Computers and Mathematics with Applications 61 (2011) 2101–2106

Contents lists available at ScienceDirect

Computers and Mathematics with Applications

journal homepage: www.elsevier.com/locate/camwa

An embedded system for real-time facial expression recognition based on the extension theory

Neng-Sheng Pai *, Shih-Ping Chang

Department of Electrical Engineering, National Chin-Yi University of Technology, Taichung, Taiwan

ARTICLE INFO

Keywords:
Embedded system
Face detection
Extension theory
Facial expression recognition

ABSTRACT

This paper presents a novel facial expression recognition scheme based on extension theory. The facial region is detected and segmented by using feature invariant approaches. Accurate positions of the lips are then extracted as the features of a face. Next, based on the extension theory, basic facial expressions are classified by evaluating the correlation functions among various lip types and positions of the corners of the mouth. Additionally, the proposed algorithm is implemented using the XScale PXA270 embedded system in order to achieve real-time recognition for various facial expressions. Experimental results demonstrate that the proposed scheme can recognize facial expressions precisely and efficiently.

1. Introduction

Continuous technological advances have significantly accelerated the processing speed of embedded systems [1], thus facilitating image processing and facial expression recognition technologies [2]. Previously impossible complicated computer-based calculations are common, allowing computers to analyze the facial expressions in real time to perform face-tracing operations [3].

Among the several face detection methods include image analysis, face template matching [4,5] and neural networks [6]. The face template approach focuses on drawing out the Eigen faces from various individuals and then building templates to proceed with matching after reducing the number of facial images using the Eigen face method. However, the computation is time consuming and the data storage requirements are enormous. The neural network approach is adopted to classify the input, which requires numerous computer-based calculations and makes the obtained data highly reliable. Given the above limitations, this work adopts a faster color-analyzing approach that distinguishes complexions from backgrounds in different color spaces.

Facial expressions have received considerable interest for quite some time [2–7]. Facial expression recognition can be summarized in a few methods, in which the well-known facial action coding system (FACS) of Ekman [7] is a notable example. Other approaches capable of recognizing facial actions include neural networks, fuzzy inference, extension deduction, and decision tree. In FACS, 44 action units are used to represent all facial actions and determine what facial expression it is. Although a neural network determines the facial expression via a Black Box, the input must be trained beforehand. The fuzzy inference and decision tree determine the facial expression through the permutation and combination of facial feature data.

Most approaches for facial expression recognition achieve recognition based on fixed-sized faces, monotone backgrounds, or manual face-tracing technology. Few studies incorporate a complete system to detect faces immediately after the images are captured and the expressions are recognized through facial features. This work presents a novel scheme to

* Corresponding author. Fax: +886 4 23924419.
E-mail addresses: pai@mail.ncut.edu.tw, pai@ncut.edu.tw (N.-S. Pai).

0898-1221/$ – see front matter
doi:10.1016/j.camwa.2010.08.082

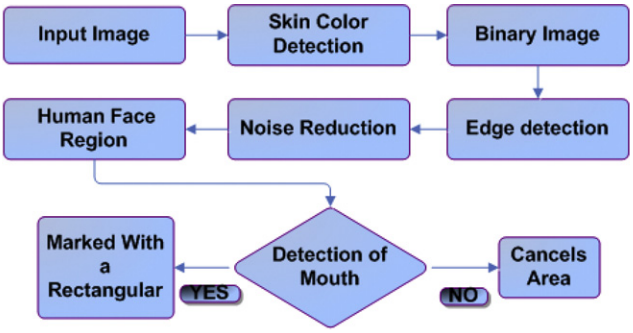

Fig. 1. The structure of the face detection method.

achieve real-time recognition of facial expressions using the extension theory [8,9], as well as classify all facial expressions using facial feature points. Additionally, a set of facial feature points is designated as a facial expression. Moreover, the input signals used for recognition can be obtained in real time by analyzing the shapes and corners of the mouth.

2. Face detection

A face tracking procedure generally includes processes such as image processing, skin color detection, lip detection and eye detection. The proposed tracking system is composed of these processes, as shown in Fig. 1.

2.1. Color space transformation and skin tone segmentation

Face detectors based on skin color segmentation is one of the most efficient means of segmenting a face. This work adopts the *YCbCr* color space for skin feature extraction to effectively distinguish the color space from brightness factor. Also, the RGB color space is converted into *YCbCr* color space, and defined as

$$\begin{bmatrix} Y \\ Cb \\ Cr \end{bmatrix} = \begin{bmatrix} 0.299 & 0.587 & 0.114 \\ -0.16874 & -0.33126 & 0.5 \\ 0.5 & -0.41869 & -0.08131 \end{bmatrix} \begin{bmatrix} R \\ G \\ B \end{bmatrix} + \begin{bmatrix} 0 \\ 128 \\ 128 \end{bmatrix}. \tag{1}$$

Here, the skin tone pixel is defined by selecting Y, Cb and Cr values within a certain region after converting *YCbCr* color space from RGB, The region values are defined as follows:

$$\begin{cases} 50 \le Y \le 230 \\ 85 \le Cb \le 130 \\ 125 \le Cr \le 165. \end{cases} \tag{2}$$

2.2. Edge detection and noise removal

By applying the sobel technique to produce the binary skin tone operation, edge detection is used to identify the edges of the areas in the image, revealing an apparent contrast with the background. Because the color space transformation and skin tone-segmenting processes result in an obvious contrast onto the face and background, the facial area can be detected precisely when performing edge detection. Additionally, facial features such as the mouth and eyes can be positioned accurately and outlined as the foundation for the following recognition.

Noise that appears in large areas, isolated points or protruding objects is eliminated using the erosion method, while small areas are effaced using the 5×5 mask median filter method. Moreover, the roughness and indents appearing on the edges of the images are reduced using the dilation method.

2.3. Facial zone and features extraction

Following elimination of the noise, connected component labeling method [10] is adopted to identify the conjoint pixels of an identical object in an image in order to obtain information from the object area. Some noise still remains after the erosion, median filter and dilation processes. Notably, a facial image can be obtained if the area proportion is taken advantage of, i.e. the length of a face must be larger than its width, and the information from the larger area is maintained. Following sieving of the facial zone, the range for the lips can be detected using the different colors between skin and lips. The color threshold values are defined in Eq. (3) using the simulation results. If the color of the lips satisfies the following condition, then the equation is the range for the lips.

$$1.65 \le (R/G) \le 2.05. \tag{3}$$

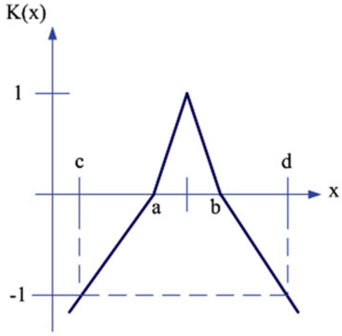

Fig. 2. The extension correlation function.

3. Extension recognition for facial expressions

Theoretical supports for extension theory of Cai are the matter element theory developed by Cai [9] and the extension set theory, in which the matter element is its logical cell. Extension nature refers to the variation possibility of a matter. The variation of matter is described based on the variation of a matter element. Elucidating the matter element theory involves investigating the extension nature, variation and variation character of matter elements. The name of a matter is defined as N and the value corresponding to the characteristic C is V. A matter element is defined as

$$R = (N, C, V). \tag{4}$$

This work presents a novel approach for recognizing the facial expressions by utilizing the extension theory. With this theory, the facial expressions are classified using facial features to represent all facial expressions, including surprise, happiness, sadness and other general ones. The input signals used for recognition can be obtained by analyzing the shapes and corners of the mouth. The models for the matter elements of facial expressions must be established first. Facial expression recognition is then proceeded with by applying extension correlation functions. The recognition algorithm procedures based on the extension theory are described as follows.

3.1. Establish matter element models for all facial expressions

In facial expression recognition, the multidimensional matter element theory can also be applied to describe all recognition forms:

$$R_i = \begin{bmatrix} N_i, & \textit{shape of lips}, & \langle a_{i1}, b_{i1}\rangle \\ & \textit{corner of mouth}, & \langle a_{i2}, b_{i2}\rangle \end{bmatrix} \quad i = 1, \dots, 4 \tag{5}$$

where R_1–R_4 refer to the facial expressions of general ones, surprise, happiness and sadness, respectively; N_i represents the various lip patterns; and $\langle a_{ij}, b_{ij}\rangle$ denote the feature values.

3.2. Evaluate the extension correlation functions

Fig. 2 illustrates the degree of relationship derived from determining whether the dot x is part of X_O using the extension correlation function. In addition to considering the location correlation of a dot and a region, that of two regions and of a dot and two regions usually need to be also considered as well. The relation between a dot and two regions is described by using a location value. By considering $X_O = \langle a, b\rangle$, $X = \langle c, d\rangle$ and $X_O \subset X$, the location value of X_O and X can be defined as

$$X_O = \begin{Bmatrix} a_{i1}, b_{i1} \\ a_{i2}, b_{i2} \end{Bmatrix}, \qquad X = \begin{Bmatrix} c_{i1}, d_{i1} \\ c_{i2}, d_{i2} \end{Bmatrix} \tag{6}$$

where $\langle a_{i1}, b_{i1}\rangle$ and $\langle c_{i1}, d_{i1}\rangle$ denote the classical domain and neighborhood domain of the feature of the shape of lips, respectively; $\langle a_{i2}, b_{i2}\rangle$ and $\langle c_{i2}, d_{i2}\rangle$ represent the classical domain and neighborhood domain of the feature of the corner of the mouth. The neighborhood domain $\langle c_{i1}, d_{i1}\rangle$ is defined as the lip width divided by lip height and is assigned as $\langle 0.1, 3.6\rangle$ using the experimental results. Here, the height of the lips is partitioned into 10 segments and the neighborhood domain $\langle c_{i2}, d_{i2}\rangle$ is defined as the average value of the left most corner and right most corner of the mouth. Therefore, the range of $\langle c_{i2}, d_{i2}\rangle$ is set as $\langle 0, 10\rangle$. The models of the matter elements for the facial expressions are defined here as the classical domain and are listed in Table 1. The correlation function can be defined as

$$K(x) = \frac{\rho(x, X_O)}{D(x, X_O, X)} \tag{7}$$

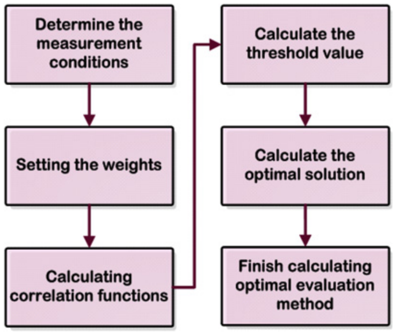

Fig. 3. The procedures for the optimal evaluation method.

where $D(x, X_O, X)$ is defined as

$$D(x, X_O, X) = \begin{cases} \rho(x, X) - \rho(x, X_O) & x \notin X_O \\ -1 & x \in X_O. \end{cases} \tag{8}$$

Rearranging Eq. (7) through the definitions of the location value Eq. (8), yields

$$K(x) = \begin{cases} -\rho(x, X_O), & x \in X_O \\ \dfrac{\rho(x, X_O)}{\rho(x, X) - \rho(x, X_O)}, & x \notin X_O \end{cases} \tag{9}$$

where

$$\rho(x, X) = \left| x - \frac{c+d}{2} \right| - \frac{1}{2}(d-c) \tag{10}$$

and

$$\rho(x, X_O) = \left| x - \frac{a+b}{2} \right| - \frac{1}{2}(b-a). \tag{11}$$

3.3. Optimal evaluation method and human facial expression recognition

The optimal evaluation method can appraise an object, which contains factors such as the matter, strategy and method. This work described an optimal evaluation method in which all related procedures are described in Fig. 3. The measurement condition is determined by defining each facial expression with an element pattern correlation function set $K_{ij}(x_j)$, $i = 1, 2, 3, 4$, $j = 1, 2$ based on Eq. (9).

The weight functions are determined from viewpoint of the importance of all relational functions. Owing to only two features addressed in this work, the weighting factors are all set to 1/2. Various facial expression correlation functions are defined as

$$\bar{K}_i = \sum_{j=1}^{2} W_{ij} K_{ij}, \quad i = 1, 2, 3, 4. \tag{12}$$

The readability and flexibility of facial expression recognition is increased by normalizing the modified correlation function as

$$\tilde{K}_i = \frac{2\overline{K_i} - \overline{K}_{i(\max)} - \overline{K}_{i(\min)}}{\overline{K}_{i(\max)} - \overline{K}_{i(\min)}}, \quad i = 1, 2, 3, 4. \tag{13}$$

The value of $\tilde{K}_i$ is located between 1 and −1. The maximum of $\tilde{K}_i$ is determined as

$$\tilde{K}_{\max} = \max_{i=1,2,3,4} \tilde{K}_i. \tag{14}$$

A larger correlation function value generally implies a higher probability of determining the human facial expression. Moreover, satisfactory recognition results are obtained using the above procedures for the proposed algorithm.

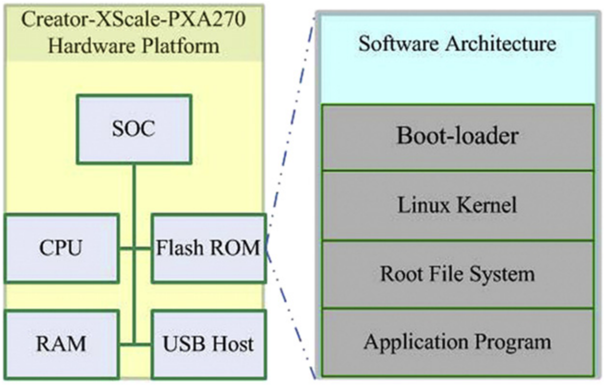

Fig. 4. The hardware and software structures of the embedded system.

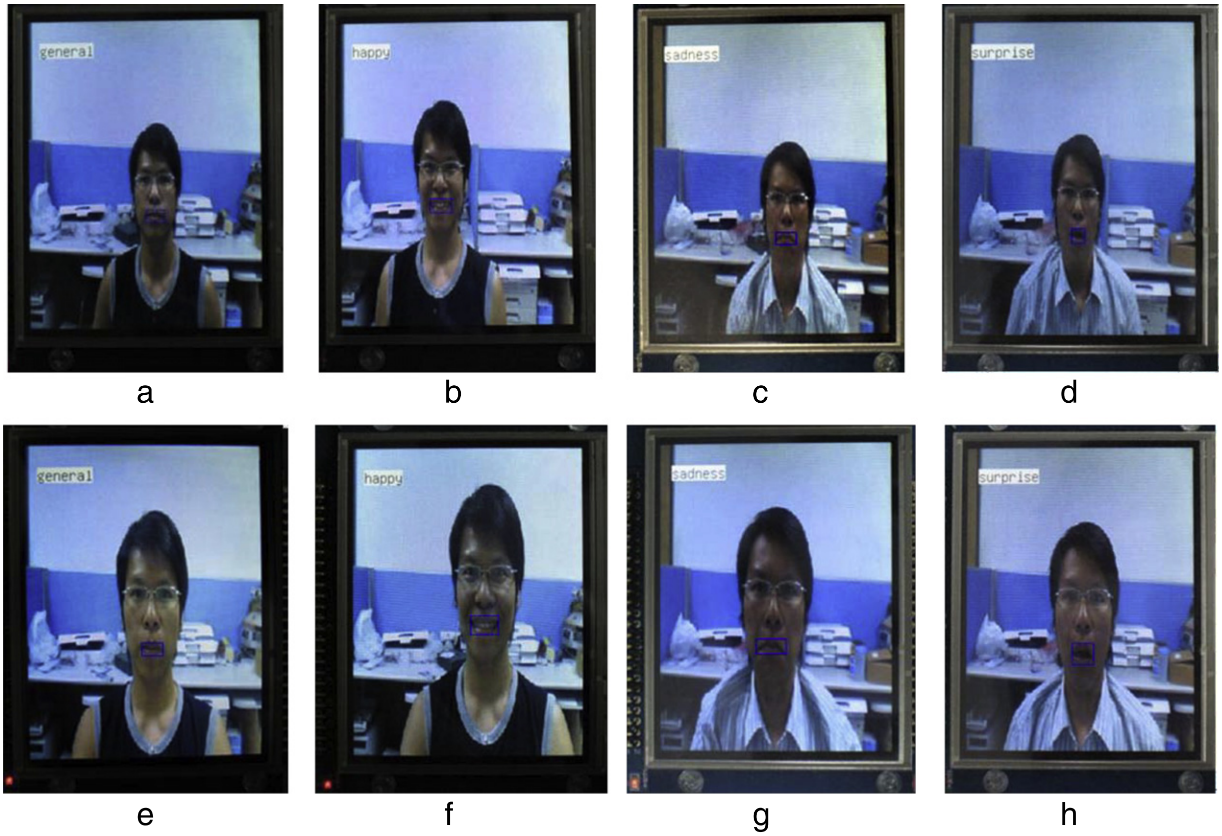

Fig. 5. Pictures of every state of facial expressions.

4. Real-time implementation of recognition on an embedded platform

The design platform adopted in this work was developed by Microtime Computer Inc. The system structure is shown in Fig. 4. The left block is used as the foundation for the hardware level, and the right block is the software level. Embedded Linux is less expensive, portable and powerful, explaining why it is adopted here as the operating system. The entire software architecture was programmed in a Flash ROM.

5. Experimental results

Fig. 5 displays every state of facial expressions detected from a short ((e)–(h)) and a long ((a)–(d)) distance. The system is highly competent in facial recognition performance, regardless of the distance. The upper left-hand corner of each picture

Table 1
Models of the matter elements for facial expressions (C_1: shape of lips, C_2: corner of mouth).

Expressions	The models of the matter elements
General	$R_1 = \begin{bmatrix} N_1 & C_{11} & \langle 2.45, 2.91\rangle \\ & C_{12} & \langle 4.1, 6.15\rangle \end{bmatrix}$
Surprise	$R_2 = \begin{bmatrix} N_2 & C_{21} & \langle 0.5, 2.1\rangle \\ & C_{22} & \langle 4.71, 6.61\rangle \end{bmatrix}$
Happy	$R_3 = \begin{bmatrix} N_3 & C_{31} & \langle 2.59, 3.56\rangle \\ & C_{32} & \langle 0.35, 4.21\rangle \end{bmatrix}$
Sadness	$R_4 = \begin{bmatrix} N_4 & C_{41} & \langle 2.04, 2.62\rangle \\ & C_{42} & \langle 6.12, 9.57\rangle \end{bmatrix}$

Table 2
The experimental results for the proposed facial expression recognition method.

		Recognition result of facial expression				Recognition rate
		General	Surprise	Happy	Sadness	
Actual expression	General	87	3	5	8	0.87
	Surprise	2	89	9	0	0.89
	Happy	6	2	91	1	0.91
	Sadness	10	6	0	84	0.84
Mean						0.878

displays the recognition results for the corresponding facial expression. One hundred test pictures are captured to recognize, with those results summarized in Table 2.

6. Conclusion

This work has presented a novel application based on extension theory to human facial expression recognition. A simple lip feature extraction scheme is also developed by using the shape of the lips and the corners of the mouth as recognition features for various facial expressions. To cope with real time face detection and recognition needs, the proposed algorithm is implemented on an embedded system to speed up facial expression recognition. Experimental results indicate that the accuracy for the expression 'sadness' is lower than that of the other 3 facial expressions. Moreover, grief is likely more difficult to express on human faces. Furthermore, the accuracy of the other facial expressions, e.g., general, surprise and happiness achieves a recognition rate of 87%.

References

[1] Christopher Hallinan, Embedded linux primer: a practical real-world approach, 2006.
[2] T. Kanade, J. Cohn, Y.-L. Tian, Comprehensive database for facial expression analysis, in: Proceedings of the 4th IEEE International Conference on Automatic Face and Gesture Recognition, 2000, pp. 6–53.
[3] M.H. Yang, D.J. Kriegman, N. Ahuja, Detecting faces in images: a survey, IEEE Transactions on Pattern Analysis and Machine Intelligence 24 (1) (2002) 34–58.
[4] Y.C. Feng, P.C. Yuen, A.K. Jain, A hybrid approach for generating secure and discriminating face template, IEEE Transactions on Information Forensics and Security 5 (2010) 103–117.
[5] Z. Liu, C. Liu, A hybrid color and frequency features method for face recognition, IEEE Transactions on Image Processing 17 (10) (2008) 1975–1980.
[6] P. Lim, L.C. De Silva, L. Jing, L.L. Ling, Multimodal approach to human-face detection and tracking, IEEE Transactions on Industrial Electronics 55 (3) (2008) 1385–1393.
[7] P. Ekman, W.V. Friesen, The Facial Action Coding System: A Technique for the Measurement of Facial Movement, Consulting Psychologists Press, San Francisco, 1978.
[8] M.H. Wang, Extension neural network-type 2 and its applications, IEEE Transactions on Neural Networks 16 (2005) 1352–1361.
[9] W. Cai, The extension set and incompatibility problem, Journal of Scientific Exploration 1 (1983) 81–93.
[10] M.B. Dillencourt, H. Samet, M. Tamminen, A general approach to connected-component labeling for arbitrary image representations, Journal of the Association for Computing Machinery 39 (1992) 253–280.