μ, and it could be calculated with expression 8.

$$\mu = \lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} X_i \tag{8}$$

The average value of the random samples which is labeled μ in expression 8 could be insteaded of $\overline{\mu}$ in expression 9.

$$\overline{\mu} = \frac{1}{n} \sum_{i=1}^{n} X_i \tag{9}$$

In this paper, the confidence interval technical has been used because the performance evaluation results which were gotten with the simulation method would be significative when it meet the need of the requirement. In one hand, the precision of the performance evaluation results could be gotten; and in the other hand, the process of simulation would be terminated when the precision of the performance evaluation results meet the need of the requirement. The scheme in detail is as follows:

According to the theorem of central limit, if the expression $X_1, X_2, \ldots, X_n$ with independent distribution, and its expectation and variance were μ and δ^2, the expression $R = \frac{1}{n} \sum_{i=1}^{n} X_i = \overline{\mu}$ would be subjected to the normal distribution when the number of the random samples is unnumberable, and the expectation and variance of the distribution were μ and δ^2. In this case, the expression $\sqrt{n} \, \frac{R - \mu}{\delta}$ could be subjected to the normal distribution, and the confidence interval of the expression $[\overline{\varphi_1}, \overline{\varphi_2}]$ which confidence level lebeled μ is 1-α could be gotten with expression 10.

$$\begin{cases} \overline{\varphi_1} = \overline{\mu} - \dfrac{S}{\sqrt{n}} t_{n-1}(\dfrac{\alpha}{2}) \\ \overline{\varphi_2} = \overline{\mu} + \dfrac{S}{\sqrt{n}} t_{n-1}(\dfrac{\alpha}{2}) \end{cases} \tag{10}$$

The expression S in expression 10 could be gotten with expression 11, and the expression $t_{n-1}(\alpha/2)$ is the sub-sites which could be subjected to the expression $\alpha/2$ of the T distribution.

$$S = \sqrt{\frac{1}{n-1} \sum_{i=1}^{n} (X_i - \overline{\mu})^2} \tag{11}$$

In this way, the confidence interval class has designed successfully, and the UML of it is shown in figure 6.

<<data>> ConfidenceInterval
<<instance variables>>
BatchSize: long
ConfidenceLevel: double
Average: double
...
<<methods>>
CalcProbabilityDistribution() :double
getAverage() :double
...
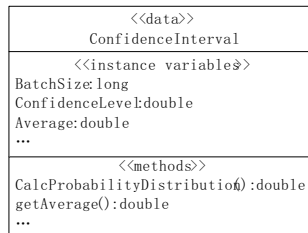
Fig. 6. Confidence interval class

The class has verified based on SystemC at system-level in the environment of the embedded speech recognition system, and it could get the expectation and confidence interval which are refer to the value by using the mathematical methods. For example, ten random samples are given: 10.1, 10.0, 9.8, 10.5, 9.7, 10.1, 9.9, 10.2, 10.3, and 9.9, and the expectation and confidence interval which confidence level is 0.95 are wished to get. By using the class, the results could be gotten, and the expectation and confidence interval are 10.05 and [9.87723, 10.2228]. It is quite refer to the value 10.05 and [9.88, 10.22] which are gotten with the mathematical methods.

3.3 The Convergence Class

The convergence class could terminate the process of simulation when the performance evaluation results meet the needs of the requirement, and there are two ways for designing. One of them is terminating the simulation when the difference of the results of the functions meet the needs of the requirement. The other is terminatng the simulation when the difference between the discrete point-estimate value of the average performance evaluation results and the adjacent point-estimate value of the random samples meet the needs of the requirement. In this paper, the convergence class has been designed by using the second one, and the UML is shown in figure 7.

The class has verified based on SystemC at system-level in the environment of the embedded speech recognition system, and it could terminate the process of simulation because the performance evaluation results meet the need of the requirement when it has simulated and analyzed for about 850 random samples successfully.
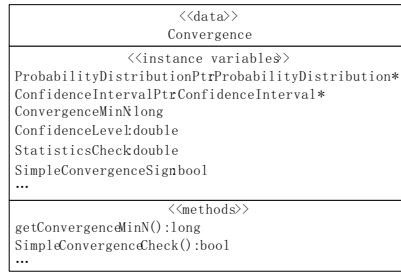
```
┌─────────────────────────────────────────────────────────┐
│                        《data》                           │
│                      Convergence                          │
├─────────────────────────────────────────────────────────┤
│                  《instance variables》                    │
│ ProbabilityDistributionPtr:ProbabilityDistribution*       │
│ ConfidenceIntervalPtr:ConfidenceInterval*                 │
│ ConvergenceMinN:long                                      │
│ ConfidenceLevel:double                                    │
│ StatisticsCheck:double                                    │
│ SimpleConvergenceSign:bool                                │
│ ...                                                       │
├─────────────────────────────────────────────────────────┤
│                     《methods》                            │
│ getConvergenceMinN():long                                 │
│ SimpleConvergenceCheck():bool                             │
│ ...                                                       │
└─────────────────────────────────────────────────────────┘
```

Fig. 7. Convergence class

3.4   The Statistics Class

The statistics class was designed successfully for using the classes which just be designed expediently and counting the numbers of the samples which could meet the need of the requirement, and the UML of it is shown in figure 8.

```
┌──────────────────────────────────────────────────┐
│                    《data》                        │
│                  Statistics                        │
├──────────────────────────────────────────────────┤
│              《instance variables》                 │
│ ConvergencePtr:Convergence *                       │
│ SimpleRequirementLow:double                        │
│ SimpleRequirementHigh:double                       │
│ ConfidenceLevel:double                             │
│ ...                                                │
├──────────────────────────────────────────────────┤
│                 《methods》                         │
│ getSimpleRequirementViolatePercent():double        │
│ ...                                                │
└──────────────────────────────────────────────────┘
```
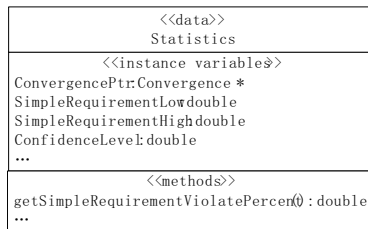
Fig. 8. Statistics class

The class has verified in the environment of the embedded speech recognition system which is based on the system-level of SystemC language, and it could work very well.

4   Applications of the Library

After the performance evaluation library has been designed, embedded speech recognition systems and the wireless AP systems have been modelled, and the performance of the systems has been evaluated. The correctness of the performance evaluation library has been verified by the model designed. In this paper, the performance evaluation results of speech recognition systems, such as the delay time and CPU occupancy rate, would be introduced in detail.

4.1   Performance of the Delay Time

The delay time of the system is the delay between a whole command received and the command has been processed.By the simulation, the relationship between delay time and the length of voice frame samples has been gotten as the figure 9.
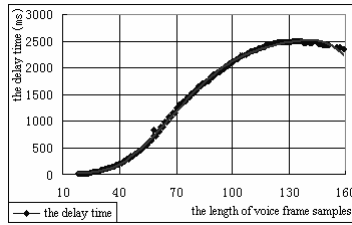
Fig. 9. The relationship between delay time and the length of voice frame samples

The delay time increases when the length of voice frame samples increase and it is below 132, and then the delay time decreases when the length of voice frame samples increase. From the evaluation results, the minimum delay time is 30 ms, the maximum delay time is 2500 ms, and the average delay time is 1316 ms.

4.2 Performance of the CPU Occupancy Rate

For making more effective use of the CPU resources for embedded systems, CPU occupancy rate of the collection and extraction process has been evaluated. By the simulation of the model, the relationship between CPU occupancy rate and the length of voice frame samples has been gotten, as shown in figure 10.
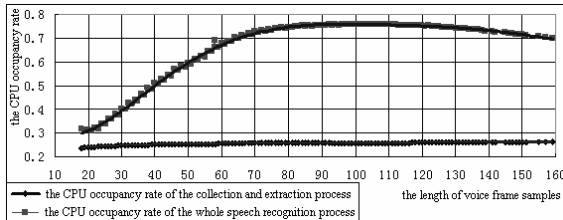
Fig. 10. The relationship between CPU occupancy rate and the length of voice frame samples

The CPU occupancy rate of the collection and extraction process as the thiner line shows, and the CPU occupancy rate of the whole speech recognition process as the bold line shows. The CPU occupancy rate of the collection and extraction process increases when the length of voice frame samples increase from 23.880% to 26.157%, the CPU occupancy rate of the whole speech recognition process increases when the length of voice frame samples increase from 30.953% to 76.035%, and the CPU occupancy rate of the whole speech recognition process decreases when the length of voice frame samples increase from 76.035% to 69.951%.

5 Conclusion

With the method for performance evaluation proposed, the performance evaluation library based on SystemC language has been designed, and it could help designers evaluate the performance of their systems. By using the design method porposed and

the library designed, the performance of embedded speech recognition systems has been modelled, and the performance parameters, such as the delay time and the CPU occupancy rate, have been gotten by simulation. For validating the correctness and practicality of the library function further, the library function has been also used in the performance evaluation of wireless AP systems, as well as in performance evaluation of complex embedded systems based on SystemC. It is of great value for extending the library of SystemC.

References

1. Chen, X., Xu, Y.N.: A Design of System On Chip Based On SystemC Language, pp. 1–6. Science Press, Beijing (2004)
2. Fang, L.B., Huang, Z.Q., Hou, Y.B., et al.: Studying on Extending POOSL and its Application on Performance Analysis. Journal of System Simulation 19(13), 2908–2912 (2007) (in Chinese)
3. Pimentel, A.D., Erbas, C., Polstra, S.: A Systematic Approach to Exploring Embedded System Architectures at Multiple Abstraction Levels. IEEE Transactions on Computers 55(2), 99–112 (2006)
4. Keutzer, K., Newton, A.R., Rabaey, J.M., Sangiovanni-Vincentelli, A.: System-level Design: Orthogonalization of Concerns and Platform-based Design. IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems 19(12), 1523–1543 (2000)
5. Coffland, J.E., Pimentel, A.D.: A Software Framework for Efficient System-level Performance Evaluation of Embedded Systems. In: Proceedings of the 2003 ACM symposium on Applied computing, March 2003, pp. 666–671 (2003)
6. Gao, H.X.: Statistical Calculation, pp. 57–74. Beijing University Press, Beijing (1995)
7. Wang, B.F.: Probability Discussed Mathematical Statistics, 3rd edn., pp. 212–242. Tongji University Press, Shanghai (1994)
8. Hull, T.E., Dobell, A.R.: Random Number Generators. SIAM Rev. 4, 230–253 (1962)
9. Askinson, A.C.: Tests of Pseudo-random Numbers. Applices Statistics 29, 164–171 (1980)

An Intelligent Computing Algorithm to Analyze Bank Stock Returns

Vincenzo Pacelli

Ph. D. - University of Rome "La Sapienza"
Assistant Professor of Economics of Financial Intermediaries,
University of Foggia, Faculty of Economics
Via Caggese, 1-71100- Foggia Italy
v.pacelli@unifg.it

Abstract. The objective of this paper is to propose an intelligent computing algorithm, represented by an artificial neural network model, to analyze the dynamics of stock prices of banks. Through the empirical application of the model developed, it is expected to obtain indications about the ability of the artificial neural network model developed to generalize the phenomenon analyzed. So the research aims to provide empirical results about the use of non-linear methods of analysis for the study of the dynamics of banks' stock prices, enriching the prospects for research in terms of methodological tools.

Keywords: Neural Networks, Forecasting, Stock Returns; Banks.

1 Introduction

The international financial crisis started in the 2008 has highlighted the inability of the financial markets to predict and analyze the phenomena that can influence the dynamics of stock prices of the banks. This is largely due to a gap of knowledge regarding both the variables and phenomena that can influence the stock prices and the techniques which are able of analyzing the available information.

This paper has a dual objective: (*a*) to analyze empirically the ability of the artificial neural network model developed to generalize the phenomenon of the dynamics of stock prices of banks through the use of some variables chosen as inputs; (*b*) to propose an evolution of the artificial neural network model developed and empirically applied.

So far the scientific literature has addressed the analysis of the relations that bind corporate variables and stock prices of banks primarily through linear methods of analysis [*Cooper, Jackson, Patterson (2003); Romagnoli (2007)*]. This leads to an obvious gap in financial literature about the analysis of the methodological tools used to study the phenomenon analyzed.

This paper aims in fact to contribute to the literature on the methodological tools used for the analysis of the efficiency of capital markets [*Fama (1970); Grossman (1976); Grossman, Stiglitz (1980); Tobin (1984); White (1988); Clemen (1989); Granger, Newbold (1989); Fama, French (1993); Grudnitski, Osburn (1993);*

Clements, Hendry (1998); Makridakis, Wheelwright, Hyndman (1998); Shachmurove, Witkowoska (2000); Lam (2004); Dutta, Jha, Laha, Mohan (2006); Moreno, Olmeda (2007)] in the field of forecasting and interpreting the dynamics of stock prices of listed companies, with a specific focus on the utilizing of artificial neural networks to the prediction of stock prices of banks, that is an issue still unexplored by the literature.

In the last twenty years there has been a growing body of literature based on the comparison of artificial neural network to traditional statistical methods of analysis in economics and finance.

Hertz, Krogh and Palmer (1991) offer a comprehensive view of neural networks and issues of their comparison to statistics.

Many economists also advocate the application of neural networks to different fields in economics and finance [*Altman, Marco, Varetto (1994); Baestaens (1994); Bierens (1994); Cheh, Weinberg, Yook (1999); Cogger, Koch, Lander (1997); Dutta, Jha, Laha, Mohan (2006); Grudnitski, Osburn (1993); Hamm, Brorsen B. Wade (2000); Hawley, Johnson, Raina (1990); Hu, Tsoukalas (1999); Jagielska (1993); Kuan Chung-Ming, White (1994); Lam (2004); Ling (1993); McNelis (2004); Refenes (1995); Refenes, White (1998); Shachmurove, Witkowoska (2000); White (1988); White (1992)*].

The non-linear and often obscure relations that govern the economic and financial variables, the presence of significant amounts of data and the failures of the conventional mathematical and statistical models have encouraged a growing development of studies on the neural networks in the economic and financial fields. Many studies were devoted in particular to the study of neural networks as models able to predict historical time series of economic and financial data [*Clemen (1989); Clements, Hendry (1998); Dutta, Jha, Laha, Mohan (2006); Granger, Newbold (1989); Lam (2004); Makridakis, Wheelwright, Hyndman (1998); Moreno, Olmeda (2007); White (1988); White (1992)*]. According to Granger (1991), non-linear relationships in financial and economic data are more likely to occur than linear relationships. Several authors have also examined the application of neural networks to financial markets, where the non-linear properties of financial data provide many difficulties for traditional methods of analysis [*Omerod, Taylor, Walker (1991); Grudnitski, Osburn (1993); Altman, Marco, Varetto (1994); Kaastra, Boyd (1995)*].

About the specific theme of this research, Yoon and Swales (1997) compare neural networks to discriminative analysis with respect to the prediction of stock price performance and find that the neural network is superior to discriminative analysis in its predictions. So neural networks can be designed and trained to discern historical data patterns in order to predict future stock price trends.

Kryzanowski, Galler and Wright (1993) and Zirilli (1997) have shown that neural networks can recognize stocks outperforming the market. Kryzanowski, Galler and Wright (1993) also trained neural networks to classify correctly 72% of the test stocks in relation to their performance. Refenes (1995) and Bansal and Viswanathan (1993) explore a dynamic version of the arbitrage pricing model (APT), replacing linear regression with neural networks to rank stocks. These authors find that, compared to regression models, neural networks better identify stocks that outperform the market.

2   The Method: The Artificial Neural Network Model

As regards the method, this research is based on the hypothesis that the intelligent computing techniques are able to discover aspects of the phenomenon of the shareholders' value added in banking otherwise invisible.

An empirical analysis was then carried out using an Artificial Neural Network Model – ANN_m.

Artificial Neural Networks are non-linear mathematical models that can be trained to map past and future values of time series data and thereby extract hidden relationship that govern the data. Artificial Neural Networks consist of multiple neurons which are extensively interconnected and organized in layers similar to those of a decision tree. These neurons work in unison to solve specific problems and, when trained through time series of data, become potent tools for analysis and forecasting.

A single artificial neuron, which is the basic element of the neural network, comprises several inputs (i_1, i_2, …, i_m) and one output (y) that can be written as follows:

$$y = f(i_n, w_n) \tag{1}$$

where, w_n are the function parameter weights of the function f.

Equation (1) is called an activation function. Each artificial neuron determines its output by applying an activation function and weights to inputs. The activation function of the ANN_m developed for this research is the symmetrical sigmoid function that can be written as follows:

$$f(y) = 1 / [1 + \exp(-k\Psi y)] \tag{2}$$

where, $\Psi > 0$ is the slope of sigmoid in its point of inflection y = 0, and k is a constant. The output (y) of the symmetrical sigmoid function is included in the range [-1, 1].

With regard to architecture of the ANN_m, it has developed a neural network multilayer fully connected and characterized by 4 hidden layers. The neurons in each layer are: 27 input neurons, 15 neurons in the first hidden layer, 10 neurons in the second hidden layer, 10 neurons in the third hidden layer, 7 neurons in the fourth hidden layer and 1 output neuron.

The response of an artificial neural network is determined by the synaptic values (weights) of the connections between the nodes of the neurons. In the same way biological nervous systems are able of learning by experience, artificial neural networks learn the reality analyzed by changing gradually their synaptic values (weights) ΔW_{ij} when trained through time series of data (pattern of input and output) and through the use of a learning algorithm.

The learning algorithm used in the training of the ANN_m developed is the back propagation algorithm with learning for cycles (on line). It is so developed an ANN_m feed-forward. In the learning for cycles, the changing of the synaptic values (weights) ΔW_{ij} of the ANN_m is calculated after each presentation of a single pattern. The new configuration of the synaptic values (weights) after a cycle of training is calculated by adding the change obtained $\Delta W_{ij}(t)$ to the previous configuration of the synaptic values $W_{ij}(t-1)$. The speed of learning is regulated by a constant η – the learning rate – which defines the portion of change that is applied to the synaptic values. The equation of learning can be write as follows:

$$W_{ij}(t) = W_{ij}(t-1) + \eta\, \Delta W_{ij}(t), \tag{3}$$

where, $0 < \eta < 1$ and, in the ANN_m developed, $\eta = 0{,}7$.

The training of the ANN_m developed was interrupted when the mean square error has become less of the value $0{,}0001$. At the end of the training of the first ANN_m (output at time "t") the mean square error was $0{,}00002$. At the end of the training of the second ANN_m (output at time "t+1") the mean square error was $0{,}00007$.

Once the learning phase was completed, the synaptic values are frozen and it is possible to study the response of the ANN_m on other patterns of data in the phase of test. The phase of test consists in the presentation of new patterns and in the calculation of activation of the nodes of the ANN_m with synaptic weights frozen.

It was carried out testing of the ANN_m through three patterns of input and output (records) that have not been used for the training. The results of the tests of the 2 ANN_m were as follows: *(a)* the mean square error for the first ANN_m (output at time "t") was $0{,}0031$; *(b)* the mean square error for the second ANN_m (output at time "t +1") was $0{,}0465$. So both the ANN_m are very precise. This means that the ANN_m developed have a good ability to generalize the phenomenon analyzed[1].

3   Description of the Data

In order to train and test the ANN_m developed, an empirical study was conducted on the ability to create and diffuse economic value by a panel of listed European banks.

The empirical analysis was conducted from 2002 to 2006 on a panel of 16 banks.

The panel of banks was selected from a population of European banks listed, active and independent at the end of December 2006[2].

Using data emerged from the empirical analysis, 2 different ANN_m have developed which are completely identical in both mathematical and scientific architecture and about the inputs used, but which are different about the output considered. The first ANN_m considers as output the per cent annual variation of the share prices of banks in the reference year of the inputs (year "t"). The second ANN_m considers as output the per cent annual variation of the share prices of banks in the year following that reference the inputs (year "t+1").

The 27 inputs used in the 2 ANN_m were calculated for the 16 banks investigated and for the 3 years analyzed (48 patterns of input and output or 48 records). The 27 inputs are as follows:

- Net income;
- per cent annual variation of the Total Capital;
- per cent annual variation of the Equity;
- per cent annual variation of the Total Capital ratio;
- per cent annual variation of the Tier 1 Capital ratio;

---

[1] A problem may be represented by a possible overfitting of the neural networks that would decrease the probability that neural networks discover the general function that describes the phenomenon analyzed. This problem can be overcome by increasing the number of the pattern of input and output used for the training and the testing of the neural networks.

[2] The empirical analysis was conducted on data from consolidated balance of banks by source Bankscope.

- per cent annual variation of the Capital Funds to Liabilities ratio;
- per cent annual variation of the Total Assets;
- per cent annual variation of the Total Earning Assets;
- per cent annual variation of the Fixed Assets;
- per cent annual variation of the Intangible Assets by bank's accounting;
- per cent annual variation of the Cost to Income ratio;
- per cent annual variation of the Net Interest Margin ratio, which can be estimated as follows:

$$\text{Net Interest Margin} = (\text{Interest Margin/Average Total Earning Assets})*100 \quad (4)$$

- Knowledge Capital Earning (KCE), which can represent an indicator about the variation of the organizational capital[3] and can be estimated as follows:

$$\text{KCE} = (\text{Net Interest Margin - Tasso EURIBOR})*\text{Total Earning Assets} \quad (5)$$

- per cent annual variation of the Deposits & Short Term Funding;
- per cent annual variation of an evaluation of Brand towards customers, which can be estimated as follows:

$$\Delta V_b = (\Delta \text{ bank "x" funding}) - (\Delta \text{ banks analyzed average funding}) \quad (6)$$

- per cent annual variation of the Dividend Pay Out ratio;
- per cent annual variation of the Price to Book Value ratio;
- per cent annual variation of the Price to Earning ratio;
- per cent annual variation of an evaluation of the reputational (or relational) capital (RCV) towards financial markets, which can be estimated as follows:

$$\text{RCV} = [(\text{P/BV}) - (\text{Roae/Ke})]*\text{Equity} \quad (7)$$

- per cent annual variation of the Return on Average Equity ratio;
- per cent annual variation of the cost of equity (K_e) which can be estimated by using the Capital Asset Pricing Model as follows:

$$K_e = r_f + \beta(r_m - r_f) \quad (8)$$

where:

- r_f is the riskless return rate and was estimated as equal to the annual average of the Euribor rate at 3 months;
- r_m is the expected market return rate for similar investments and was estimated as equal to 9,16% which is the minimum annual return rate – between 2002 and 2006 – of the Italian banking sector index by Borsa Italiana;
- β is the beta of the share and expresses the historical relation between the variation of market returns and the variation of the share returns. In other words, the beta of the share measures the sensitivity of the share trend to the variations of the market index. The beta of a one-year share has been considered as correlated with the Mib 30 Index (for Italian banks) and with the DJ Euro Stoxx 50 index (for Foreign banks).

---

[3] See Lev (2001); Aiaf (2003), pp. 40-41.

- per cent annual variation of the economic capital value, which was estimated by using a revenue method as follows:

$$\text{Economic Capital Value} = \text{Net Income}/ K_e \tag{9}$$

- per cent annual variation of the annual average of the Euribor rate at 3 months;
- per cent annual variation of the Beta;
- annual Volatility of shares prices, which was estimated as follows:

$$V_p = [(P_{Max} - P_{Min})/(P_{Dec})]*100 \tag{10}$$

where:

o P_{Max} is the high annual value of the share prices;
o P_{Min} is the low annual value of the share prices;
o P_{Dec} is the share price at December of the year analyzed.

- per cent variation – between 2 year – of the Net income;
- upgrade or downgrade of Fitch Rating Long Term.

The inputs' portfolio developed provides – as a whole – an estimate of the dynamics of the variables which determine the supply and demand of the economic capital of a bank. The ANN_m developed, therefore, aims to enhance the impact of each variable of input considered on the bank share prices (output of the ANN_m).

Once developed and tested the 2 ANN_m, in order to determine the effects on outputs caused by the variations of each input, it is proceeded – for the 2 ANN_m – as follows.

Increased individually by 10% the values of each input of the 2 ANN_m, keeping time to time constant the values of the other 26 input, it was studied the deviations of the output produced by the ANN_m as a result of the variation of the input from the output expected, namely that provided by the ANN_m without variations of inputs.

$$\text{First } ANN_m: O_t = f[(1,10 * I_1); K\ I_{2\text{-}27}] \tag{11}$$

$$\text{Second } ANN_m: O_{t+1} = f[(1,10 * I_1); K\ I_{2\text{-}27}] \tag{12}$$

In order to be used to train the ANN_m, the patterns of input and output must be appropriately amended (normalised) through their transformation into the range [-1, 1] as follows:

$$I = I_{min} + [(I_{max}-I_{min})*(D-D_{min})/(D_{max}-D_{min})] \tag{13}$$

where:

- D_{min} and D_{max} are the extreme values of the ranges – namely the maximum and minimum values – of the variables of input and output;
- D is the value of each input and output;
- I_{max} and I_{min} are the new extreme values of the range or "-1" and "1".

4 The Empirical Results

This paragraph introduces the empirical results produced by the application of the 2 ANN_m developed in accordance with the methodology outlined in paragraphs 3 and 4.

A first result is about the ability of the 2 ANN_m developed to generalize the phenomenon analyzed. As previously underlined, the results of the tests of the 2 ANN_m were as follows: *(a)* the mean square error for the first ANN_m (output at time "t") was 0,0031; *(b)* the mean square error for the second ANN_m (output at time "t +1") was 0,0465. So both the ANN_m are very precise. This means that the ANN_m developed have a good ability to generalize the phenomenon analyzed.

A second result is about the influence of the variables used as input in pricing of the economic capital of the banks. The analysis of the data reveals that the variables used as input of the 2 ANN_m do not have a significant influence on the share prices of the banks (output of the ANN_m).

As a result, in fact, of the increase of 10% of each input, the per cent variations of the outputs of the 2 ANN_m compared with the outputs expected are included among the maximum limit of 6,04% (as a result of the variation of the equity at time "t") and the minimum limit of -5,80% (as a result of the variation of the Price to Earning ratio at time "t").

According to the ANN_m developed, therefore, the inputs analyzed do not influence the share prices of the banks, both in the year of the inputs' accounting and in the following one. And I am not so surprised by this. As it is possible to deduce from the portfolio of the inputs analyzed, the 2 ANN_m have been developed using as inputs information – or data – characterized by a quantitative nature and freely accessible through the public disclosure of the banks or the financial markets. In other words, the 2 ANN_m have been developed using as inputs indicators characterized by a not innovative nature – these being derived from information already incorporated in the share prices – as required by the economic theory assuming a good efficiency of the markets.

What then emerges represents a not surprising fact which leads to evolve the artificial neural network model (ANN_m) developed in the direction that is illustrated in Appendix A.

5   Concluding Remarks

The empirical results emerging through the application to the panel of the banks analyzed of the non-linear method of analysis described in the paper seem to reveal the following relevant aspects:

1.   The variables characterized by a purely quantitative nature and measurable using public and not innovative information do not seem to express a significant influence on the share prices of the banks. This is because these variables are already reflected and incorporated in the share prices, assuming a good efficiency of the markets;

2.   The ANN_m developed have a good ability to generalize the phenomenon analyzed.

The two considerations highlighted above lead indirectly to a third consideration:

3.   There are some banks' variables – often characterized by an invisible and intangible nature – which contribute to the creation of the value and the price of the economic capital of a bank but are difficult to measure or account through linear methods of analysis that consider only quantitative and public data. So it would be superficial not analyze these intangibles variables because the empirical evidence shows that these intangibles have a fundamental importance both in determining the

expectations of investors in the banking shares and in determining the competitive advantage of the bank in the market.

The three considerations highlighted above lead to a fourth consideration which is essentially methodological and is about the effectiveness of an integrated approach, which is based on the joint use of linear and non-linear methods of analysis to study the phenomenon analyzed in this paper. Through this research, in fact, it emerges that the artificial neural network model developed have a good ability to generalize the phenomenon.

As highlighted previously, in fact, the artificial neural networks are particularly suited to analyze and interpret – revealing hidden relationships that govern the data – complex and often obscure phenomena and processes, which are, for example, those governing the dynamics of the share prices in financial markets. In order to be efficient, an artificial neural network must be properly developed and trained. About the adequacy of the ANN_m's training, there are two weaknesses of the ANN_m developed in this research that are about the number of the records (patterns of inputs and output) used to train the ANN_m and the purely quantitative and public nature of the inputs analyzed. As highlighted previously the artificial neural networks have as many chances to be efficient as more numerous are the series of data (records) used to train the ANN_m. In this research the number of the records used to train the ANN_m was relatively small.

The second weakness of the ANN_m developed is about the purely quantitative and public nature of the inputs analyzed. In order to develop an efficient ANN_m, in fact, it would be necessary to consider and analyse all the phenomena which are able to influence – both directly and indirectly – the share prices of the banks.

However if the first weakness can be easily remedied by increasing the number of the banks analyzed and extending the period of analysis in order to increase the number of the records, there would be further difficulties – from the viewpoint of a researcher outside the bank – where you would overcome the second weakness. Many phenomena that can significantly influence the share prices of the banks have, in fact, essentially a qualitative nature and a profile of confidentiality which hinders their disclosure. Many of these phenomena, therefore, are hardly measurable and, consequently, analyzable because of their qualitative nature. And where quantifiable, the information on these phenomena have a profile of confidentiality which hinders their disclosure.

These arguments, however, are not reasons enough to hinder the theoretical evolution of the ANN_m developed (see Appendix A). Obviously the artificial neural networks, which aim to predict bank stock returns through a comprehensive study of the possible relevant phenomena, will have inevitably a disclosure and, therefore, an utility strictly within the bank, as means of strategic planning and internal control of the management. This is because only the management of the bank has all those confidential and qualitative information that should be used as input of an artificial neural network which has such an ambitious aim.

References

1. Aiaf.: Intangibles: metodi di misurazione e valorizzazione. In: Quaderno di ricerca 113 (2003)
2. Altman, E.I., Marco, G., Varetto, F.: Corporate Distress Diagnosis.: Comparisons Using Linear Discriminant Analysis and Neural Networks (The Italian Experience). Journal of Banking and Finance 18, 505–529 (1994)

3. Baestaens, D.E.: Neural Network Solutions for Trading in Financial Markets. Financial Times Management, London (1994)
4. Bansal, R., Viswanathan, S.: No Arbitrage and Arbitrage Pricing: A New Approach. Journal of Finance 48(4), 1231–1262 (1993)
5. Bierens, H.J.: Comment on Artificial Neural Networks: An Econometric Perspective. Econometric Reviews 13 (1994)
6. Cheh, J.J., Weinberg, R.S., Yook, K.C.: An Application of an Artificial Neural Network Investment System to Predict Takeover Targets. Journal of Applied Business Research 15(4), 33–45 (1999)
7. Clemen, R.T.: Combining Forecast. A Review and Annotated Bibliography. International Journal of Forecasting 5, 559–583 (1989)
8. Clements, M.P., Hendry, D.F.: Forecasting economic time series. Cambridge University Press, Cambridge (1998)
9. Cogger, K.O., Koch, P.D., Lander, D.M.: A Neural Network Approach to Forecasting Volatile International Equity Markets. In: Hirschey, M.M., Wayne, M. (eds.) Advances in financial economics, Greenwich, Conn. and London, pp. 117–157. JAI Press (1997)
10. Cooper, M.J., Jackson III, W.E., Patterson, G.A.: Evidence of predictability in the cross-section of bank stock returns. Journal of Banking and Finance 27, 817–850 (2003)
11. Dell'Atti, S., Pacelli, V.: Prezzo e valore nelle banche. Un'analisi empirica nei principali gruppi bancari italiani. Banche e Banchieri 6 (2006)
12. Dutta, G., Jha, P., Laha, A.K., Mohan, N.: Artificial Neural Network Models for Forecasting Stock Price Index in the Bombay Stock Exchange. Journal of Emerging Market Finance 5 (2006)
13. Fama, E.F.: Efficient capital market: a review of theory and empirical work. Journal of Finance (May 1970)
14. Fama, E.F., French, K.R.: Common Risk Factors in the Returns on Stocks and Bonds. Journal of Financial Economics 33 (1993)
15. Granger, C.W.J.: Developments in the Nonlinear Analysis of Economic Series. Scandinavian Journal of Economics 93(2) (1991)
16. Granger, C.W.J.: Newbold.: Forecasting Economic Time Series. Academic Press, Inc., New York (1989)
17. Granger, C.W.J., Ramanathan, R.: Improved Methods of Combining Forecasts. Journal of Forecasting 3, 197–204 (1984)
18. Grossman, S.J.: On the efficiency of competitive stock markets where traders have different information. The Journal of Finance XXXI, 2 (1976)
19. Grossman, S.J., Stiglitz, J.E.: On the impossibility of informatively efficient markets. The American Economic Review 70 (June 1980)
20. Grudnitski, G., Osburn, L.: Forecasting S&P and Gold Futures Prices: An Application of Neural Networks. Journal of Futures Markets 13(6), 631–643 (1993)
21. Hamm, L., Brorsen, W.B.: Trading Futures Markets Based on Signals from a Neural Network. Applied Economics Letters 7(2), 137–140 (2000)
22. Hawley, D.D., Johnson, J.D., Raina, D.: Artificial Neural Systems: A New Tool for Financial Decision-Making. Financial Analysis Journal, 63–72 (November/December 1990)
23. Hertz, J., Krogh, A., Palmer, R.G.: Introduction to the Theory of Neural Computation, Redwood City. Addison-Wesley, Reading (1991)
24. Hu, M.Y., Tsoukalas, C.: Combining Conditional Volatility Forecasts Using Neural Networks: An Application to the EMS Exchange Rates. Journal of International Financial Markets, Institutions & Money 9(4), 407–422 (1999)

25. Jagielska, I.: The application of neural networks to business information systems. In: Proceedings of 4th Australian Conference on Information Systems, Brisbane, pp. 565–574 (1993)
26. Kaastra, I., Boyd, M.S.: Forecasting Futures Trading Volume Using Neural Networks. Journal of Futures Markets 15(8) (December 1995)
27. Kryzanowski, L., Galler, M., Wright, D.W.: Using Artificial Neural Networks to Pick Stocks. Financial Analysts Journal 49(4), 21–27 (1993)
28. Kuan, C.C.M., White, H.: Artificial Neural Networks: An Econometric Perspective. Econometric Views 13(1), 1–91 (1994)
29. Lam, M.: Neural network techniques for financial performance prediction: integrating fundamental and technical analysis. Decision Support Systems 37 (2004)
30. Lev, B.: Intangibles: Management, Measurement, and Reporting. Brookings Insitution Press, Washington (2001)
31. Ling, C.S.: Choosing the right neural network model for trading. In: Proceedings of First Symposium on Intelligent Systems Applications, Singapore, pp. 26–33 (1993)
32. Makridakis, S., Wheelwright, S.C., Hyndman, R.J.: Forecasting - Methods and Applications. John Wiley & Sons, New York (1998)
33. McNelis, P.D.: Neural Networks in Finance: Gaining Predictive Edge in the Market. Academic Press, Inc., New York (2004)
34. Moreno, D., Olmeda, I.: Is the predictability of emerging and developed stock markets really exploitable? European Journal of Operational Research 182 (2007)
35. Omerod, P., Taylor, J.C., Walker, T.: Neural Networks in Economics. In: Taylor, M.P. (ed.) Money and Financial Markets. Blackwell, Cambridge (1991)
36. Refenes, A.P.: Neural Networks in the Capital Markets. John Wiley & Sons, New York (1995)
37. Refenes, A. P., White, H.:Neural Networks and Financial Economics. International Journal of Forecasting 17 (1998)
38. Romagnoli, A.: Indici di bilancio e rendimenti di borsa: un'analisi per le banche italiane. Temi di discussione della Banca d'Italia 648, Roma (2007)
39. Shachmurove, Y., Witkowoska, D.: Utilizing Artificial Neural Network Model to Predict Stock Market, Caress Working Paper. 11 (2000)
40. Tobin, J.: On the efficiency of the financial system. In: Lloyds Bank Review, Luglio (1984)
41. White, H.: Economic Prediction Using Neural Networks: The Case of IBM Stock Prices. In: Proceedings of the Second Annual IEEE Conference in Neural Networks, pp. 451–458. IEEE Press, New York (1988)
42. White, H.: Artificial Neural Networks: Approximation and Learning Theory. Blackwell, Oxford (1992)
43. Zirilli, J.S.: Financial Prediction Using Neural Networks. International Thompson Computer Press, London (1997)

Appendix A

In this Appendix it is proposed an evolution of the ANN_m described in paragraph 3, with similar mathematical architecture but different input[4].

Table 1. New Input for the Artificial Neural Network Model

	Input of the Artificial Neural Network
	I. Indicators about the variation of the financial capital and the financial capital adequacy to risk weighted assets
	I a). Indicator about the variation of the equity
	I b). Indicator about the variation of the tier 1 and total capital
	I c). Indicators about the variation of the financial capital adequacy to risk weighted assets
	II. Indicators about the variation of the tangible and visible capital
	III Indicators about the variation of the intangible and invisible capital
	IIIa). Indicators about the variation of the organizational capital
	IIIb). Indicators about the variation of the reputational (relational) capital
	IIIb 1). Indicators about the variation of the reputation towards customers
	IIIb 2). Indicators about the variation of the reputation towards financial markets
	IIIb 3). Indicators about the variation of the reputation towards employees
	IIIb 4). Indicators about the variation of the reputation towards suppliers
	Output of the Artificial Neural Network
Output 1	*% Value of the Bank's Economic Capital between t_0 e t_1*
Output 2	*% Price of the Bank's Economic Capital between t_0 e t_1*

The objective of this ANN_m is to analyse the impact of each input analyzed to the economic value and to the market value of the bank (2 outputs of this ANN_m).

By the analysis of the portfolio of the inputs considered (inputs representative of phenomena within the bank)[5], it is known as many of these inputs have a purely qualitative nature and therefore it is difficult to quantify their value. In order to estimate the variation – between t_0 and t_1 – of each input considered, the management of the bank can express evaluations on a scale from "-10" to "+10".

A value of "-10" indicates a maximum negative variation.

A value of "+10" indicates a maximum positive variation.

A value of "0" indicates no variation in the period considered.

The period between t_0 and t_1 may be 3 years in order to consent to the investments in intangibles to produce their effects on the value and the price of the economic capital of a bank.

---

[4] In the description of the input of the ANN_m, then, it is omitted the analysis of the mathematical architecture, for which you refer to in paragraph 3.

On an Ant Colony-Based Approach for Business Fraud Detection[*]

Ou Liu[1], Jian Ma[2], Pak-Lok Poon[1], and Jun Zhang[3]

[1] School of Accounting and Finance, The Hong Kong Polytechnic University,
Hong Kong, China
{afliuou,afplpoon}@inet.polyu.edu.hk
[2] Department of Information Systems, City University of Hong Kong,
Hong Kong, China
isjian@cityu.edu.hk
[3] Department of Computer Science, Sun Yat-sen University,
Guangzhou, China
junzhang@ieee.org

Abstract. Nowadays we witness an increasing number of business frauds. To protect investors' interest, a financial firm should possess an effective means to detect such frauds. In this regard, artificial neural networks (ANNs) are widely used for fraud detection. Traditional back-propagation-based algorithms used for training an ANN, however, exhibit the local optima problem, thus reducing the effectiveness of an ANN in detecting frauds. To alleviate the problem, this paper proposes an approach to training an ANN using an ant colony optimization technique, through which the local optima problem can be solved and the effectiveness of an ANN in fraud detection can be improved. Based on our approach, an associated prototype system is designed and implemented, and an exploratory study is performed. The results of the study are encouraging, showing the viability of our proposed approach.

Keywords: Ant colony optimization, artificial neural network, fraud detection.

1 Introduction

In today's dynamic business environment, we have seen a great flood of business frauds, including telecom subscription frauds, credit card frauds, and financial statement frauds. Countless fraudulent cases (such as the Enron Saga) have been reported in the literature [1, 2, 6, 11, 14]. Fraud detection is no longer an option to protect investors' interest. The enactment of the Sarbanes-Oxley Act in 2002 clearly shows that fraud detection is under the spotlight in the business sector.

Information technologies, if properly used, can help detect and manage frauds in an efficient and effective manner [2, 6, 16]. Numerous methods and technologies have been proposed for the implementation of Fraud Management Systems (FMSs) [2, 9,

---

14, 16]; among which rule-based scoring models [12, 15], statistical tools [7, 10], and artificial neural networks (ANNs) [8, 14] are the common ones.

Although ANNs have some advanced features and are widely used for fraud detection (more details are given in Section 2), traditional back-propagation-based algorithms used for ANN training have the local optima problem which reduces the effectiveness of ANNs in detecting frauds. To alleviate the problem, this paper proposes an approach to training an ANN using an ant colony optimization technique, which can solve the local optima problem and improve the effectiveness of ANNs in fraud detection.

Our literature review reveals that training an ANN with an ant colony-based approach for business fraud detection is a seldom-visited (or even yet-to-be-explored) area. Motivated by this observation, this paper discusses our attempt to explore the viability of this approach for fraud detection. Section 2 outlines some previous work on fraud detection methods. Section 3 discusses the details of using an ant colony-based approach for training ANNs for fraud detection. Section 4 illustrates a prototype system built on the proposed approach and discusses an exploratory study to evaluate the viability of that approach. Finally, Section 5 summarizes the paper.

2   Previous Work on Fraud Detection Methods

This section outlines some previous work on fraud detection methods. They include rule-based scoring models, statistical tools, and artificial neural networks.

A scoring model is a set of rules that describes the characteristics of some particular known fraud types [12, 15]. Scoring models work well when the fraud types are simple and manageable in the scale and complexity of rule parameters. Even for simple fraud types, however, they can be very laborious to define all the rules. In addition, scoring models have no self-learning capability and are unable to swiftly adapt to new fraud types.

Several statistical tools are used to detect frauds [7, 10]. Least-squares regression analysis is used in fraud audits [10]. In this regard, practical guidelines are provided to those auditors who are interested in detecting frauds without focusing on the associated mathematical details. In [7], a hybrid financial analysis model is presented, involving static and trend analysis models to construct and train a back-propagation neural network model. The experiments in [7] show that the proposed model not only provides a high predication rate but also outperforms other models including discriminant analyses, decision trees, and back-propagation neural networks alone. However, pure statistical tools can hardly provide an adaptive learning capability in fraud management [7].

Artificial neural networks (ANNs) with self-learning capabilities are proposed to overcome the above problems [3, 4, 7]. ANNs learn by experience, generalize from previous experiences to new ones, and can make decisions. ANNs are motivated by information-processing units as neurons in the human brain that a neural network is made up of artificial neurons [3, 4, 7]. A neural network can be thought of as a black-box non-parametric classifier. Neural networks are therefore more flexible. ANNs have become an important technology for pattern recognition and implementation of a FMS.

Despite of their growing popularity, the performance of ANNs largely depends on how the networks are trained [13]. We note that most of the current ANN training for fraud detection is designed based on a back-propagation (BP) algorithm. This type of algorithms, however, has a well-known problem, that is, difficulty in escaping local optima. This is because all BP algorithms used for training ANNs initialize the starting point in an *n*-dimensional space randomly. If the starting point is located in a local valley, a BP algorithm probably converges on a *local* solution. Many modifications have been made to overcome this problem by finding the *global* solution [13]. One of them is to train an ANN with an ant colony-based approach [5], which is the central theme of this paper.

In Section 3, we propose an approach to training an ANN using an ant colony optimization technique, which can solve the local optima problem and improve the effectiveness of ANNs in fraud detection.

3   An Ant Colony-Based Approach for Fraud Detection

Ant Colony Optimization (ACO) is a metaheuristic technique based on the research of ant group behaviors in the natural world [5]. It imitates ants' behaviors in establishing shortest paths from their nest to feeding sources and back. Individual ants are supposed to interact with each other by some chemical pheromone released by them. When an ant finds a food source, it releases chemical pheromone on the ground. The quantity of pheromone depends on the quantity and quality of the food source. The pheromone vanishes over time. When another ant looks for food, it moves in a random manner basically. However, if it detects chemical pheromone in the environment around, it will have higher probability to move toward the direction with denser pheromone. This in turn reinforces the trail with more pheromone, and attracts more ants to the food source. The indirect communication between the ants via the pheromone trails allows them to find shortest paths between their nest and food sources. The collective behavior brings ACO with characteristics of positive feedback, parallel computing, robustness, and global optimization [5].

The proposed approach to fraud detection utilizes ant colony algorithm to optimize ANN settings. It actually includes three steps:

1.   initial authentication (often by means of PINs and passwords)
2.   automatic fraud detection by ANN
3.   further review by human users

Step 1 is a preliminary process to check the user identity and authentication. It can use PINs, passwords, or IP addresses in the Internet; fingerprint in some terminals; and SIM card identity or calling site in telecommunication. This step can reveal a large number of identity misuse problems (such as identity theft).

Step 2 is the core part of this approach. While step 1 is a signature-based analysis process, step 2 is a behavior-based process that needs more intelligent methods to analyze the behavior data. Users' transaction data (like telecom subscribers' calling detail records) are collected and analyzed in order to build up their profiles (or

behavioral models). If any abnormal transaction occurs, the case will be highlighted for manual review in step 3.

However, "abnormal" transactions are just potential fraudulent events, and sometimes they are just occasional usage of legal users. So in step 3, those abnormal transactions detected in step 2 are sent to the relevant employees for further review. Thus this three-step approach forms a system approach that provides overall controls to the person in charge of fraud detection.

In step 2, ANN is the major analysis method we described in the paper. We use the classic ANN model with K input nodes, H hidden nodes and J output nodes. Suppose a set of T training samples is available, the target is to optimize the error function:

$$E(w) = \frac{1}{2} \sum_{i=1}^{T} \sum_{j=1}^{J} (T_{ij} - O_{ij})^2 \tag{1}$$

where T_{ij} is the jth element of the ith target vector while O_{ij} is the jth element of the ith output vector.

The major steps of our algorithm are shown as below:

```
/* initialization */
represent the problem in the form of a weighted connected graph;
foreach edge x do
    set initial pheromone for x;
    end_foreach;

/* main processing loop */
while not reach convergence do
    foreach node n do
        randomly allocate ants to n;
        end_foreach;
    foreach ant a do
        a walks to next node depending on the probability distribution;
        compute the length of the path traveled by a;
        release a quantity amount of pheromone;
        end_foreach;
    update the phermone intensity;
    end_while;

/* output the best route and train ANN */
output the best route parameters as the initial weights of ANN;
use a BP algorithm to train ANN;
```

Due to the characteristics of ACO, the proposed algorithm can overcome the local optima problem and easily reach the global optimum. This helps to improve the performance of ANN in fraud detection.

4 A Prototype: Business Intelligent Fraud Management System (BiFMS)

The proposed approach is implemented in the Business Intelligent Fraud Management System (BiFMS) for telecommunication operators. BiFMS is a specially designed system that has a standard interface with billing systems. It collects subscriber usage data (for example, calling detail record, billing record, and payment record) and analyzes them with business intelligent algorithms to extract behavior models. The system can be used in preventing, detecting, and predicting fraud behaviors of the subscribers. In BiFMS, advanced business intelligence methods and algorithms (for example, ANN and rough fuzzy set) are used for profiling, detecting, and predicting.

The BiFMS consists of six components:

(a) **Data Collection and Processing:** Data collection and processing is a module that collects data from the billing system of telecommunication network, processes them into a certain format and stores them into the database. The data includes, for example, subscriber information, call site and operator information, service and account information, phone information, calling detail records, billing records, and payment record.

(b) **Profiling Module:** Profiling module uses statistical and advanced business technologies (including ANN, ACO, genetic programming (GP), and rough fuzzy set) to extract behavior model from subscriber usage data such as calling detail record and billing record. The resultant behavior model will be used in detection and prediction. A screenshot of customer analysis interface in profiling module is shown as Fig. 1.

Fig. 1. A screenshot of customer analysis interface in profiling module

(c) **Fraud Detector:** Fraud detector module is the heart of the fraud management system. This module can be divided into three procedures: Fraud detection, Subscriber scoring and Alert. Fraud detection uses rule-based detector and intelligent detector. Rule-based detector use simple rules to detect many simple fraudulent, such as SIM stuffing and SIM cloning. Intelligent detector uses business intelligent technologies (for example, ANN, ACO, genetic programming, and rough fuzzy set) to identify and detect complicated frauds. Detectors for different kinds of frauds are independent to each other, so the user can decide which package should be added to the system according to the statistics data of a certain area. When the detector finds a suspicious call, it will subtract a certain points from the total score of a phone, called "scoring". The score represents the risk of a subscriber being a fraudster. The score of every phone will be checked to see if the phone needs to be put into the blacklist. When a fraud is detected, an alert will be sent both to the operator through graphical user interface and subscriber with SMS, Email or other possible methods.

(d) **Case Manager:** Case manager is a module that builds and tracks cases, and will give statistics information about fraud types to the operator. When an alert was generated, the case builder will find all the records and alerts that might have relations with this alert and gather them together to form a case. The case tracker will integrate all cases of the same phone or patterns to form a case queue. The case queue will provides statistics information of frauds and will help operators to find complicated fraud types or combined crime.

(e) **Report Manager:** Report manager is a module that presents the basic data and other advanced data to the operator or customer for various purposes, such as self care, basic data view, and service plan evaluation. Intelligent report manager gives users the ability to present the information they are interested in with the format they want. A screenshot of report interface is shown as Fig. 2. Possible fraud alerts are also highlighted with red color.

Fig. 2. A screenshot of report interface with fraud alerts

(f) System Administration: System administration is the module that controls the access to the system, generates log files of every user and helps supervisor to configure the system and scheduling tasks. It provides an overall assessment of the risk and security level of the company business environment. A screenshot of system administration interface is shown as Fig. 3.

Fig. 3. A screenshot of system administration interface

Inside this system, we compare the ant colony-based algorithm with traditional BP algorithm. As shown in Table 1, compared with BP algorithm, the ant colony algorithm has better performance in both global optimization rate and convergence time.

Table 1. Comparison between BP algorithm and ant colony algorithm

	Global optimization rate	Average number of convergence steps	Average CPU time
BP algorithm	70%	1035	3.5 sec
Ant colony algorithm	95%	783	2.9 sec

5 Summary

In this paper, a novel approach to business fraud detection is proposed. The core part of this approach is based on the utilization of artificial neural networks which is trained by an ant colony algorithm. The approach can easily reach a global optimization. It is implemented in the Business Intelligent Fraud Management System (BiFMS)

which is designed for telecommunication operators. It can help telecommunication operators to detect subscription frauds. It can be generalized to other areas such as credit card frauds.

References

1. Bihina Bella, M.A., Eloff, J.H.P., Olivier, M.S.: A Fraud Management System Architecture for Next-Generation Networks. Forensic Science International 185, 51–58 (2009)
2. Burge, P., Shawe-Taylor, J.: An Unsupervised Neural Network Approach to Profiling the Behavior of Mobile Phone Users for Use in Fraud Detection. Journal of Parallel and Distributed Computing 61, 915–925 (2001)
3. Cerullo, M.J., Cerullo, V.: Using Neural Networks to Predict Financial Reporting Fraud: Part 1. Computer Fraud and Security 1999, 14–17 (1999)
4. Chen, H.J., Huang, S.Y., Kuo, C.L.: Using the Artificial Neural Network to Predict Fraud Litigation: Some Empirical Evidence from Emerging Markets. Expert Systems with Applications 36, 1478–1484 (2009)
5. Dorigo, M., Stutzle, T.: Ant Colony Optimization. MIT Press, Cambridge (2004)
6. Estévez, P.A., Held, C.M., Perez, C.A.: Subscription Fraud Prevention in Telecommunications Using Fuzzy Rules and Neural Networks. Expert Systems with Applications 31, 337–344 (2006)
7. Huang, S., Tsai, C., Yen, C., Cheng, Y.: A Hybrid Financial Analysis Model for Business Failure Prediction. Expert Systems with Applications 35, 1034–1040 (2008)
8. Khashei, M., Hejazi, S.R., Bijari, M.: A New Hybrid Artificial Neural Networks and Fuzzy Regression Model for Time Series Forecasting. Fuzzy Sets and Systems 159, 769–786 (2008)
9. Kim, G.: Sarbanes-Oxley, Fraud Prevention, and IMCA: A Framework for Effective Controls Assurance. Computer Fraud and Security, 12–16 (2003)
10. Mercer, C.J.: Fraud Detection via Regression Analysis. Computers and Security 9, 331–338 (1990)
11. Pollard, C.: Telecom Fraud: The Cost of Doing Nothing Just Went Up. Computers and Security 24, 437–439 (2005)
12. Quah, T.S., Sriganesh, M.: Real-Time Credit Card Fraud Detection Using Computational Intelligence. Expert Systems with Applications 35, 1721–1732 (2008)
13. Sexton, R.S., Dorsey, R.E.: Reliable Classification Using Neural Networks: A Genetic Algorithm and Backpropagation Comparison. Decision Support Systems 30, 11–22 (2000)
14. Shawe-Taylor, J., Howker, K., Burge, P.: Detection of Fraud in Mobile Telecommunications. Information Security Technical Report 4, 16–28 (1999)
15. Viaene, S., Ayuso, M., Guillen, M., Gheel, D.V., Dedene, G.: Strategies for Detecting Fraudulent Claims in the Automobile Insurance Industry. European Journal of Operational Research 176, 565–583 (2007)
16. Yen, C.: Warning Signals for Potential Accounting Frauds in Blue Chip Companies: An Application of Adaptive Resonance Theory. Information Sciences 177, 4515–4525 (2007)

Markov Chain Monte Carlo on Asymmetric GARCH Model Using the Adaptive Construction Scheme

Tetsuya Takaishi

Hiroshima University of Economics,
731-0192 Hiroshima, Japan
takaishi@hiroshima-u.ac.jp

Abstract. We perform Markov chain Monte Carlo simulations for a Bayesian inference of the GJR-GARCH model which is one of asymmetric GARCH models. The adaptive construction scheme is used for the construction of the proposal density in the Metropolis-Hastings algorithm and the parameters of the proposal density are determined adaptively by using the data sampled by the Markov chain Monte Carlo simulation. We study the performance of the scheme with the artificial GJR-GARCH data. We find that the adaptive construction scheme samples GJR-GARCH parameters effectively and conclude that the Metropolis-Hastings algorithm with the adaptive construction scheme is an efficient method to the Bayesian inference of the GJR-GARCH model.

Keywords: Markov Chain Monte Carlo, Bayesian inference, GJR-GARCH model, Metropolis-Hasting algorithm.

1 Introduction

Price returns of financial assets such as stock indexes, exchange rates show various interesting properties which can not be derived from a simple assumption that the price returns follow the geometric Brownian motion. Those properties are now classified as stylized facts [1,2]. Typical examples of the stylized facts are (i) fat-tailed distribution of return (ii) volatility clustering (iii) slow decay of the autocorrelation time of the absolute returns.

In empirical finance the volatility is an important quantity to measure risk. In order to forecast future volatility it is important to make models which mimic the properties of the volatility. The most successful model is the Generalized Autoregressive Conditional Heteroscedasticity (GARCH) model by Engle [3] and Bollerslev [4], which can capture the property of volatility clustering and shows fat-tailed distributions.

In the original GARCH model [3,4], the volatility process is symmetric under positive and negative observations. However the volatility often shows higher response against negative shocks, which was first observed by Black [5]. This phenomenon is called the leverage effect. In order to incorporate this asymmetry

effect into the model, some extended GARCH models have been proposed [6,7,8, 9,10,11]. In this study we use the GJR-GARCH model by Glosten *et al.* [7] which introduces the asymmetry into the volatility process by an artificial indicator function.

In order to infer model parameters from financial data we employ Markov Chain Monte Carlo (MCMC) methods based on the Bayesian inference. Since there is no unique way to implement the MCMC scheme various MCMC methods for the GARCH models have been proposed [12,13,14,15,16,17,18,19]. In a survey on the MCMC methods of the GARCH models [17] it is shown that Acceptance-Rejection/ Metropolis-Hastings (AR/MH) algorithm works better than other algorithms. In the AR/MH algorithm the proposal density is assumed to be a multivariate Student's t-distribution and the parameters to specify the distribution are estimated by the Maximum Likelihood (ML) technique. Recently an alternative method to estimate those parameters without relying on the ML technique was proposed [20,21,22]. In the method the parameters are determined by using the data generated by an MCMC method and updated adaptively during the MCMC simulation. We call this method "adaptive construction scheme".

The adaptive construction scheme was tested for GARCH and QGARCH models [20,21,22] and it is shown that the adaptive construction scheme can significantly reduce the correlation between sampled data. In this study we apply the adaptive construction scheme to the GJR-GARCH model and study the efficiency of the adaptive construction scheme.

2   GJR-GARCH Model

The GJR-GARCH model [7] is given by

$$y_t = \sigma_t \epsilon_t, \tag{1}$$

$$\sigma_t^2 = \omega + \alpha y_{t-1}^2 + \lambda \{1_{y_{t-1}<0}\} y_{t-1}^2 + \beta \sigma_{t-1}^2, \tag{2}$$

where ϵ_t is an independent normal error $\sim N(0,1)$ and y_t are observations. The indicator function $1_{y_{t-1}<0}$ is 1 if $y_{t-1} < 0$, and otherwise 0. This indicator function introduces the asymmetry in the time series and it generates higher volatilities after negative shocks than positive ones.

3   Bayesian Inference

Using the Bayes' rule the posterior density $\pi(\theta|y)$ with n observations $y = (y_1, y_2, \ldots, y_n)$ is given by

$$\pi(\theta|y) \propto L(y|\theta)\pi(\theta), \tag{3}$$

where $L(y|\theta)$ is the likelihood function. $\pi(\theta)$ is the prior density for θ. The functional form of $\pi(\theta)$ is not known a priori. In this study we assume that the prior density $\pi(\theta)$ is constant.

The likelihood function for the GJR-GARCH model is given by

$$L(y|\theta) = \Pi_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma_t^2}} \exp\left(-\frac{y_t^2}{2\sigma_t^2}\right), \tag{4}$$

where $\theta = (\alpha, \beta, \omega, \lambda)$ stands for the GJR-GARCH parameters. σ_t^2 is given by eq.(2).

Using $\pi(\theta|y)$ the GJR-GARCH parameters are inferred as the expectation values given by

$$\langle\theta\rangle = \frac{1}{Z} \int \theta\pi(\theta|y)d\theta, \tag{5}$$

where $Z = \int \pi(\theta|y)d\theta$ is a normalization constant irrelevant to MCMC estimations.

3.1   Markov Chain Monte Carlo

In general it is difficult to evaluate eq.(5) analytically. The MCMC technique gives a method to estimate eq.(5) numerically. The basic procedure of the MCMC method is as follows. First we sample θ drawn from a probability distribution $\pi(\theta|y)$. Sampling is done by a technique which produces a Markov chain. After sampling some data, we evaluate the expectation value as an average value over the sampled data $\theta^{(i)}$,

$$\langle\theta\rangle = \lim_{k\to\infty} \frac{1}{k} \sum_{i=1}^{k} \theta^{(i)}, \tag{6}$$

where k is the number of the sampled data. The statistical error for k independent data is proportional to $\frac{1}{\sqrt{k}}$. In general, however, the data generated by the MCMC process are correlated. As a result the statistical error will be proportional to $\sqrt{\frac{2\tau}{k}}$ where τ is the autocorrelation time among the sampled data.

3.2   Metropolis-Hastings Algorithm

The MH algorithm [24] is a generalized version of the Metropolis algorithm [23]. Let us consider to generate data x from a probability distribution $P(x)$. The MH algorithm consists of the following steps. First starting from x, we propose a candidate x' which is drawn from a certain probability distribution $g(x'|x)$ which we call proposal density. Then we accept the candidate x' with a probability $P_{MH}(x, x')$ as the next value of the Markov chain:

$$P_{MH}(x, x') = \min\left[1, \frac{P(x')}{P(x)} \frac{g(x|x')}{g(x'|x)}\right]. \tag{7}$$

If x' is rejected we keep the previous value x. We repeat these steps.

When the proposal density does not depend on the previous value, i.e. $g(x|x') = g(x)$ we obtain

$$P_{MH}(x, x') = \min\left[1, \frac{P(x')}{P(x)} \frac{g(x)}{g(x')}\right]. \tag{8}$$

4   Adaptive Construction Scheme

The efficiency of the MH algorithm depends on how we choose the proposal density. By choosing an adequate proposal density for the MH algorithm one can reduce the correlation between the sampled data. The posterior density of GARCH parameters often resembles to a Gaussian-like shape. Thus one may choose a density similar to a Gaussian distribution as the proposal density. Such attempts have been done by Mitsui, Watanabe [16] and Asai [17]. They used a multivariate Student's t-distribution in order to cover the tails of the posterior density and determined the parameters to specify the distribution by using the ML technique. Here we also use a multivariate Student's t-distribution but determine the parameters through MCMC simulations.

The (p-dimensional) multivariate Student's t-distribution is given by

$$
g(\theta) = \frac{\Gamma((\nu + p)/2)/\Gamma(\nu/2)}{\det \Sigma^{1/2}(\nu\pi)^{p/2}}
$$
$$
\times \left[1 + \frac{(\theta - M)^t \Sigma^{-1}(\theta - M)}{\nu}\right]^{-(\nu+p)/2}, \tag{9}
$$

where θ and M are column vectors,

$$
\theta = \begin{bmatrix} \theta_1 \\ \theta_2 \\ \vdots \\ \theta_p \end{bmatrix}, M = \begin{bmatrix} M_1 \\ M_2 \\ \vdots \\ M_p \end{bmatrix}, \tag{10}
$$

and $M_i = E(\theta_i)$. Σ is the covariance matrix defined as

$$
\frac{\nu\Sigma}{\nu - 2} = E[(\theta - M)(\theta - M)^t]. \tag{11}
$$

ν is a parameter to tune the shape of Student's t-distribution. When $\nu \to \infty$ the Student's t-distribution goes to a Gaussian distribution. At small ν Student's t-distribution has a fat-tail. Since eq.(9) is independent of the previous value of θ, eq.(8) is used in the MH algorithm.

For the GJR-GARCH model, $p = 4$, $\theta = (\theta_1, \theta_2, \theta_3, \theta_4) = (\alpha, \beta, \omega, \lambda)$, and Σ is a 4×4 matrix. The unknown parameters in M and Σ are determined by using the data obtained from MCMC simulations. First we make a short run by the Metropolis algorithm and accumulate some data. Then we estimate M and Σ. Note that there is no need to estimate M and Σ accurately. Second we perform an MH simulation with the proposal density of eq.(9) with the estimated M and Σ. After accumulating more data, we recalculate M and Σ, and update M and Σ of eq.(9). By doing this, we adaptively change the shape of eq.(9) to fit the posterior density.

5 Numerical Study

In this section we study the performance of the adaptive construction scheme by using artificial GJR-GARCH data generated with known parameters. The GJR-GARCH parameters are set to $\alpha = 0.03$, $\beta = 0.85$, $\omega = 0.05$ and $\lambda = 0.1$. We have generated 2000 data by the GJR-GARCH process with these parameters. Fig.1 shows the time series of the 2000 data.
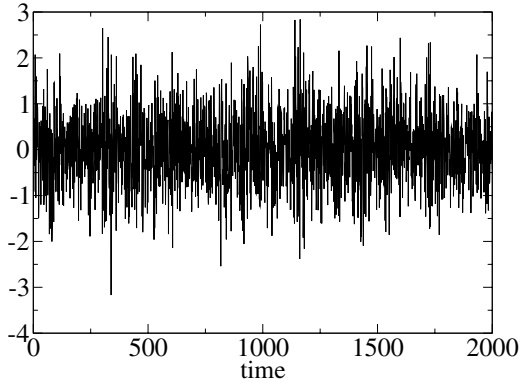
Fig. 1. The artificial times series generated by the GJR-GARCH process with $\alpha = 0.03$, $\beta = 0.85$, $\omega = 0.05$ and $\lambda = 0.1$

The MCMC method with the adaptive construction scheme is implemented as follows. First we make a short run by a standard Metropolis algorithm. The first 5000 data are discarded as burn-in process or thermalization. Then we accumulate 1000 data for M and Σ estimations. The estimated M and Σ are substituted to $g(\theta)$. In this study we take $\nu = 10$. We re-start a run by the MH algorithm with $g(\theta)$. Every 1000 updates we re-calculate M and Σ and update $g(\theta)$. We accumulate 100000 data for analysis.

For comparison we also use a standard Metropolis algorithm to infer the GJR-GARCH parameters and accumulate 100000 data for analysis. In this study the Metropolis algorithm is implemented as follows. We draw a candidate θ' by adding a small random value $\delta\theta$ to the present value θ:

$$\theta' = \theta + \delta\theta, \tag{12}$$

where $\delta\theta = d(r - 0.5)$. r is a uniform random number in $[0, 1]$ and d is a constant to tune the Metropolis acceptance. We choose d so that the acceptance becomes greater than 50%.

The results of the parameters inferred by the MH algorithm with the adaptive construction scheme and the Metropolis algorithm are summarized in Table 1. We see that both algorithms well reproduce the values of the input parameters within the standard deviation. Furthermore the obtained values and the standard
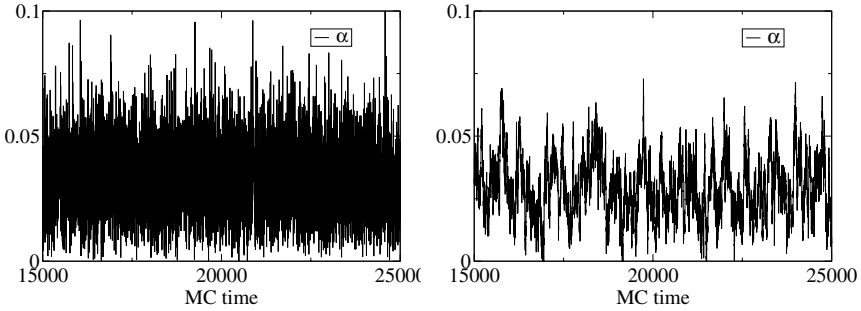
Fig. 2. Monte Carlo history of α from the MH algorithm with the adaptive construction scheme(left) and the Metropolis algorithm(right)

Table 1. Results of GJR-GARCH parameters. SD and SE stand for standard deviation and statistical error respectively.

	α	β	ω	λ
true	0.03	0.85	0.05	0.1
Adaptive	0.03285	0.85540	0.04522	0.08719
SD	0.0015	0.040	0.019	0.026
SE	0.00011	0.00025	0.00011	0.00066
2τ	2.8 ± 0.2	3.3 ± 0.7	4.6 ± 0.8	2.6 ± 0.2
Metropolis	0.0323	0.855	0.0454	0.0895
SD	0.0015	0.038	0.018	0.026
SE	0.0007	0.004	0.0018	0.0015
2τ	320 ± 100	1050 ± 350	990 ± 330	350 ± 110

deviations from both algorithms coincide well each other. This observation is not surprising because both algorithms are performed using the same artificial data and thus the posterior density is the same for both.

Fig.2 shows Monte Carlo time histories of the sampled α from the adaptive construction scheme (left) and Metropolis algorithm(right). We see that the data sampled by the Metropolis algorithm are substantially correlated. The similar behavior is also seen for the sampled data for other parameters.

The correlations between the data can be measured by the autocorrelation function (ACF). The ACF of certain successive data $\theta^{(i)}$ is defined by

$$ACF(t) = \frac{\frac{1}{N}\sum_{j=1}^{N}(\theta^{(j)} - \langle\theta\rangle)(\theta^{(j+t)} - \langle\theta\rangle)}{\sigma_\theta^2}, \tag{13}$$

where $\langle\theta\rangle$ and σ_θ^2 are the average value and the variance of θ respectively.

Fig.3 compares the ACF of α sampled from the adaptive construction scheme (left) and the Metropolis algorithm(rigth). The ACF of the adaptive construction scheme decreases quickly as Monte Carlo time increases. On the other hand the ACF of the Metropolis algorithm decreases very slowly, which indicates that the correlation between the sampled data is very large.
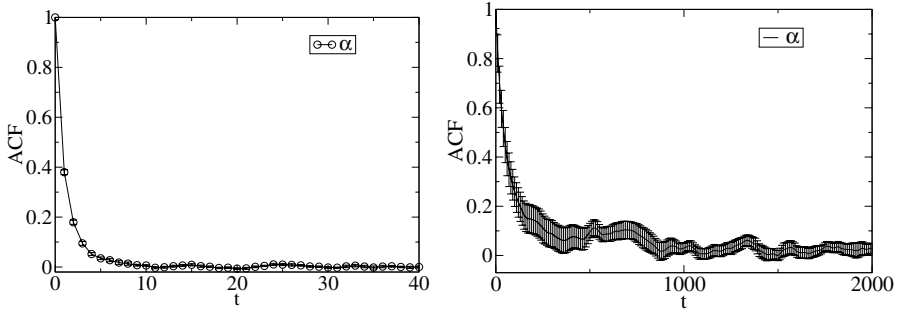
Fig. 3. Autocorrelation function of α sampled by the adaptive construction scheme(left) and the Metropolis algorithm(right)
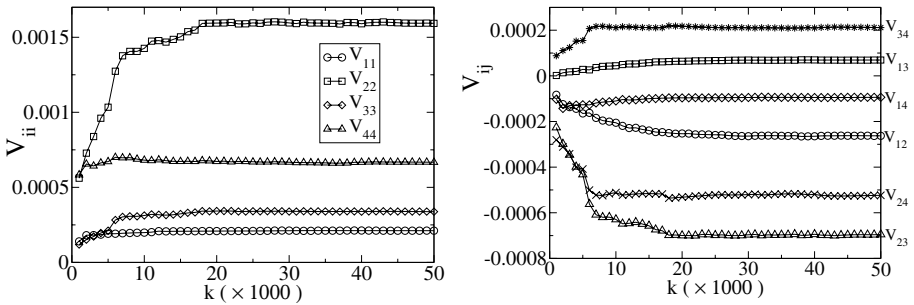
Fig. 4. The diagonal(left) and off-diagonal(right) elements of V as a function of the data size

To quantify the correlation between the data we calculate the autocorrelation time (ACT) τ defined by

$$\tau = \frac{1}{2} + \sum_{i=1}^{\infty} ACF(i). \tag{14}$$

2τ is also called "inefficiency factor". Results of τ are summarized in Table 1. We find that the ACT from the adaptive construction scheme have much smaller τ than those from the Metropolis simulations. For instance τ of α parameter from the adaptive construction scheme is decreased by a factor of 100 compared to that from the Metropolis algorithm. Furthermore the values of τ are about $2-3$ which are similar to the results of the AR/MH algorithm [16,17]. These results prove that the adaptive construction scheme is an efficient algorithm for sampling de-correlated data. The differences in τ also explain that the statistical errors from the adaptive construction scheme are much smaller than those from the Metropolis simulations.

Fig. 4 shows the convergence property of the diagonal and off-diagonal elements of the covariance matrix V. Here V is defined by $V = E[(\theta - M)(\theta - M)^t]$.
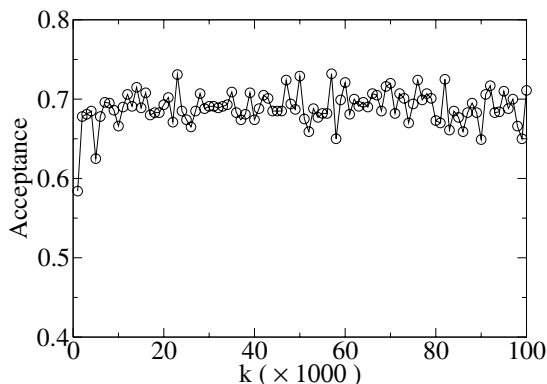
Fig. 5. Acceptance at MH step with the adaptive proposal density

We see that all the diagonal elements of V quickly converge to certain values as the simulations are proceeded.

Fig. 5 shows the acceptance at the MH algorithm with the adaptive proposal density of eq. (9). Each acceptance is calculated every 1000 updates and the calculation of the acceptance is based on the latest 1000 data. Surprisingly at the first stage of the simulation the acceptance already reaches a plateaus of about 70%. This indicates that the accuracy of the parameters of the proposal density are high enough for the MH algorithm already at the first stage.

6   Conclusions

We have performed the MCMC simulations of the Bayesian inference on the GJR-GARCH model which is one of asymmetric GARCH models. The MCMC was implemented by the MH algorithm with the Student's t-distribution and the parameters of the Student's t-distribution were updated adaptively during the simulations. The autocorrelation times of the data sampled by the adaptive construction scheme are found to be very small. The obtained values of τ are similar to those of the AR/MH algorithm in [16,17]. Thus the adaptive construction scheme has the similar efficiency with the AR/MH algorithm. It is concluded that the adaptive construction scheme is also an efficient MCMC technique for the Bayesian inference of the GJR-GARCH model. The previous studies on the GARCH and QGARCH models [20,21,22] also reached the same conclusion.

It is also found that the acceptance of the MH algorithm quickly reaches a plateau of about 70% at the beginning of the simulation. This indicates that the parameters of the Student's t-distribution for the MH algorithm are calculated precisely enough by using the small data sampled at the beginning of the simulation. This observation may suggest that one can stop updating the parameters at some point of the simulation and use the same proposal density after that.

Acknowledgments

References

1. Mantegna, R., Stanley, H.E.: Introduction to Econophysics. Cambridge University Press, Cambridge (1999)
2. Cont, R.: Empirical Properties of Asset Returns: Stylized Facts and Statistical Issues. Quantitative Finance 1, 223–236 (2001)
3. Engle, R.F.: Autoregressive Conditional Heteroskedasticity with Estimates of the Variance of the United Kingdom inflation. Econometrica 60, 987–1007 (1982)
4. Bollerslev, T.: Generalized Autoregressive Conditional Heteroskedasticity. Journal of Econometrics 31, 307–327 (1986)
5. Black, F.: The Pricing of Commodity Contracts. Journal of Financial Economics 3(1-2), 167–179 (1976)
6. Nelson, D.B.: Conditional Heteroskedasticity in Asset Returns: A New Approach. Econometrica 59, 347–370 (1991)
7. Glosten, L.R., Jaganathan, R., Runkle, D.E.: On the Relation Between the Expected Value and the Volatility of the Nominal Excess on Stocks. Journal of Finance 48, 1779–1801 (1993)
8. Ding, Z., Granger, C.W.J., Engle, R.F.: A Long Memory Property of Stock Market Returns and a New Model. Journal of Empirical Finance 1, 83–106 (1993)
9. Zakoïan, M.: Threshold Heteroscedastic Models. Journal of Economic Dynamics and Control 18, 931–955 (1994)
10. Engle, R.F., Ng, V.: Measuring and testing the impact of news on volatility. Journal of Finance 48, 1749–1778 (1993)
11. Sentana, E.: Quadratic ARCH models. Review of Economic Studies 62, 639–661 (1995)
12. Bauwens, L., Lubrano, M.: Bayesian inference on GARCH models using the Gibbs sampler. Econometrics Journal 1, c23–c46 (1998)
13. Kim, S., Shephard, N., Chib, S.: Stochastic volatility: Likelihood inference and comparison with ARCH models. Review of Economic Studies 65, 361–393 (1998)
14. Nakatsuma, T.: Bayesian analysis of ARMA-GARCH models: Markov chain sampling approach. Journal of Econometrics 95, 57–69 (2000)
15. Vrontos, I.D., Dellaportas, P., Politis, D.N.: Full Bayesian inference for GARCH and EGARCH models. Journal of Business and Economic Statistics 18, 187–198 (2000)
16. Mitsui, H., Watanabe, T.: Bayesian analysis of GARCH option pricing models. J. Japan Statist. Soc. (Japanese Issue) 33, 307–324 (2003)
17. Asai, M.: Comparison of MCMC Methods for Estimating GARCH Models. J. Japan Statist. Soc. 36, 199–212 (2006)
18. Takaishi, T.: Bayesian Estimation of GARCH model by Hybrid Monte Carlo. In: Proceedings of the 9th Joint Conference on Information Sciences (2006), CIEF-214, doi:10.2991/jcis.2006.159

19. Ardia, D.: Financial Risk Management with Bayesian Estimation of GARCH models. Springer, Heidelberg (2008)
20. Takaishi, T.: An Adaptive Markov Chain Monte Carlo Method for GARCH Model. arXiv:0901.0992v1
21. Takaishi, T.: Bayesian Estimation of GARCH Model with an Adaptive Proposal Density. In: New Advances in Intelligent Decision Technologies. Studies in Computational Intelligence, vol. 199, pp. 635–643 (2009)
22. Takaishi, T.: Bayesian Inference on QGARCH Model Using the Adaptive Construction Scheme. In: Proceedings of 8th IEEE/ACIS International Conference on Computer and Information Science, pp. 525–529 (2009), doi:10.1109/ICIS.2009.173
23. Metropolis, N., et al.: Equations of State Calculations by Fast Computing Machines. J. of Chem. Phys. 21, 1087–1091 (1953)
24. Hastings, W.K.: Monte Carlo Sampling Methods Using Markov Chains and Their Applications. Biometrika 57, 97–109 (1970)

Robot Visual Servo through Trajectory Estimation of a Moving Object Using Kalman Filter

Min-Soo Kim[1], Ji-Hoon Koh[1], Ho Quoc Phuong Nguyen[1], and Hee-Jun Kang[2,*]

[1] Graduate School of Electrical Engineering, University of Ulsan,
680-749, Ulsan, South Korea
[2] School of Electrical Engineering, University of Ulsan,
680-749, Ulsan, South Korea
hjkang@ulsan.ac.kr

Abstract. In this paper, a robot visual servo control algorithm is proposed by combining the conventional image based robot visual servoing algorithm with a trajectory estimation algorithm of a moving object using Kalman filter. The erroneous image information of a moving object due to the imprecise camera characteristics is compensated by applying Kalman filter to the process model of a moving object. The robot visual servo control algorithm is simulated, implemented and discussed with a Samsung FARA AT-2 robot and a MV50 Camera for its effectiveness, in both cases of with/without a trajectory estimation algorithm of a moving object using Kalman filter.

Keywords: Robot Visual Servo Control, Image based Visual Servo Control, a Moving Object Model, Kalman Filter.

1   Introduction

In order to allow a robot to have more flexibility and more precision for the various tasks, the robot visual servo control (or robot visual servoing) has been investigated by many researchers[1-5]. Generally, the robot visual servo control is divided into two categories such as position based control and image based control by defining the error signal. Position based control uses the error measured in 3 dimensional Cartesian space, while image based control uses the error in 2 dimensional image plane[2].

In position based control, features are extracted from the image and used with the geometric model of the target to determine the kinematic relationship between the target and the robot. This method requires more precise models of a robot and camera system so that it needs the complicate robot and camera calibration process. In image based control, the error is measured in image plane so that the process of pose estimation is omitted. It means that this method is less sensitive to the preciseness of the robot and camera model. Therefore, the image based control method may reduce computational delay and eliminate errors in modeling and camera calibration. Even though the image based control have the advantages over the position based control as prescribed above, it still has many difficulties for successful control.

---

* Corresponding author.

Due to camera characteristics such as image signal quantization error, lens distortion, charge coupled device senor, the various noises may inevitably occur to lead the pixel level error. This noise contained signal brings the imprecise the estimation of the error. In order to relieve this problem, Chang[6] suggested 3 dimensional motion parameter estimation method for tracking a maneuvering target with Kalman filter, Broida and Chellappa[7] suggests a dynamic model of a rigid object with Kalman filter. These approaches have been done for position based control.

In this paper, a robot visual servo control algorithm is proposed by combining the conventional image based robot visual servoing algorithm[3] with a moving object model of LaValle[9] by applying the extended Kalman filter. Computer simulation and real implementation has been performed with FARA AT/2 robot and a MV50 camera for the effectiveness of this algorithm. The results are discussed in terms of the pixel error radii in the image plane from the algorithms with and without Kalman filters.

2   Image Based Visual Servo Control

In this study, the RMRC(Resolved Motion Rate Control) structure[3] is used with the image Jacobian which relates the joint angle differential to the image space differential as shown in Fig. 1. In Fig. 1, $^{t6}J_q$ and $^{t6}J_c$ are Jacobians which relate differentials of the robot joint motion and the position of the robot mounted camera to the differentials of the robot tip. $^{I}J_c$ is Jacobian which shows the differential relationship between image space and camera pose space.
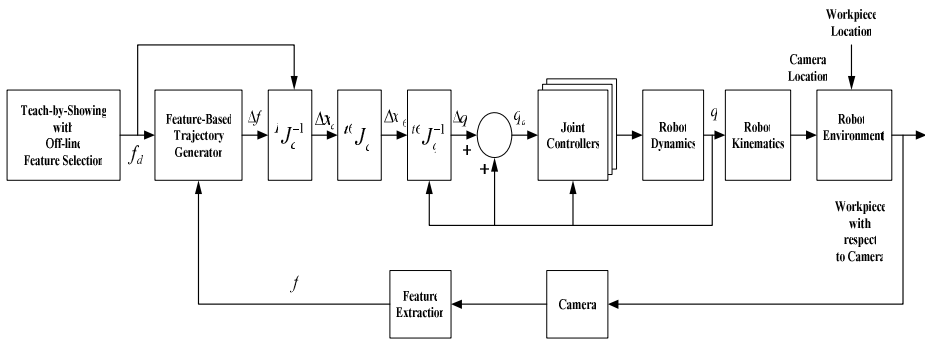
Fig. 1. The Structure of RMRC(Resolved Motion Rate Control)

3   Visual Servo Control with the Extended Kalman Filter

3.1   A Moving Object Model in Image Plane

When the velocity and acceleration of a moving object are given for the current instance, the position for the next instance in Cartesian space, (P_x, P_y), can be estimated as follows[8]

$$\hat{P}_{x+\delta t} = \hat{P}_x + \hat{V}_x \delta t + \frac{1}{2}\hat{A}_x \delta t^2 \qquad (1)$$

$$\hat{P}_{y+\delta t} = \hat{P}_y + \hat{V}_y \delta t + \frac{1}{2}\hat{A}_y \delta t^2 \qquad (2)$$

Where δt is the sampling period, and $(\hat{P}_x, \hat{P}_y)$, $(\hat{V}_x, \hat{V}_y)$, $(\hat{A}_x, \hat{A}_y)$ are the estimated position, velocity, and acceleration of the moving object, respectively.

The motion of an object displayed in the image plane can be predicted as discrete time-varying equations[9, 11] by dividing it into a linear velocity, v_k, and an angular velocity ω_k. The resulting equations are as follows:

$$\delta x_{k+\delta t,k} = v_k \delta t \cos\left(\theta_k + \frac{1}{2}\omega_k \delta t\right) \approx v_k \cos\left(\theta_k\right)\delta t - \frac{1}{2}\omega_k v_k \sin\left(\theta_k\right)\delta t^2 \qquad (3)$$

$$\delta y_{k+\delta t,k} = v_k \delta t \sin\left(\theta_k + \frac{1}{2}\omega_k \delta t\right) \approx v_k \sin\left(\theta_k\right)\delta t + \frac{1}{2}\omega_k v_k \cos\left(\theta_k\right)\delta t^2 \qquad (4)$$

$$\delta\theta_{k+\delta t,k} = \omega_k \delta t \qquad (5)$$

And also, the noises and uncertainties in the measuring the linear and angular velocities can be modeled as follows:

$$\delta v_{k+\delta t,k} = \xi_v \qquad (6)$$

$$\delta\omega_{k+\delta t,k} = \xi_\omega \qquad (7)$$

where ξ_v and ξ_ω are zero-mean Gaussian random variables.

In order to apply a model of the moving object to the Kalman filter, Eqs. (3)-(7) can be expressed in form of the discrete time state transition equations and its observation model such as

$$x_k = \Phi_{k,k-1}x_{k,k-1} + w_{k-1} \qquad (8)$$

$$z_k = H_k x_k + v_k \qquad (9)$$

where z_k, Φ_k and H_k are a measurement vector, a state transition matrix and an observation matrix, respectively. And also w_{k-1} and v_k are a state transition error and measurement noises.

Applying the Eqs. (3)-(7) to the form of Eqs. (8)-(9) gives as follows:

$$
\begin{bmatrix} x_k \\ y_k \\ \theta_k \\ v_k \\ \omega_k \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & \delta t \cos(\theta_{k-1}) & -\frac{1}{2} v_{k-1} \delta t^2 \sin(\theta_{k-1}) \\ 0 & 1 & 0 & \delta t \sin(\theta_{k-1}) & \frac{1}{2} v_{k-1} \delta t^2 \cos(\theta_{k-1}) \\ 0 & 0 & 1 & 0 & \delta t \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x_{k-1} \\ y_{k-1} \\ \theta_{k-1} \\ v_{k-1} \\ \omega_{k-1} \end{bmatrix} + \begin{bmatrix} 0 \\ 0 \\ 0 \\ \xi_v \\ \xi_\omega \end{bmatrix} \tag{10}
$$

$$
\begin{bmatrix} x_k \\ y_k \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} x_k \\ y_k \\ \theta_k \\ v_k \\ \omega_k \end{bmatrix} + \begin{bmatrix} \gamma_x \\ \gamma_y \end{bmatrix} \tag{11}
$$

3.2 State Estimation of a Moving Object by Kalman Filter

In the given state estimation problem of a moving object with Kalman Filter, the measurement vector is the position of a object in x-y plane which could be obtained from vision image. Based on this measurement, the state variables of the object as x-y position, its direction, its linear/angular velocities shown in Eqs. (10)-(11) would be estimated by using the Kalman Filter in Fig. 2.
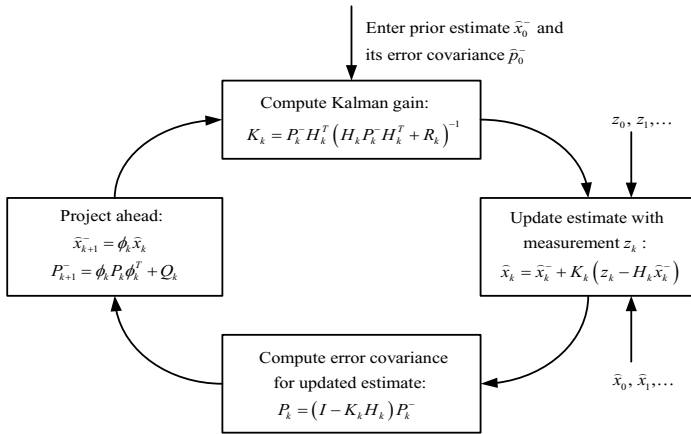
Fig. 2. Kalman Filter Block Diagram

4 Computer Simulation for Robot Visual Servo Control

In order to show the effectiveness of the state estimation with Kalman filter, a computer simulation of a 6 dof manipulator visual servo control has been performed. Its algorithm is constructed with combination of the image based robot visual servo control in Sec. 2 and the state estimation of a moving object in Sec.3. Fig. 3 shows the

simulation setup for a robot visual servo control. Gaussian random noise covariances are intentionally included to the linear velocity, v_k, and the angular velocity ω_k of the moving object trajectory. They are 1 pixel/sec and 1deg/sec, respectively. Other state transition noise covariances are assumed to be zero. And also its measurement noise covariance is 1 pixel/sec. Therefore, the resulting state transition noise covariance matrix, Q_k, and the measurement noise covariance, R_k, are given in Eq. (12) as

$$Q_k = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & \pi/180 \end{bmatrix}, \quad R_k = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \tag{12}$$
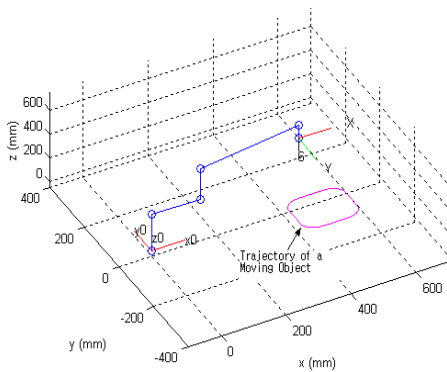
Fig. 3. A Simulation Setup for a Robot Visual Servo Control
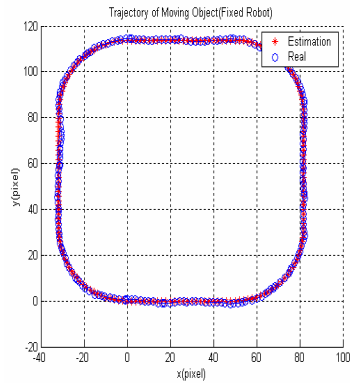
Fig. 4. The trajectory of a moving Object in Image Plane of a Fixed Camera

The considered real trajectory and its resulting estimated trajectory are shown in Fig. 4 and the differences of their states(x_k, y_k, θ_k, v_k and ω_k) as estimation errors are shown in Fig. 5. Since the observation model deals with only x_k and y_k, their estimation errors are relatively smaller as in between -2 and 2 pixel than the other states. In order to show the effectiveness of the above estimation method, a robot visual servo control simulation has been performed with and without the estimation algorithm. The simulation results with/without the estimation are shown in Fig. 6 and Fig. 7, respectively. The maximum pixel errors in the image plane for both cases are 6.6 pixels shown in Fig. 6 and 3.3 pixels shown in Fig. 7. It means that the suggested estimation has good effectiveness to reduce the tracking error under the various noisy environments.
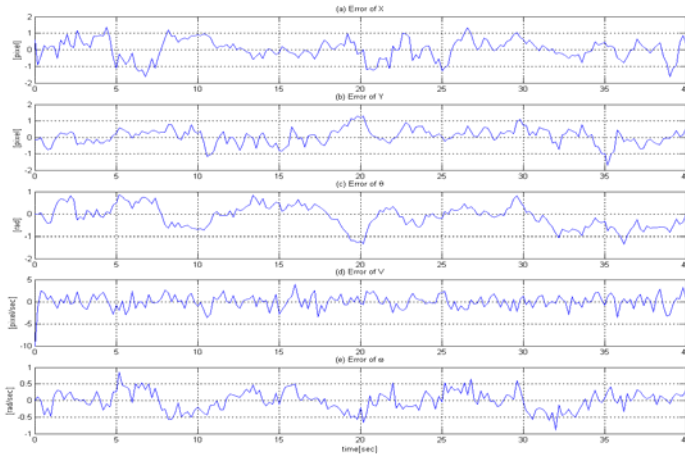
Fig. 5. The Resulting Estimation Error of Each State
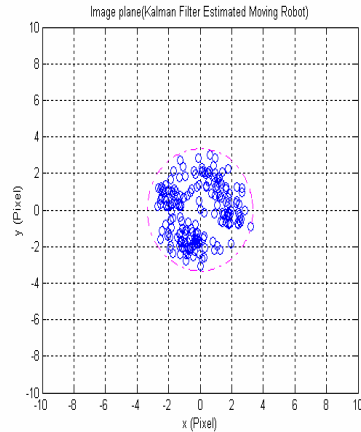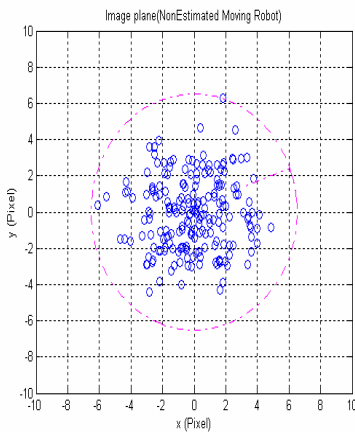
Fig. 6. Pixel Error of Robot Visual Tracking without the Estimation

Fig. 7. Pixel Error of Robot Visual Tracking with the Estimation

In order to check the robustness to the uncertainties of the system, the relationship between the above maximum error radius and the arbitrarily assigned the noise covariance value has been investigated. The covariance noise matrices of the state transition model, Q_k, and measurement model, R_k, in Eq. (12) are multiplied by weight α and the robot visual servo control simulation has been performed as α changes from 0 to 10 by 0.1. The result is shown in Fig. 8 and the slopes of the maximum error radius in both cases are comparable. It shows that moving object model based estimation algorithm is more effective in the robot visual servo control as the more measurement noise exists.

5   Real Implementation for Robot Visual Servo Control

Real Implementation of the proposed robot visual servo control algorithm has been performed like the computer simulation. The total system shown Fig. 9 consists of FARA AT 2 robot, MV50 CCD Camera with 320 by 240 resolution and Meteor II frame grabber and MMC(Multi Motion Controller) based PC controller. In order to reduce the light effect to the camera image, the black background and a white object were used. 8 bit gray level was used to reduce the image processing time. In real implementation, image processing speed is 5~8 frames/sec and its resulting visual control frequencies are 4~7 Hz for no estimation case and 2~5 Hz for the estimation case.
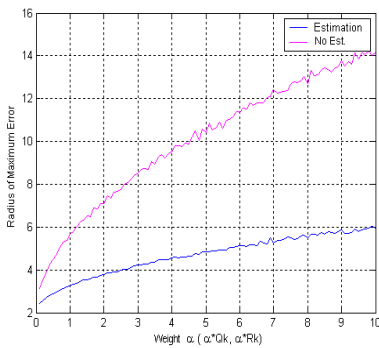
Fig. 8. Maximum Error Radius as the change of the weight α
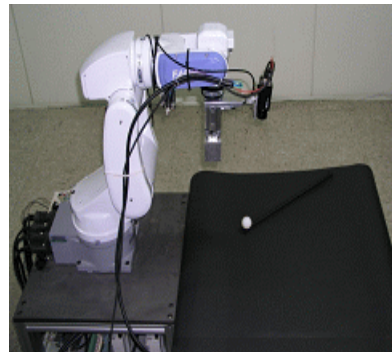
Fig. 9. Real Implementation of the proposed Robot Visual Servo
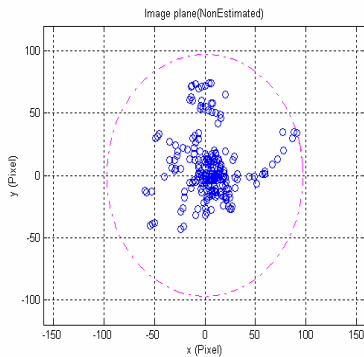
Fig. 10. Pixel Error of Robot Visual Tracking without the Estimation
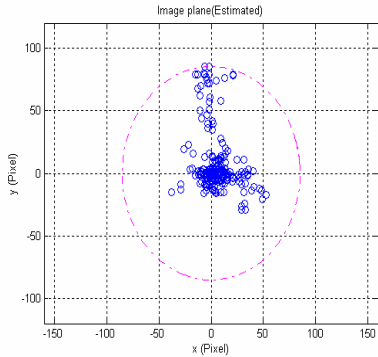
Fig. 11. Pixel Error of Robot Visual Tracking with the Estimation

The maximum error radii of the both cases as seen in the computer simulation are shown in Fig. 10 and Fig. 11. The results in both cases are 97.2 pixels and 85.2 pixels, respectively. The maximum pixel error is greatly increased when compared in computer simulation case. It is reasoned that the real time control problem might be more serious than the measurement noise. The algorithm without the estimation could more or less relieves the additional computational effort. However, the algorithm with the estimation algorithm is still more effective for reducing the maximum error radius by about 12 pixels in real implementation.

6   Conclusion

In this paper, a robot visual servo control algorithm is proposed by combining the conventional image based robot visual servoing algorithm with a trajectory estimation algorithm of a moving object using Kalman filter. The erroneous image information of a moving object due to the imprecise camera characteristics is compensated by applying Kalman filter to the process model of a moving object.

From the simulation of the robot visual servo control algorithm, the pixel error due to the measurement noise could be considerably reduced by using Kalman Filter based on the moving object model. And also, the proposed algorithm is shown to be more effective as the measurement noise grows bigger. Real implementation results shows that the algorithm overcomes the additional computational efforts and has more error reducing capability.

Acknowledgement

References

1. Sutanto, H., Sharma, R., Varma, V.: Image based Autodocking without Calibration. In: Proceedings of the 1997 IEEE International Conference on Robotics and Automation, pp. 974–979 (1997)
2. Corke, P.I.: Visual control of robot manipulators – a review. In: Hashimotor, K. (ed.) Visual Servoing, pp. 1–32. World Scientific, Singapore (1993)
3. George Lee, C.S., Feddema, J.T., Robert, O.M.: Feature-Based Visual Servoing of Robotic Systems. In: Visual Servoing. World Scientific Series in Robotics and Automated Systems, pp. 105–138 (199)
4. Weiss, L., Sanderson, N.C.: Dynamic Sensor-based Control of Robots with Visual Feedbace. IEEE Trans. on Robotics and Automation. 3(5), 404–417 (1987)
5. Thuilot, B., Martinet, P., Cordesses, L., Gallice, J.: Position based Visual Servoing: Keep the Object in the Field of Vision. In: IEEE Int. Conf. on Robotics and Automation, pp. 1624–1629 (2002)
6. Chang, K.C., Lee, H.J., Chung, C.G.: Adaptively Estimation Motion Parameters for Tracking a Maneuvering Target Using Image. Journal of Robotics & Automation 3, 43–48 (1989)

7. Broida, R.J., Chellappa, R.: Estimating the Kinematics and Structure of a Rigid Object from a Sequence of Monocular Images. IEEE Trans. Patten Analysis and Machine Intelligence 3(6), 497–513 (1991)
8. Berg, R.F.: Estimation and Prediction for Maneuvering Trget Trajectories. IEEE Transactions on Robotics and Automation AC-28, 294–304 (1983)
9. LaValle, S.M., Sharma, R.: On Motion Planning in Changing, Partially Predictable Environments. The International Journal of Robotics Research 16(6), 775–805 (1997)
10. Feddema, J.T., Lee, C.S.G.: Adaptive Image Feature Prediction and Control for Visual Tracking with a Hand-Eye Coordinated Camera. IEEE Trans. Syst. Man Cybernet 20, 1172–1183 (1990)
11. Park, J.-H., Yun, J.-M., Lee, J.-M.: Trajcetory Estimation of a Moving Object using Kalman filter and Kohonen networks. Robotica 25, 567–574 (2007)

Erratum to: Quantum Quasi-Cyclic Low-Density Parity-Check Codes

Dazu Huang[1,2], Zhigang Chen[1], Xin Li[1,2], and Ying Guo[1]

[1] School of Information Science and Engineering, Central South University, Changsha, 410083 China
[2] Department of Information Management, Hunan College of Finance and Economics, Changsha, 410205, China

DOI 10.1007/978-3-642-04070-2_119

The paper entitled "Quantum Quasi-Cyclic Low-Density Parity-Check Codes'', on pages 18-27 of this volume, has been retracted, because a large portion of the contents had been taken from the paper "Quantum Quasi-Cyclic LDPC Codes'' by Manabu Hagiwara and Hideki Imai, published by arXiv.org in the year 2007.

Author Index