International Journal of Innovative Computing, Information and Control
Volume **11**, Number **2**, April **2015**

ICIC International ©2015 ISSN 1349-4198
pp. 707–717

FACE DETECTION USING TSK-TYPE FUZZY CEREBELLAR MODEL ARTICULATION CONTROLLER NETWORK

Jyun-Guo Wang[1], Shen-Chuan Tai[1] and Cheng-Jian Lin[2]

[1]Institute of the Computer and Communication Engineering
National Cheng Kung University
No. 1, University Road, Tainan 701, Taiwan
jyunguo.wang@gmail.com; sctai@mail.ncku.edu.tw

[2]Department of Computer Science and Information Engineering
National Chin-Yi University of Technology
No. 57, Sec. 2, Zhongshan Road, Taiping Dist., Taichung 41170, Taiwan
cjlin@ncut.edu.tw

Received May 2014; revised September 2014

Abstract. *Face detection is a topic concerning businesses that use security surveillance systems, and accurate detection is critical for the subsequent analysis of images used in facial recognition. In this study, we propose a face-detection method that involves using lighting compensation, skin-color analysis, and a TSK-type fuzzy cerebellar model articulation controller (TSK-type FCMAC) network. The proposed TSK-type FCMAC network has some advantages, such as include its fast rapid learning property, good generalization generalizability capability, easy to ease of implementation, and small lower required computing computer-memory requirements. Our experimental results confirm that the proposed method outperforms other models and effectively overcomes the problems of face variation and complex backgrounds.*
Keywords: Skin tone, Color transformation, Lighting compensation, TSK-type cerebellar model articulation controller, Fuzzy system

1. **Introduction.** Image processing is a critical research field, and it is widely used in various applications, such as visual surveillance, pattern classification, and image retrieval. Because employing surveillance systems is a non-contact detection method, face detection is suitable for recognizing and tracking objects. Moreover, users do not feel that their privacy is invaded. Consequently, face detection systems are among the most successful and crucial image-processing applications [1,2]. Many face detection techniques, such as the knowledge-based method, invariant-feature method, template-matching method, and appearance-based method, have been proposed in the literature [3] to solve face detection problems. The knowledge-based method involves a rule-based approach and an encoded analysis of the features that constitute a typical face. Generally, some specific rules have been designed to capture the relational characteristic of facial components. The invariant-feature method has been used to identify structural features, such as facial components, textures, and skin color. The aforementioned methods were designed to identify typical structural features among various types of faces under distinct ambient conditions. The template-matching method involves describing the correlations between a current input image and standard patterns that are considered typical human facial components. The appearance-based method was proposed to capture diverse variations in facial appearance from a set of training images. Because of extreme variations in facial expressions and complex backgrounds, these face-detection techniques are prone to extracting incorrect information from numerous facial expressions. Therefore, to address these problems

and enhance recognition performance, many face detection methods [4-6] based on a specific face model have been proposed to account for varying conditions, including face rotation, extreme variation in facial expressions, and complex backgrounds. Ai *et al.* [4] employed support vector machines and template-matching approaches to analyze face images. Karungaru *et al.* [5] applied a genetic algorithm-based template-matching method to perform image classification. Wang and Yuan [6] proposed a novel human face detection method by using an evolutionary computation technique to cluster skin-like color pixels and to segment each face-like region from color images under complex conditions. However, variations in the size of the regions involved in the aforementioned methods lead to judgment errors in a face image. To overcome this problem, some skin color models have been developed using color spaces such as RGB color space [7,8], HSV or HSI color space [9,10], YCbCr color space [11-13]. High-performance algorithms are required to overcome some of the drawbacks of these approaches. Therefore, the color conversion and characteristics analysis are very important for preprocessing on face detection. In addition, a robust classifier is a key that influences overall recognition performance.

In recent decades, artificial neural networks have exhibited adequate performance for applications in various fields. The most widely used neural network models are the multilayer perception (MLP) model and cerebellar model articulation controller (CMAC). Recently, many types of architecture model, such as the hierarchical CMAC (HCMAC) and fuzzy CMAC (FCMAC), have been proposed. However, these methods still have some drawbacks; for example, the large rule base required causes memory overload, the advantage of mapping by performing local approximation of the rule hyper-planes is not fully exploited, and the globally optimal solution cannot be determined. Therefore, to overcome these drawbacks, we propose a novel TSK-type FCMAC network that solves face-detection problems. In addition, we compare the proposed method with other classifiers, such as a standard FCMAC [14-17]. The proposed TSK-type FCMAC is a local network, where, for a given input vector, only a few of the network's nodes (or hypercube cells) will be activated and will effectively contribute to the corresponding network output. The proposed TSK-type FCMAC network has advantages, such as a rapid learning property, good generalization capability, ease of implementation, and low computer-memory requirements. Finally, we employed the proposed TSK-type FCMAC model to solve face detection problems in complex environments.

The remainder of this paper is organized as follows. Section 2 describes the feature extraction method for the face detection system. Sections 3 and 4 present the proposed model and corresponding online learning algorithm, respectively. Section 5 details the performance of the proposed approach when it was subjected to a benchmark test for comparison with other structures involved in skin detection applications. Finally, we offer our conclusion in Section 6.

2. **Analysis of the Feature Extraction.**

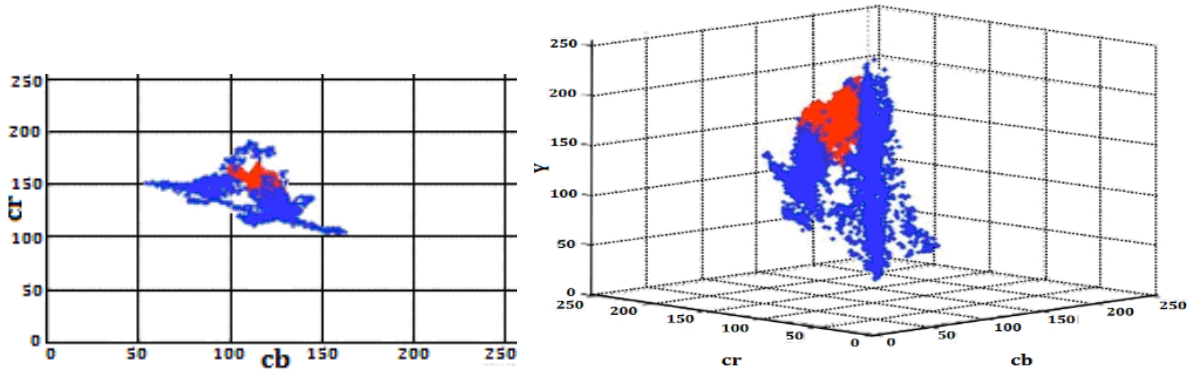
Feature extraction has a critical role in and is directly related to the performance of face-recognition systems. Therefore, many researchers have proposed various types of method for improving recognition results through analysis of textures, shapes, colors, or other key features. In this paper, we propose a method that involves using skin-tone characteristics to perform face detection. Because skin tone can provide numerous characteristics for segmenting local characteristics in color images, color image processing is discussed and incorporated into the following three steps: (1) lighting compensation, (2) color segmentation, and (3) skin tone detection. These steps are described briefly as follows.

(1) Lighting compensation

Because differences in lighting variation strongly influence the quality of subsequent image processing, the original color intensity must be adjusted by introducing a "reference white" to normalize the color appearance.

(2) Color segmentation

Many color spaces are used in analyzing images (e.g., RGB, HSV, and YCbCr). The RGB color space is not the most used color space in skin detection methods because it is easily influenced by lighting conditions and, thus, is unsuitable for color segmentation in well-lit environments. The HSV color space is more sensitive to diversification of lighting, although it generates some unwanted errors. The YCbCr space provides good separation between luminance and chrominance information and faster transformation speed. According to research result demonstrates that skin tone and background features have excellent segregation characteristics in the YCbCr color space, indicating that skin tone possessed uniqueness and a characteristic of aggregation when it is distributed in Cb and Cr subspaces. Based on the results shown in Figure 1, we used the YCbCr color space in this study.



(a) (b)

FIGURE 1. Skin and non-skin distribution: (a) 2-D projection in the CbCr subspace, and (b) the YCbCr space

(3) Skin tone detection

We randomly selected numerous skin and non-skin pixels in color images obtained from the California Institute of Technology (CIT) face database.

Processing is performed in five stages. First, lighting compensation is conducted to collect color information from the images and, thus, remove improper lighting conditions caused by chrominance. Second, the default RGB color space is transformed into the YCbCr color space. Third, the skin-tone cluster characteristics are determined in the YCbCr color space. Fourth, the Y, Cb, and Cr information of the color images are adopted as features into the TSK-type FCMAC model for training. Fifth, through computation of the proposed TSK-type FCMAC model, the output is identified as skin or non-skin information to detect facial regions.

3. **Structure of the Proposed TSK-Type FCMAC Network.** As shown in Figure 2, the network structure of the proposed TSK-type fuzzy CMAC is composed of the following layers: (1) the input space partition; (2) association memory selection; and (3) defuzzification. Similar to the conventional CMAC model, the proposed TSK-type FCMAC network approximates a nonlinear function $y = f(x)$ by using two primary mappings, i.e., $S(x)$ and $P(\alpha)$. These two mappings were realized using fuzzy operations.

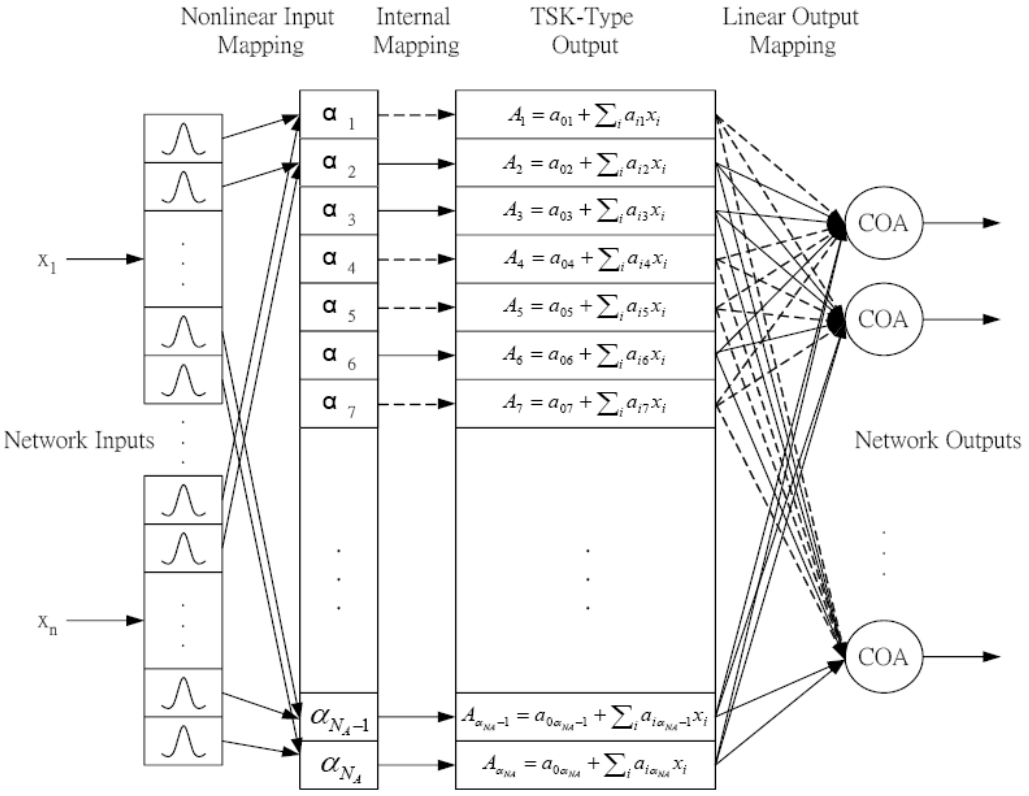


FIGURE 2. Architecture of the proposed TSK-type FCMAC network

The function maps each point x in the input space onto an associated vector that has N_L nonzero elements ($N_L < N_A$). Here, $0 \leq \alpha \leq 1$ for all components in α ($\alpha = \alpha_1, \alpha_2, \ldots, \alpha_n$), was derived from the composition of the receptive field functions and sensory inputs. In the hypercube, we compute by the input state x that hypercube value is calculated by product operation through the strength of the receptive field functions for each input state. The mathematical functions of each node and layer l are detailed as follows. The net input of the ith node in layer l is represented as $u_i^{(l)}$ and the output value is represented as $O_i^{(l)}$.

Layer 1 (input layer): The inputs, which are crisp values and $\bar{x} = (x_1, x_2, \ldots, x_n)$ are inputted into this layer, while the corresponding outputs are computed as

$$O_i^{(1)} = u_i^{(1)}, \text{ and } u_i^{(1)} = x_i. \tag{1}$$

To normalize the input range, each node in this layer scales input x_i, $i = 1, \ldots, n$, within the range of $[-1, 1]$. No weight requires adjustment in this layer.

Layer 2 (fuzzification layer): This layer is used to perform the fuzzification operation. Each node in this layer is a Gaussian membership function. For the ith fuzzy set A_j^i on the input variable x_i, $i = 1, \ldots, n$, a Gaussian membership function is defined by

$$O_{ij}^{(2)} = \exp\left(-\frac{\left[u_i^{(1)} - m_{ij}\right]^2}{\sigma_{ij}^2}\right), \text{ and } u_i^{(2)} = O_i^{(1)} \tag{2}$$

where m_{ij} and σ_{ij} are respectively the mean and variance of the Gaussian membership function of the jth term of the ith input variable x_i.

Layer 3 (spatial firing layer): Each node represents a hypercube cell that computes the firing strength of the rules for this layer. For the obtained spatial firing strength α_j,

each node performs a fuzzy meet operation on the inputs, as it receives from the previous layer via an algebraic product operation. Thus, the output function of each inference node can be calculated as

$$\alpha_j = O_j^{(3)} = \prod_i u_{ij}^{(3)}, \text{ and } u_{ij}^{(3)} = O_{ij}^{(2)} \tag{3}$$

where the $\prod_i u_{ij}^{(3)}$ of a rule node represents the firing strength of its corresponding rule.

Layer 4 (consequent layer): Each node is an optional node formed by a function of a linear combination of input variables in this layer. Each linear combination is inferred to produce a partial fuzzy output by applying the value of its corresponding association memory selection vector as matching degree. The formula can be expressed as

$$O_j^{(4)} = \alpha_j \left(a_{0j} + \sum_{i=1}^{N_D} a_{ij} x_i \right) \tag{4}$$

where both a_{0j} and a_{ij} denote the scalar values, N_D is the number of input dimensions, and x_i represents the ith input dimension.

Layer 5 (output layer): The partial fuzzy output is defuzzified into a scalar output by using the centroid of area (COA) approach. Subsequently, the actual output y is derived as follows

$$y = \frac{\sum_{j=1}^{NL} \alpha_j \left(a_{0j} + \sum_{i=1}^{N_D} a_{ij} x_i \right)}{\sum_{j=1}^{NL} \alpha_j} \tag{5}$$

where N_L denotes the number of hypercube cells.

4. **Learning Algorithm for the Proposed TSK-Type FCMAC Network.** The proposed TSK-type FCMAC model simultaneously evolves to form two types of learning (i.e., structure and parameter learning), as shown in Figure 3.

A. Structure learning scheme

The first task in structured learning involves determining whether to generate a new hypercube cell and identifying the number of hypercube cells extracted from the training data. The product operation of the firing strength obtained directly from $\prod_i u_{ij}^{(3)}$ can be used as the degree measure as follows:

$$\alpha_j = \prod_i u_{ij}^{(3)} \tag{6}$$

where u_{ij} represents the degree of the firing strength of the input vector for $i = 1, \ldots, N_D$ and the associated vector $\alpha_j \in [0, 1]$. The criterion for the degree measure for generating a new hypercube cell of new incoming data $\vec{x} = (x_1, x_2, \ldots, x_n)$ is expressed in Equation (7), which is employed to find the maximum degree $\alpha_{\max}$

$$\alpha_{\max} = \max_{1 \le j \le N_L} \alpha_j(\vec{x}) \tag{7}$$

where N_L is the number of nonzero elements of the associative vector. In addition, we set the threshold value $\bar{\alpha} \in [0, 1]$ as a prespecified threshold that should decay during the learning process to limit the size of the proposed TSK-FCMAC network. When $\alpha_{\max} \le \bar{\alpha}$, a new hypercube cell is generated.

B. Parameter learning scheme

Four parameters must be tuned (i.e., m_{ij}, σ_{ij}, a_{0j}, and a_{ij}) by using the supervised gradient descent method. Our goal is to minimize the cost function E, i.e.,

$$E(t) = \frac{1}{2}[y^d(t) - y(t)]^2 \tag{8}$$

where $y^d(t)$ is the desired output and $y(t)$ is the actual output at time t. Subsequently, the parameter learning algorithm, which is based on backpropagation, can be described as follows:

$$a_{0j}(t+1) = a_{0j}(t) + \Delta a_{0j} = a_{0j}(t) + \eta \cdot e \cdot \left(\frac{\partial y}{\partial a_{0j}}\right) \tag{9}$$

and

$$a_{ij}(t+1) = a_{ij}(t) + \Delta a_{ij} = a_{ij}(t) + \eta \cdot e \cdot \left(\frac{\partial y}{\partial a_{ij}}\right) \tag{10}$$

where a_{0j} and a_{ij} are the proper scalar and coefficient of the ith input dimension, respectively; and j is the jth element of the TSK-type output vector for $j = 1, 2, \ldots, N_L$ (where N_L denotes the number of hypercube cells); and η is the learning rate (between 0 and 1). The elements of the TSK-type output vectors are updated according to the amount

$$\Delta a_{0j} = \eta \cdot e \cdot \frac{\partial y}{\partial a_{0j}} = \eta \cdot e \cdot \frac{\alpha_j}{\sum_{j=1}^{N_L} \alpha_j} \tag{11}$$

and

$$\Delta a_{ij} = \eta \cdot e \cdot \frac{\partial y}{\partial a_{ij}} = \eta \cdot e \cdot \frac{x_i \alpha_j}{\sum_{j=1}^{N_L} \alpha_j} \tag{12}$$

where η is the learning rate, between 0 and 1, and e is the error between the desired and actual output, $e = y^d - y$. The receptive field functions are updated as follows

$$m_{ij}(t+1) = m_{ij}(t) + \Delta m_{ij} = m_{ij}(t) + \eta \cdot e \cdot \left(\frac{\partial y}{\partial \alpha_j} \cdot \frac{\partial \alpha_j}{\partial m_{ij}}\right) \tag{13}$$

and

$$\sigma_{ij}(t+1) = \sigma_{ij}(t) + \Delta \sigma_{ij} = \sigma_{ij}(t) + \eta \cdot e \cdot \left(\frac{\partial y}{\partial \alpha_j} \cdot \frac{\partial \alpha_j}{\partial \sigma_{ij}}\right) \tag{14}$$

where i denotes the ith input dimension for $i = 1, 2, \ldots, n$; m_{ij} denotes the mean of the receptive field functions, and σ_{ij} denotes the variance of the receptive field functions. The parameters of the receptive field functions are updated according to the amount

$$\begin{aligned}\Delta m_{ij} &= \eta \cdot e \cdot \frac{\partial y}{\partial \alpha_j} \cdot \frac{\partial \alpha_j}{\partial m_{ij}} \\ &= \eta \cdot e \cdot \frac{\left(a_{0j} + \sum_{i=1}^{N_D} \alpha_{ij} x_i\right) \sum_{j=1}^{N_L} \alpha_j - \sum_{j=1}^{N_L} \alpha_j \left(a_{0j} + \sum_{i=1}^{N_D} \alpha_{ij} x_i\right)}{\left(\sum_{j=1}^{N_L} \alpha_j\right)^2} \cdot \alpha_j \cdot \frac{2(x_i - m_{ij})}{\sigma_{ij}^2}\end{aligned} \tag{15}$$

and

$$\begin{aligned}\Delta \sigma_{ij} &= \eta \cdot e \cdot \frac{\partial y}{\partial \alpha_j} \cdot \frac{\partial \alpha_j}{\partial \sigma_{ij}} \\ &= \eta \cdot e \cdot \frac{\left(a_{0j} + \sum_{i=1}^{N_D} \alpha_{ij} x_i\right) \sum_{j=1}^{N_L} \alpha_j - \sum_{j=1}^{N_L} \alpha_j \left(a_{0j} + \sum_{i=1}^{N_D} \alpha_{ij} x_i\right)}{\left(\sum_{j=1}^{N_L} \alpha_j\right)^2} \cdot \alpha_j \cdot \frac{2(x_i - m_{ij})^2}{\sigma_{ij}^3}\end{aligned} \tag{16}$$

where η is the learning rate of the mean and variance of the receptive field functions.

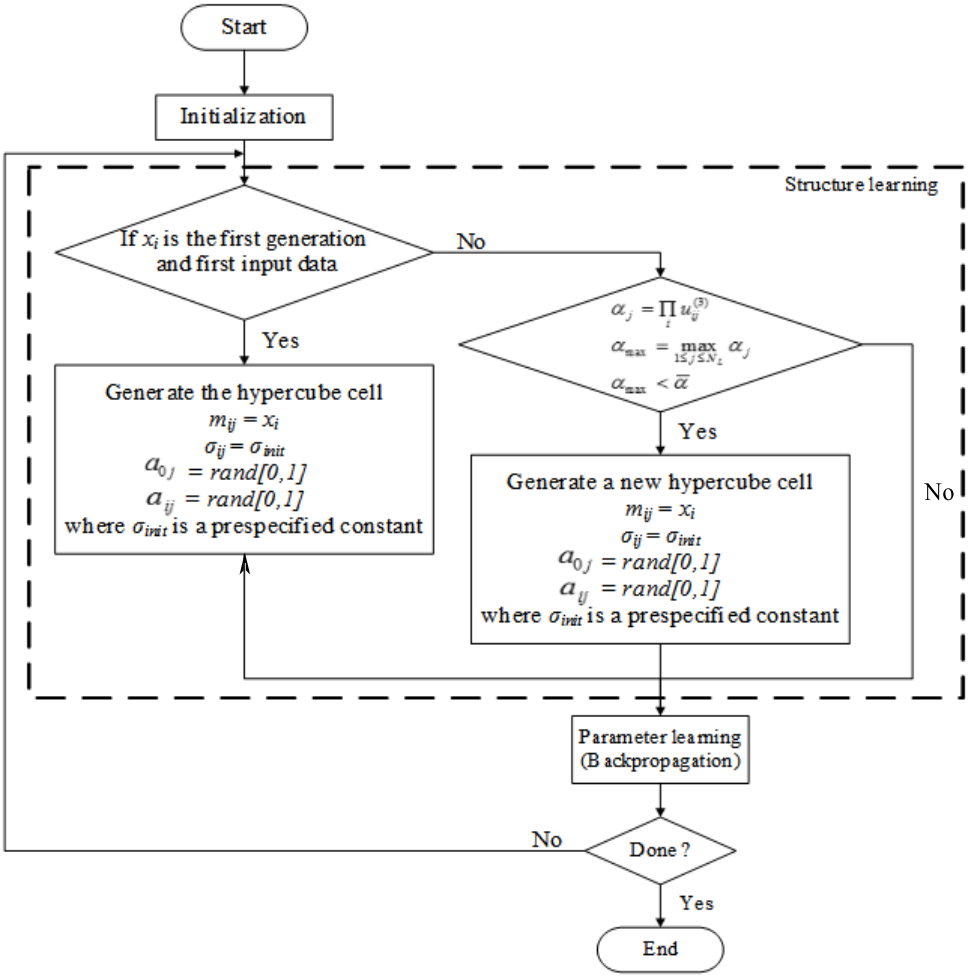
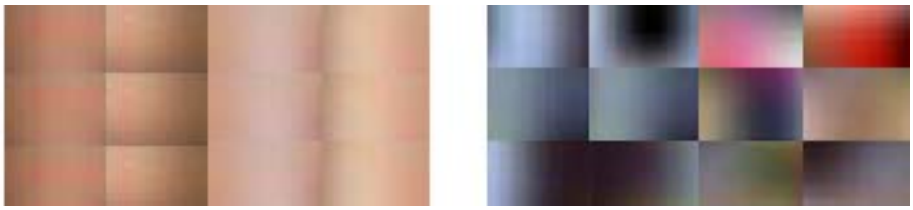


FIGURE 3. Flowchart of the structure and parameter learning for the proposed TSK-type FCMAC network

5. **Experimental Results.** In this paper, we analyzed various types of environment obtained from the California Institute of Technology (CIT) database to confirm that the proposed method outperforms other methods. The images analyzed in this study included 450 color images of 27 people in various environments with distinct lighting, backgrounds, and facial expressions. Figure 4 shows some examples of the various environments in the original images obtained from the CIT database. A two area method was employed to classify each feature as skin or non-skin for the input feature vector of the skin detection method from Y, Cb, and Cr; a pixel-based classification method was used to classify each skin and non-skin image independently from its neighbors. Figure 5 shows the training data and test data from the faces and non-faces used to measure the metric performance. For the input features of three dimension (i.e., Y, Cb, and Cr), we randomly selected 6000 training data and 6000 test data from the skin and non-skin color images. The initial threshold value and the learning rate were set at $\bar{\alpha} = 1 * 10^{-9}$ and $\eta = 0.0005$, respectively. As shown in Figure 6, in the conducted experiments, we calculated the mean from the results of 10 runs to plot the learning curve, whereas Table 1 presents the mean results of 10 runs, with the mean, best and worst root-mean-square error (RMSE) values. The individual results of the 10 independent runs, including the number of hypercube cells and accuracy rates, are shown in Table 2, where the hypercube cell represents the minimal number of rules of operation required for application. As shown in Table 3, the proposed

FIGURE 4. Examples from the original CIT face database



(a) (b)

FIGURE 5. Feature sets: (a) skin sets, (b) non-skin sets

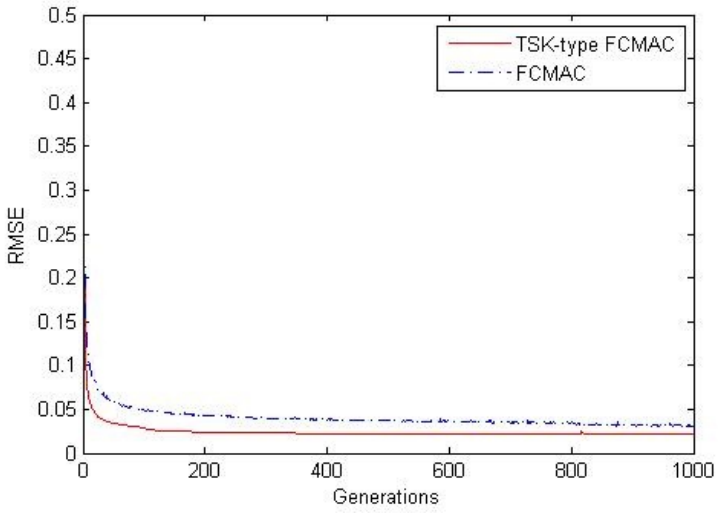


FIGURE 6. 10-run averaged learning curves

approach outperformed the traditional FCMAC method based on the mean results of 10 runs test. Finally, Figures 7 and 8 show the simulated image location results of precise skin detection; thereby observe the pros and cons comparison between the conventional FCMAC model and the model proposed in this study.

TABLE 1. RMSE comparison in training

Models/Items	*Mean RMSE*	*Best RMSE*	*Worst RMSE*
FCMAC	*0.0368*	*0.0205*	*0.0593*
TSK-type FCMAC	*0.0330*	*0.0164*	*0.0390*

TABLE 2. Numbers of hypercube cells and accuracy rates for 10 independent runs

Models/Experiments	*1*	*2*	*3*	*4*	*5*
Hypercube cell (FCMAC)	*8*	*9*	*10*	*10*	*9*
Hypercube cell (TSK-type FCMAC)	*6*	*6*	*7*	*7*	*6*
Accuracy (%) (FCMAC)	*86.9667*	*87.0000*	*87.4500*	*87.8000*	*87.3667*
Accuracy (%) (TSK-type FCMAC)	*91.5667*	*91.4667*	*91.0167*	*92.7380*	*91.2800*
Models/Experiments	*6*	*7*	*8*	*9*	*10*
Hypercube cell (FCMAC)	*9*	*10*	*9*	*9*	*10*
Hypercube cell (TSK-type FCMAC)	*6*	*6*	*6*	*7*	*7*
Accuracy (%) (FCMAC)	*86.1667*	*88.5500*	*88.0333*	*90.1833*	*90.1333*
Accuracy (%) (TSK-type FCMAC)	*90.0900*	*90.5000*	*90.7000*	*90.0800*	*91.1333*

TABLE 3. Comparison result of the correct rate in testing

Models/Items	*Mean*	*Best*	*Worst*
FCMAC	*87.9650%*	*90.1833%*	*87.0000%*
TSK-type FCMAC	*91.0570%*	*92.7380%*	*90.0900%*

FIGURE 7. Results of the FCMAC model

6. **Conclusions.** The experimental results show that the proposed model can overcome the limits of the facial region size and solve complex background image problems. Using lighting compensation to reduce the extrinsic influence of lighting and selecting a strong skin-tone color model (e.g., Y, Cb, Cr) of the YCbCr color space as the network's input feature and via the TSK-type FCMAC model based on backpropagation algorithm

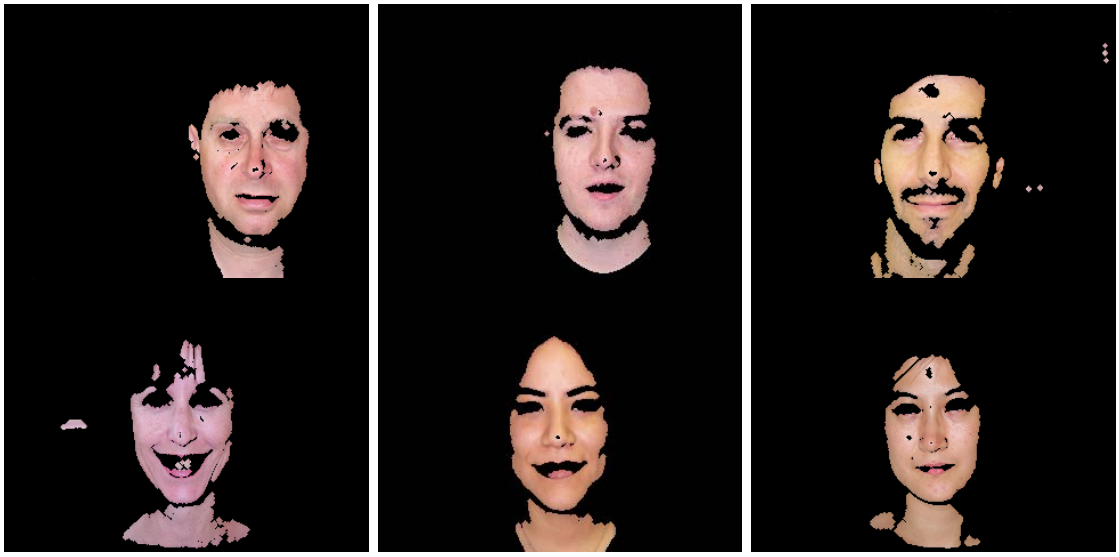


FIGURE 8. Results of our proposed TSK-type FCMAC model

to adjust the parameters appropriately, thereby enhancing the ability to identify a globally optimal solution. Thus, the proposed model is more efficient than the conventional FCMAC model and requires fewer hypercube cells.

REFERENCES

[1] W. Zhao, R. Chellappa, P. J. Phillips and A. Rosenfeld, Face recognition: A literature survey, *ACM Computing Surveys*, vol.35, no.4, pp.399-458, 2003.
[2] X. Lu, Image analysis for face recognition, *Personal Notes*, 2003.
[3] M. H. Yang, D. J. Kriegman and N. Ahuja, Detecting faces in images: A survey, *IEEE Trans. Pattern Anal. Mach. Intell.*, vol.24, no.1, pp.34-58, 2002.
[4] H. Ai, L. Liang and G. Xu, Face detection based on template matching and support vector machines, *Intl. Conf. Image Process*, vol.1, pp.1006-1009, 2001.
[5] S. Karungaru, M. Fukumi and N. Akamatsu, Face recognition using genetic algorithm based template matching, *IEEE International Symposium on Communications and Information Technology*, vol.2, pp.1252-1257, 2004.
[6] Y. J. Wang and B. Z. Yuan, A novel approach for human face detection from color images under complex background, *Pattern Recognition*, vol.34, pp.1983-1992, 2001.
[7] M. J. Seow, D. Valaparla and V. K. Asari, Neural network based skin color model for face detection, *The 32nd Applied Imagery Pattern Recognition Workshop*, pp.141-145, 2003.
[8] D. Chai, S. L. Phung and A. Bouzerdoum, A Bayesian skin/non-skin color classifier using non-parametric density estimation, *International Symposium on Circuits and Systems*, vol.2, pp.464-467, 2003.
[9] Q. Qiao, J. Zhou and K. He, Face detection in color images based on sub-image fusion and SVM, *IEEE International Conference on Mechatronics and Automation*, vol.2, pp.626-630, 2005.
[10] H. Lee, S. Hong, K. Oh, E. Kim and M. Park, TSK fuzzy modeling approach for face detection, *International Joint Conference SICE-ICASE*, pp.3941-3944, 2006.
[11] S. Kim, T. Jung, Y. Park and J. Kim, Adaptive skin detection system on DTV, *Proc. of 2004 International Symposium on Intelligent Signal Processing and Communication Systems*, pp.586-591, 2004.
[12] J. Yang, Z. Fu, T. Tan and W. Hu, Adaptive skin detection using multiple cues, *International Conference on Image Processing*, vol.2, pp.901-904, 2004.
[13] Z.-M. Lu, X.-N. Xu and J.-S. Pan, Face detection based on vector quantization in color images, *International Journal of Innovative Computing, Information and Control*, vol.2, no.3, pp.667-672, 2006.

[14] D. Y. Wang, C. J. Lin and C. Y. Lee, A new pseudo-Gaussian-based recurrent fuzzy CMAC model for dynamic systems processing, *International Journal of Systems Science*, vol.39, no.3, pp.289-304, 2008.

[15] S. F. Su, Z. J. Lee and Y. P. Wang, Robust and fast learning for fuzzy cerebellar model articulation controllers, *IEEE Trans. Systems, Man, and Cybernetics, Part B: Cybernetics*, vol.36, no.1, pp.203-208, 2006.

[16] T. F. Wu, P. S. Tsai, F. R. Chang and L. S. Wang, Adaptive fuzzy CMAC control for a class of nonlinear systems with smooth compensation, *IEE Proc. of Control Theory and Applications*, vol.153, no.6, pp.647-657, 2006.

[17] K. H. Cheng, Adaptive fuzzy CMAC-based nonlinear control with dynamic memory architecture, *Journal of the Franklin Institute*, vol.348, no.9, pp.2480-2502, 2011.