Facial Expression Recognition
Using A CMAC Network with Clustering Memory

Yu-Yi Liao
National Cheng Kung University
Email:liao.yuyi@gmail.com

Jzau-Sheng Lin
National Chin-Yi University of Technology
Email:jslin@chinyi.ncit.edu.tw

Shen-Chuan Tai
National Cheng Kung University
Email:sctai@mail.ncku.edu.tw

Abstract

In this paper, a facial expression recognition system based on CMAC (Cerebellar Model Articulation Controller) with a clustering memory is presented. Firstly, the facial expression features were automatically extracted and preprocessed from given still images in the JAFFE database in which the frontal view of faces were contained. A 2D DCT was then used to focus the key information of expression characteristics in order to decrease the size of images. Thirdly, a block size of the lower frequency of DCT coefficients is rearranged as input vectors with binary manner to send into the proposed CMAC that can rapidly obtain output using non-linear mapping with a look-up table in training or recognizing phase. Finally, the experimental results demonstrated recognition rates with various block size of coefficients in lower frequency and learning rates to show promising recognition results.

Key words: CMAC, Facial expression recognition, Feature extraction

1. Introduction

Human being is a social animal and their facial expressions play a significant role in communication. Facial expressions are facial movements in response to the internal affective state, psychological state, and cognitive activity of a man. In recent years facial expression recognition has become an active research topic and a number of methods have been proposed in the literature [1-3, 7-9, 13, 14].

Ekman and Friesen [10, 11] proposed six basic emotions, the standard of classification of recognition, that possess each a distinctive content together with a unique facial expression in 1971 [10] and developed the well-known Facial Action Coding System (FACS) for facial expression description [2, 11]. The FACS provided 46 Action Units (or AUs) to describe any facial expression of human being and indicates eyebrows, eyes and mouth are the key feature for facial expression recognition [7, 14].

The two dimensional discrete cosine transform (or 2D DCT) is often used in signal and image compression, because it can concentrate the most of signal information into low frequency components. It also had been utilized in 2D facial expression recognition in recent years. In references 7 and 14, they used the entire facial image to obtain lower-frequency 2D DCT coefficients and used to train a neural network.

The most of neural networks spend more time at training phase except CMAC which proposed by Albus [4-6]. CMAC is a supervised neural network of associative memory based on table look-up method. The advantages of CMAC are fast learning, simple computation, local generalization, non-linear mapping, and can be easily implemented by the hardware. In this paper, a new recognition technique is proposed, which used the 2D DCT over the entire face image as a facial expression feature extractor and the CMAC with clustering memory as a classifier. The proposed recognition system used the CMAC with clustering memory in order to speed up the processing time of training and recognition phases.

The paper is organized as follows. Section 2 gives a brief review of the traditional CMAC architecture. Following, 2D DCT to extract facial features and the CMAC network used to the facial expression recognition are presented in Section 3. Section 4 describes the preprocessing to de-noise and normalize the face images for feature extraction. The experimental results will be demonstrated at section 5. Finally, Section 6 gives the conclusions

2. The traditional CMAC neural network

The architecture of the traditional CMAC, as shown in Fig. 1, is essentially a four level with three subsequent mappings supervised neural network. The first one projects each vector s in input space X into co-relational t cell in virtual associative memory A with a non-linear mapping function $S(s)$.

$$S : X \overset{S(s)}{\Rightarrow} A \tag{1}$$

where X and A are S-dimensional input vector space and the total number of cells respectively.

The second mapping utilizes the hash coding technique to compress the values of associative memory space A into the address in physical memory space W, for reducing the memory space.

$$H : A \overset{H(t)}{\Rightarrow} W \tag{2}$$

where $W = \lfloor w_1, w_2, \cdots, w_p \rfloor$ is P-dimensional address space or weight vector. $H(t)$ is a non-linear mapping function. P is an arbitrarily chosen association coefficient which is the important property of the CMAC network called generalization parameter. The CMAC weight adjustment can be defined as

$$w_{(i+1)} = w_{(i)} + \frac{\beta(y_d - y)}{g} \tag{3}$$

where β, y_d, y and g are the learning rate, expected output, existent output, and generalization constant respectively. The last mapping sums up all of the weight values in physical memory Y as the output y of CMAC.

$$y = t^T W = \sum_{i=1}^{P} w_i S_i(s) \tag{4}$$

where w_i is the weight and $S_i(s)$ is the mapping function for the ith component of an input vector.

3. The 2D DCT and the proposed CMAC network

The facial expression recognition system based on 2D DCT and the CMAC network with a clustering memory, as shown in Fig. 2, is proposed in order to recognize facial expression, in which a difference image by subtracting a neutral image from a given expression image was obtained as an input vector of the CMAC. The difference images retain the key characteristics of recognition, but they still have an amount of data. The 2D DCT is often used in reducing the size of data, because it has a powerful compression property, that is, most of the signal information tend to be concentrated in a few low-frequency DCT coefficients.

The formula of 2D DCT was used as following

$$X(k_1, k_2) = \frac{4C(k_1 k_2)}{M^2} \sum_{i=0}^{M-1} \sum_{j=0}^{M-1} p(i, j) \cos\frac{(2i+1)k_1\pi}{2M} \cos\frac{(2j+1)k_2\pi}{2M}$$

$$k_1, k_2 = 0,1,2,\ldots, M-1 \tag{5}$$

and

$$\begin{cases} C(k) = 1 & \text{for } \ k \neq 0 \\ C(0) = \dfrac{1}{\sqrt{2}} & \text{otherwise} \end{cases}$$

where $p(i, j)$ is the gray-level of 2D difference image. Therefore, a square of the magnitudes of low-frequency 2D DCT coefficients ($X(k_1, k_2) : k_1, k_2 = 0, 1, \cdots, L-1$) was quantized into a size of $S = L^2 * N$ bits of binary number and represented as an S-bits input vector to be fed into the proposed CMAC network.

An S-bit input vector is encoded to quantize into c clusters which including b bits, namely $S = c * b$, for extending the input space without overlap. The output layer is k-layer parallel memories to memories k facial expressions. In training phase, an S-bit test input vector only excited and adjusted the weight in the *i*th address by encoding, quantizing the excite weights in proper memories, and then sum up the exciting address. The weight can be adjusted with the steeping descent rule as follows:

$$w(k+1) = w(k) + \eta \Delta w(k) \tag{6}$$

where

$$\Delta w(k) = -\frac{\partial E}{\partial y}, \quad E = \frac{1}{2}\sum_{i=1}^{c}(y_d - y)^2 \tag{7}$$

and

$$w(k+1) = w(k) + \eta(y_d - y) \tag{8}$$

where E, η, y_d, and y are the error function, learning rate, expected output, and existent output respectively. The CMAC network used in this paper was successfully applied to the character recognition in 2005 [13].

In recognition phase, an output value for all objects is obtained by exciting the weight memories of all layers, and the output value "1" is represented the input vector recognized.

4. Preprocessing for facial feature extraction

As shown in Fig. 3(a), the images in the JAFFE database contain the frontal view of face of Japanese females which are included the variant contrast, noises, and angle of heads. Therefore, the images were needed preprocessing to normalize and extract the facial expression features for locating the eyes at the same horizontal line to yield a "head-in-the-box" format image as shown in Fig. 3(f). The partially preprocessing steps are described as follows.

Remove noises: First of all, we had to remove noises from the image and used median filter to remove noise and preserve edges simultaneously [12]. It replaced the pixel value with the median of the gray levels in the neighborhood of that pixel. The filtering result is shown in Fig. 3(b).

Enhance the contrast: Enhancement refers to increasing or sharpening of image features, like contrast, boundaries, edges, etc. As shown in Fig. 3(c), logarithmic transformations were used frequently for image enhancement and also employed on this paper [12]. Log transformation maps a narrow range of low gray-level values in the input image into a wider range of output levels, so it shows considerable ability to compress the dynamic range of images with large variations in pixel values

Detect the face edge: Edges carry a lot of information about the various regions in the image. We utilized Sobel operator to find out the edge of face in the image and the result is shown in Fig. 3(d). Sobel method is a 3*3 neighborhood based gradient operator and has superior noise-suppression characteristics [12].

Rotate the image: The images in database have various angles of heads, so this step is to rotate the image by an angle $-\theta$,

$$\theta = \arctan\left[\frac{lr - rr}{rc - lc}\right] \tag{9}$$

where (lr, lc) are the left-eye row and column coordinates and (rr, rc) are the right-eye row and column coordinates respectively. The rotated result is shown in Fig. 3(e).

Crop the image: FACS decomposed various facial actions into the related muscle activities and it provides the main features of facial expression relative to the neutral face. This step is to crop the size of 256*256 images into the small images which only have the main expression features including eyebrows, eyes and mouth for obtaining the good difference image that only contain the distinguish feature between expression and neutral. As shown in Fig. 3(f), the two eyes and mouth are taken as the reference points to place the eyes of all of the images at the same horizontal line and to cut them into a size of 104*104 manners, which preserves the relative feature positions on the face.

5. Experimental results

The facial express patterns were obtained from JAFFE database and the experiments were simulated in a Pentium-IV personal computer by using of the language MATLAB.

Facial expression database: The four facial expressions (neutral, happy, sadness and surprise) of 10 Japanese female models in JAFFE database were used in our experiments. Each person was recoded three or four images for each expression. The size of each image is 256*256 pixels. In our simulation experiments, two images of each expression were used for network training and the remainder images were used for testing, namely 60 subjects for training and 30 subjects for testing. After preprocessing phase, all of the images were already normalized and cropped the size into 104*104. The JAFFE database was also used in references 3, 8, 9, 15, and 16. Fig. 4 shows the samples of cropped images from the JAFFE database.

The CMAC with clustering memory: Three factors will affect recognition rate in our experiments. First, the 2D DCT transforms a difference image, obtained by subtracting a neutral image from a given expression image, from a spatial domain into the frequency domain where the lower-frequency coefficients keep the crucial and relevant information of recognition. It is important to determine the block size L^2 of the lower-frequency DCT coefficients because it yields a great quantity training data and obtained better recognition rate with a large memory if the block size is large and opposite situation if block size is small. Therefore, we quantized the different block size of lower-frequency coefficients to binary numbers of $S = L^2 * N(N = 20)$ bits as an input vector and sent to the proposed CMAC network. Each input vector had 10 runs with varied initial value of weights memory that is encoded to quantize into the various cluster size. The results of experiment, shown in Table 1, reveal that the size of 8*8 is the best block size because it is not only obtained a better recognition rate (93.33%) but also use less memory than the block size of 14*14.

The second factor is the cluster size. The proposed CMAC encoded the $S = L^2 * N(N = 20)$ bits input vectors to quantize to c cluster which containing b bits, that is, $S = c * b$. If the cluster size is small, the CMAC will have more clusters. That is the components of error in exciting weights will be distributed on more places in memory. Then the error will not clearly affect the output to obtain better recognition rate. Therefore, the used CMAC has the capacity of tolerate or against learning interference to promote the recognition rate. Table 1 shows the recognition rate with various cluster size and the model of 4-bit cluster yields better recognition rate about 93.33%.

The last factor is learning rate. The gradient descent is very slow with a small learning rate as well as the gradient decent is rapid and oscillates widely if the learning rate is bigger. The different learning rates also influence the recognition rate, as shown in Table 1; learning rate with a value of 0.05 could obtain better recognition rate about 93.33%.

Finally, we take a closer look at better performance of the proposed CMAC that correspond to the block size of 8*8 with a 4-bit cluster is selected, and a suitable learning rate is 0.05. Tables 2, 3, and 4 demonstrate that the confusion matrix for three expressions of better recognition rate about 93.33% obtained by the used CMAC with a 4-bit cluster and the block sizes 8*8, 12*12, and 14*14 respectively.

6. Conclusion

To sum up, from the experiment results lead to the conclusion that the facial expression characteristics had been automatically extracted and cropped the frontal view of face images in the JAFFE database from 256*256 to 104*104 for reducing the size of images. The 2D DCT transform difference image from spatial domain to frequency domain to yield an 8*8 block of lower frequency coefficients as the input signals that were quantized as a binary manner and fed into the CMAC. Finally, the CMAC used a proper size of 4-bit cluster and a learning rate 0.05 has successfully obtained a better recognition rate about 93.33%.

Acknowledgment

The authors thank Dr. Michael J. Lyons for providing the face expression database and more information at [17].

References

[1] Claude C, Chibelushi and Fabrice Bourel, " Facial Expression Recognition: A Brief Tutorial Overview, " *In CVonline: On-Line Compendium of Computer Vision* (2003)

[2] Gianluca Donato, Marian stewart Bartlett, Joseph C. Harger, Paul Edman, and Terrence J. Sejnowski, "Classifying Facial Actions," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 21, No. 10 pp. 974~989 (1999)

[3] Guodong Guo and Charles R. Dyer, "Learning From Examples in the Small Sample Case: Face Expression Recognition," *IEEE Transaction on Systems, Man, and Cybernetics—Part B: Cybernetics*, Vol. 35 No. 3 pp. 477~488(2005)

[4] James S. Albus, "A Theory of Cerebellar Function," *Mathematical Biosciences*, Vol. 10, pp. 25-61 (1971).

[5] James S. Albus, "A New Approach to Manipulator Control: The Cerebellar Model Articulation Controller (CMAC)," *Journal Dynamic System Measurement and Control, Transactions ASME 97*, pp. 220-227 (1975).

[6] James S. Albus, "Data Storage in the Cerebellar Model Articulation Controller (CMAC)," *Journal Dynamic System Measurement and Control, Transactions ASME 97*, pp. 228-233 (1975).

[7] L. Ma and K. Khorasani, "Facial Expression Recognition Using Constructive Feedforward Neural networks," *IEEE Transactions on Systems. Man. And Cybernetics—Part B: Cybernetics*. Vol., 34 No. 3 pp. 1588~1594 (2004)

[8] M. J. Lyons, S. Akamatsu, M. Kamachi, and J. Gyoba. "Coding facial expressions with Gabor wavelets,"

Third IEEE Intl. Conf. on Auto. Face and Gesture Recog., pp. 200~205(1998)

[9] M. J. Lyons, J. Budynek, and S. Akamatsu, "Automatic classification of single facial images," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 21, no. 12, pp. 1357~1362(1999).

[10] Paul Ekman and W.V. Friesen, "Constants Across cultures in the Face and Emotion," *Journal of Personality and Social Psychology*, 17(2): pp. 124~129(1971)

[11] Paul Ekman and W. Friesen, "Facial Action Coding system: A Technique for the Measurement of Facial Movement," *Palo Alto, Calif.: Consulting Psychologists Press*(1978)

[12] Rafael C. Gonzalez and Richard E. Woods, "Digital Image Processing", *Prentice-Hall*, (2002)

[13] Shao-Han Liu and Jzau-Sheng Lin, "Character Recognition Based on CMAC with Field Programmable Gate Array," *Journal of Technology*, Vol. 20, No. 4 pp. 357~366 (2005)

[14] Yegui Xiao, N. P. Chandrasiri, Yoshiaki Tadokoro, and Masaomi Oda, "Recognition of facial expressions using 2D DCT and neural network", *Electronics and Communication in Japan*, vol.82, no 7, pp. 1-11, 1999

[15] Z. Zhang, M. Lyons, M. Schuster, and S. Akamatsu, "Comparison between geometry-based and Gabor-wavelets-based facial expression recognition using multi-layer perceptron," *Proc. 3rd IEEE Int. Conf. Automatic Face and Gesture Recognition*, pp. 454–459(1998)

[16] Z. Zhang, "Feature-based facial expression recognition: Sensitivity analysis and experiments with a multi-layer perceptron," *Pattern Recognition and Artificial Intelligence*, vol. 13, no. 6, pp. 893~911 (1999).

[17] The Japanese Female Facial Expression Database: http://www.kasrl.org/jaffe .html

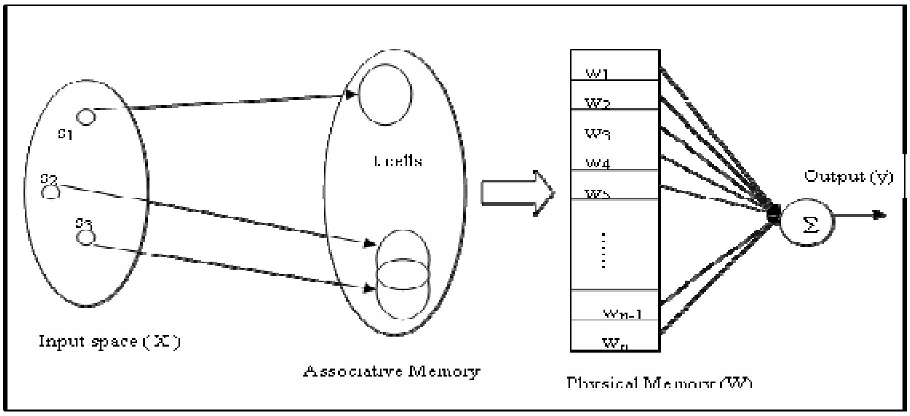

Fig. 1: Architecture of the traditional CMAC

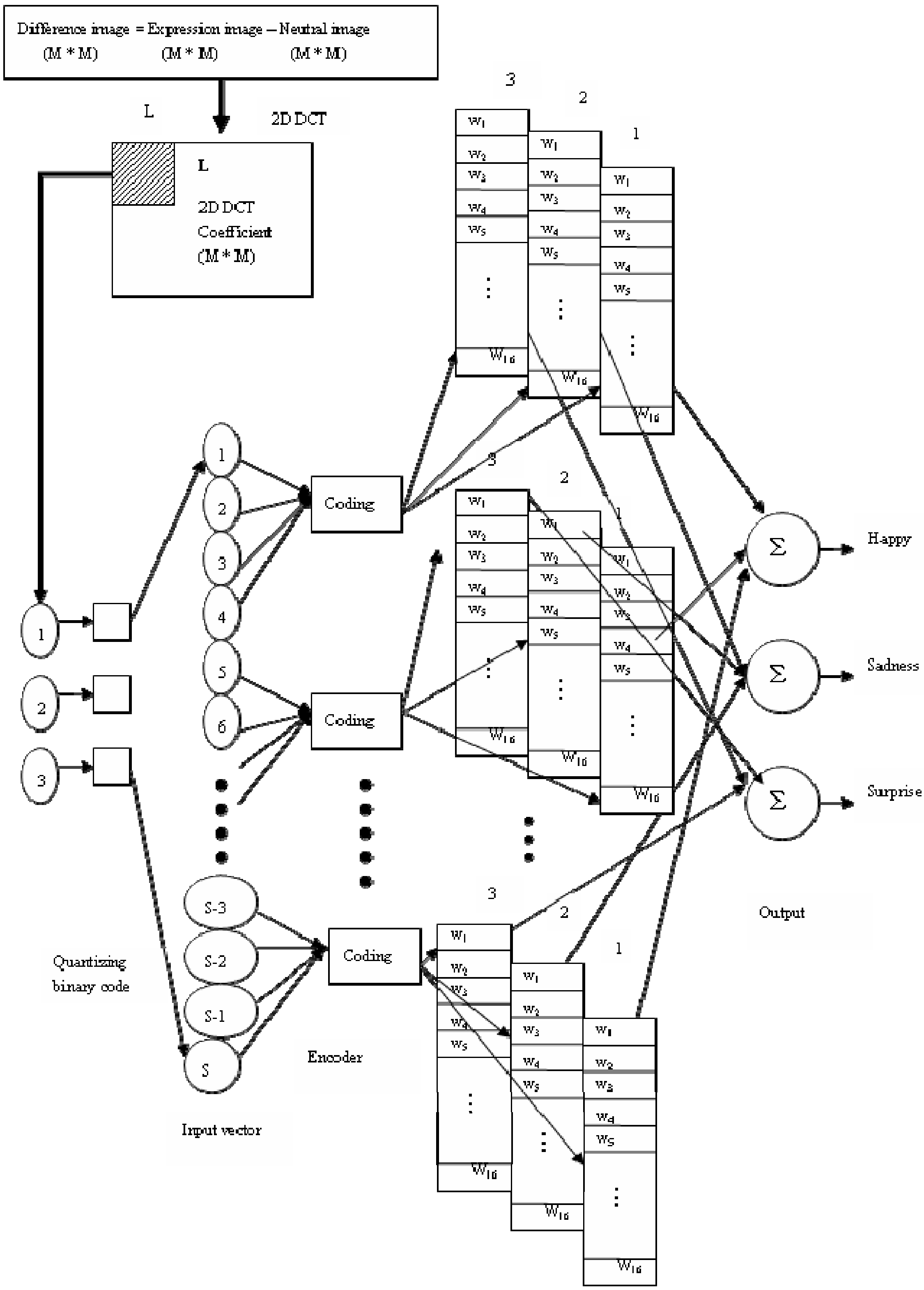

Fig. 2: Architecture of the facial expression recognition of CMAC with clustering memory

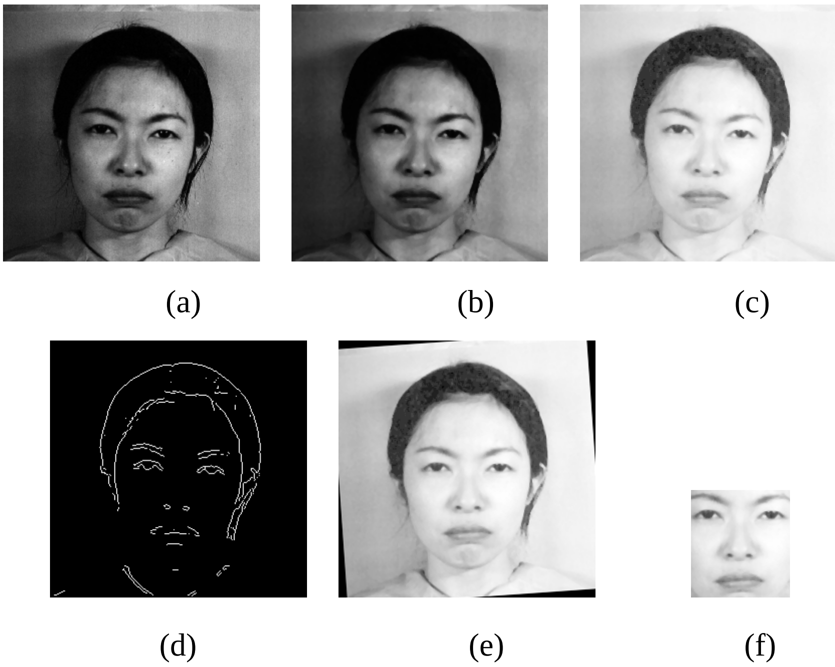

Fig. 3: The results of preprocessing: (a) original image, (b) removed noises, (c) enhanced contrast, (d) detected edges, (e) rotated image, and (f) cropped image

Neutral Happy Sadness surprise

Fig. 4: samples of cropped image of the JAFFE database

Table 1 Recognition rate versus the block size, cluster size, and learning rate obtained testing the proposed CMAC

Cluster size	Learning rate	Testing Recognition rate of different block size.			
		14*14	12*12	8*8	6*6
10 bits	0.005	83.33%	86.67%	83.33%	76.67%
	0.05	83.33%	83.33%	80%	76.67%
	0.15	70%	76.67%	70%	63.33%
5 bits	0.05	**93.33%**	90%	90%	90%
	0.15	**93.33%**	86.67%	73.33%	86.67%
4 bits	0.05	**93.33%**	**93.33%**	**93.33%**	86.67%
	0.15	**93.33%**	90%	86.67%	86.67%

Table 2 Confusion matrix of each expression obtained from the best recognition rate 93.33% (the size 8*8 of block and the size 4-bit of cluster size)

	Happy	Sadness	surprise
Happy	10	0	0
Sadness	0	10	0
surprise	1	1	8
Mean recognition rate 93.33%			

Table 3 Confusion matrix of each expression obtained from the best recognition rate 93.33% (the size 12*12 of block and the size 4-bit of cluster size)

	Happy	Sadness	surprise
Happy	10	0	0
Sadness	1	9	0
surprise	1	0	9
Mean recognition rate 93.33%			

Table 4 Confusion matrix of each expression obtained from the best recognition rate 93.33% (the size 14*14 of block and the size 4-bit of cluster size)

	Happy	Sadness	surprise
Happy	10	0	0
Sadness	1	9	0
surprise	1	0	9
Mean recognition rate 93.33%			