運用人臉特徵及類神經網路於人臉表情辨識系統
A Facial Expression Recognition System Using Facial Feature Extraction and Artificial Neural Networks

江妍瑤* 張元翔 李忠驊 楊靜杰
Yan-Yau Jiang*, Yuan-Hsiang Chang, Chung-Hua Li, Jin-Jie Yang

摘要

臉部表情是人類表示情緒狀態的重要特徵。本研究整合臉部特徵（例如：眉毛、眼睛等）位置變化的情況來定義表情，發展出一套「臉部表情辨識」系統，從臉部影像辨別各種表情（例如：無表情、微笑、生氣、悲傷和快樂）。首先利用影像處理技術自動擷取臉部特徵，將其對一參考軸正規化，且將這些經正規化特徵作為類神經網路之輸入，而類神經網路之輸出即為表情辨識結果。本系統使用本研究室建立之560個臉部影像的表情資料庫，其中280個為訓練樣本，280個為測試樣本。另外並採用JAFFE表情資料庫作為測試樣本。研究結果顯示，本系統可達到約88%的辨識率，JAFFE表情資料庫辨識率約87%。實驗結果顯示本系統對於臉部表情可達到初步成功之辨識結果。

關鍵詞：類神經網路，表情辨識，特徵擷取，影像處理

Abstract

Facial expression is an important human characteristic that indicates a person's motional state. In this paper, facial expression is defined as the combined results of position changes in facial features (e.g., eyebrows, eyes, etc.). We propose a facial expression recognition system to distinguish various facial expressions (i.e., neutral, smile, angry, sad, or happy) using face images. Facial features are first automatically extracted using image processing techniques and normalized with respect to a central axis. These normalized features are then used as inputs to an artificial neural network (ANN) and the ANN outputs are used to indicate the potential facial expression. During the experiment, 560 face images were collected in our laboratory, among which, 280 images were formed as the training set, and the remaining 280 images were used for testing. In addition, the JAFFE database was used as testing samples. Our results demonstrated that the system has achieved a reasonable detection rate of approximately 88% in our database and 87% in the JAFFE database. In summary, the system was shown to recognize given facial expressions with preliminary success.

Keywords: artificial neural network, facial expression, feature extraction, image processing

I. INTRODUCTION

Facial expression is an important non-verbal communication that often conveys emotional state of human [1, 2]. Facial expression is generally referred to as the combined result of motions or position changes in facial features, such as eyebrows, eyes, nose, mouth, and/or face muscle. With continuous advances of computer technology to date, computers have played an important role in human's daily life. Therefore, an ultimate goal of today's computer scientists is to achieve a friendly human-computer interaction [3, 4]. With this regard, facial features or expressions are the most important clues for the computer to recognize and interact with its human users. Therefore, recognition and analysis of facial features and facial expressions have drawn the attention and become the research of interest by many computer scientists. Face images are typically captured from a person using digital camera, webcam, camcorder, or video camera. These face images typically contain the person's facial information, including his or her facial features that can be used to identify the person himself/herself. In addition, the face images may contain the person's facial expression that can be used to determine the person's emotional state, thus providing an extra clue to the computers.

中原大學資訊工程系
*Corresponding author. Email: riverrock@gmail.com
Department of Information and Computer Engineering, Chung Yuan Christian University, Chung Li, Taiwan, R.O.C.
Manuscript received 18 April 2007; revised 5 June 2007; accepted 21 December 2007

During the past years, researchers have focused on two main aspects of the face-related researches, namely the face recognition and the facial expression classification. Face recognition [5, 6] is a technique to identify the person in face images. The main processes generally include image processing techniques to locate regions of interest that may contain possible human faces in images, and artificial intelligence techniques to recognize relevant facial features for personal identification. Typical applications include personal authorization, building entrance control, or video surveillance system. Instead, facial expression classification [7-9] is a technique to identify the person's facial expression that may reflect the person's emotional response. Facial expression classification is however based on the results of face recognition [10-13]. Its applications may include medical care, facial expression simulation and human-computer interaction.

The objective of this research was to utilize a facial feature extraction method and an artificial neural network for the computerized analysis, and to construct a *facial expression recognition system* to quantitatively characterize human's facial expression. This system may allow us to produce a potential clue to the computer such that an effective response from the computer could be generated, thus achieving a relatively interactive human-computer interface.

II. MATERIAL AND METHODS

Figure 1 shows the simplified flow diagram of the facial expression recognition system. The system included several main processes, i.e., (1) face image acquisition; (2) facial feature extraction; (3) artificial neural network; and (4) facial expression recognition. Given a digital face image, facial features were extracted first. These facial features and their relationship were normalized with respect to a reference, namely the central axis. The normalized facial features were then used as inputs to an artificial neural network (ANN). As a result, the outputs of the ANN were used to indicate the likelihood of a specific facial expression (i.e., neutral, smile, angry, sad, or happy) for the given face image. Detail description of each main process follows.

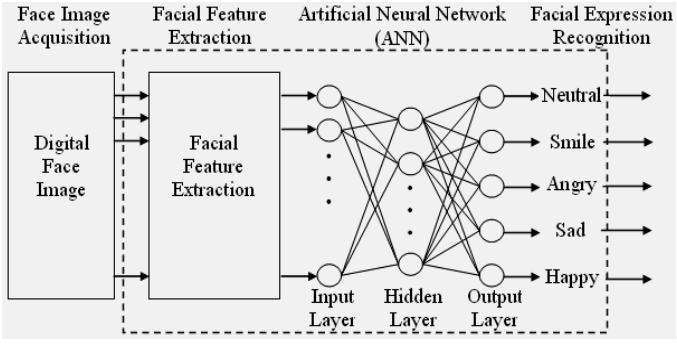

Fig. 1 Simplified flow diagram of the facial expression recognition system

1. *Face Image Acquisition*

To form our facial expression database, a digital camera (Sony Cyber-Shot DSC-F505V) was used to capture the digital images. A total of 28 people (25 males and 3 females), including members of our laboratory and students from our department, were invited to participate in this research. During the image acquisition, each person was asked to face the camera in a frontal position. Therefore, each face image contains single face from a person. In addition, each person was asked to repeat 4 times for 5 different facial expressions (i.e., neutral, smile, angry, sad, or happy). A typical example is given in Figure 2. During the image acquisition, two independent observers were also invited to subjectively determine if the given facial expression is acceptable. In this study, the definitions of facial expressions as described by Stathopoulou and Tsihihrintzis [16] were used as the criteria for subjective classification by the observers in our facial expression database. Here, neutral expression can be referred to as no expression as well. Other facial expressions (e.g., smile, angry, sad, etc.) are variations depart from the neutral expression, resulting in position changes of facial features [16]. All images were obtained indoor with fluorescent light condition and white background in our laboratory and were acquired with 640 × 480 pixels per image. As a result, a total of 560 face images were acquired. The digital images were stored and processed using bitmap file format with no compression.

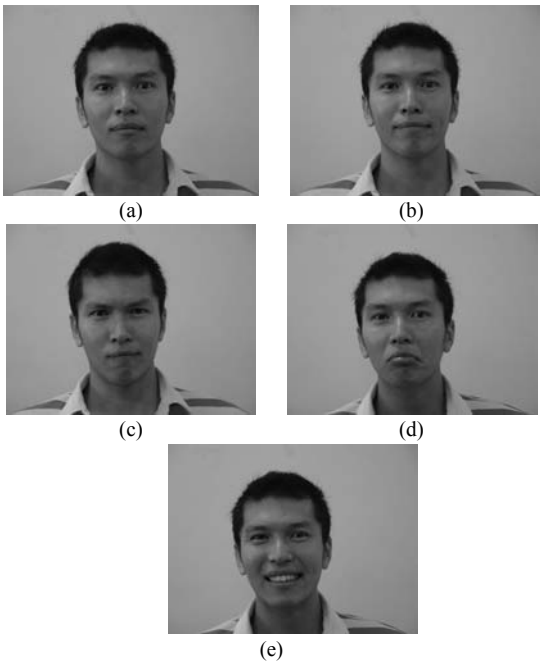

Fig. 2 Various facial expressions of interest: (a) neutral or no expression; (b) smile; (c) angry; (d) sad; or (e) happy

2. Face Feature Extraction

Facial expression is defined as the combined results of one or more position changes of facial features, such as eyebrows, eyes, nose, mouth, and/or face muscle. To quantitatively characterize the position changes of facial features, 17 facial features were first defined as given in Figure 3. These features were defined at specific positions such as corners for eyebrows, eyes, nose, and mouth, because facial expression generally occurs when these positions vary in positions [14-16].

In our system, automatic facial feature extraction was implemented and can be described herein [17]. First, the facial region was partitioned into four independent regions, namely (1) eyebrows; (2) eyes; (3) nose; and (4) mouth regions, as shown in figure 4. Image processing techniques were applied to each of the four regions for the automatic extraction of facial features of interest. The techniques included: (1) color conversion; (2) facial region partition; (3) automatic thresholding; and facial feature search. Figure 5 demonstrates an example of the automatic extraction of facial features where 17 facial features were extracted accordingly.

Because different people may depict various face sizes in digital images, a normalization technique was applied prior to the facial expression recognition. Suppose the Euclidean distance between two points P_1, and P_2 is defined by:

where $P_1 = (x_1, y_1)$ and $P_2 = (x_2, y_2)$ are the corresponding image coordinates for the two points. We chose a relatively large distance in a face, called *central axis*, as the reference axis for normalization. The central axis was defined as the center point (midpoint between the *B* and *C*) extended to the point *O* as shown in Figure 6. Suppose the Euclidean distance of the central axis is defined as $C_{d,}$ we then computed the following 14 facial feature distances with respect to the central axis as follows [14-16]:

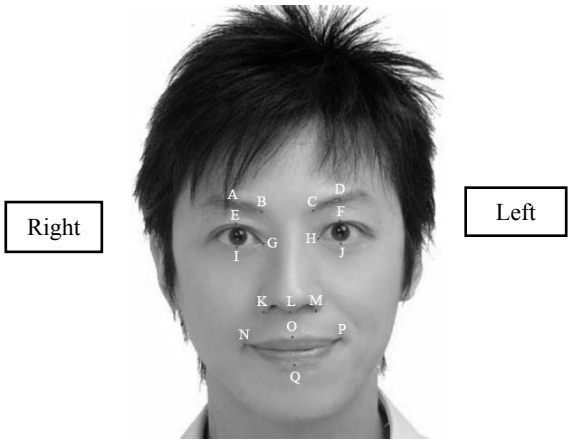

Fig. 3 The 17 facial features (marked *A* through *Q*) as extracted for the facial expression recognition system. The facial features were marked in the alphabetic order from top to bottom, and from right to left according to the observed face

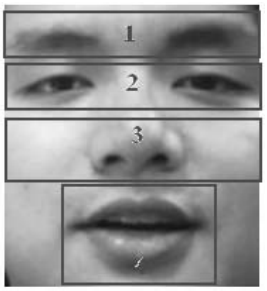

Fig. 4 Four expression areas

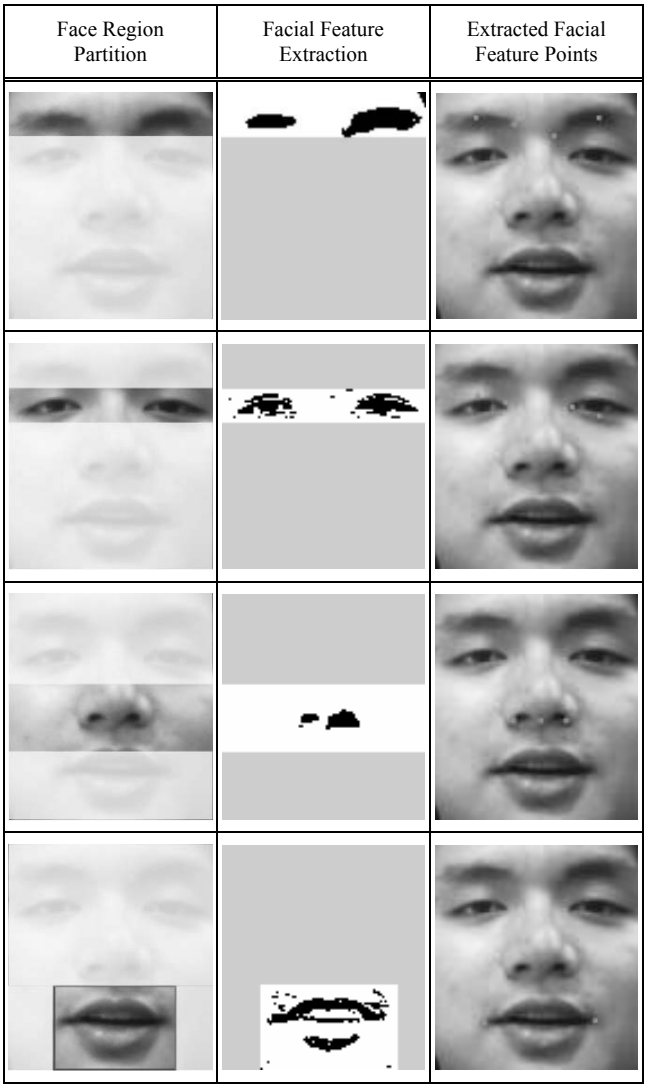

Fig. 5 Automatic extraction of facial features

$$
\begin{aligned}
&\text{Feature distance 1:} \quad f_1 = d(B, C)/\, C_d; \\
&\text{Feature distance 2:} \quad f_2 = d(A, E)/\, C_d; \\
&\text{Feature distance 3:} \quad f_3 = d(B, G)/\, C_d; \\
&\text{Feature distance 4:} \quad f_4 = d(C, H)/\, C_d; \\
&\text{Feature distance 5:} \quad f_5 = d(D, F)/\, C_d; \\
&\text{Feature distance 6:} \quad f_6 = d(E, I)/\, C_d; \\
&\text{Feature distance 7:} \quad f_7 = d(F, J)/\, C_d; \\
&\text{Feature distance 8:} \quad f_8 = d(I, N)/\, C_d; \\
&\text{Feature distance 9:} \quad f_9 = d(J, P)/\, C_d; \\
&\text{Feature distance 10:} \quad f_{10} = d(K, N)/\, C_d; \\
&\text{Feature distance 11:} \quad f_{11} = d(M, P)/C_d; \\
&\text{Feature distance 12:} \quad f_{12} = d(L, O)/\, C_d, \\
&\text{Feature distance 13:} \quad f_{13} = d(O, Q)/C_d; \\
&\text{Feature distance 14:} \quad f_{14} = d(N, P)/C_d;
\end{aligned} \tag{1}
$$

$$d(P_1, P_2) = \sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2} \tag{2}$$

where $d(P_1, P_2)$ is the Euclidean distance for the two given points (A through Q). Therefore, all facial feature distances were normalized with respect to the distance of the central axis C_d. In neutral expressions, all these feature distances were therefore normalized to the range between 0 and 1.

3. Artificial Neural Network

Because facial expression is a combined result from various facial feature changes, techniques to analyze and integrate multi-dimensional features for a generalized output are desired. The artificial neural network is a "black-box" approach [18-19] and is used to automatically infer the facial expression given the multi-dimensional features. In this research, the artificial neural network was a three-layer feed-forward network [20]. There were one input layer with 14 neurons, one hidden layer with 9 neurons, and one output layer with 5 neurons. Inputs were the 14 facial feature distances as defined in Equation (1). Each output neuron, however, yielded the likelihood (or probability) for each of the 5 different facial expressions. During the training, the objective was to minimize the error function E as defined by:

$$E = \left(\frac{1}{2}\right)\sum_j \left(T_j - Y_j\right)^2 \tag{3}$$

where T_j are the target outputs of the jth neuron, and Y_j are the actual outputs of the jth neuron in the output layer. The supervised learning was used to update the inter-connected weights for the network using the generalized delta rule [19]:

$$w_{ij}(n+1) = w_{ij}(n) + \Delta w_{ij}(n) \tag{4}$$

where the inter-connected weight w_{ij} from ith neuron to the jth neuron was updated by Δw_{ij} iteratively. For each training cycle, errors were back-propagation from the output layer to the input layer. In addition, the number of neurons in the hidden layer was chosen using the pruning strategy [19]. In practice, the learning rate (η = 0.1) and the momentum (α = 0.9) were chosen during the training of the ANN.

4. Facial Expression Recognition

For the network training, the following target outputs T_j , j =1,2,…5, were given for various facial expressions accordingly: i.e., T [1..5] = {1, 0, 0, 0, 0} for "neutral"; T [1..5] = {0, 1, 0, 0, 0} for "smile"; T [1..5] = {0, 0, 1, 0, 0} for "angry"; T [1..5] = {0, 0, 0, 1, 0} for "sad"; or T [1..5] = {0, 0, 0, 0, 1} for "happy", respectively. To determine which facial expression category the outputs should be classified (neutral, smile, angry, sad, or happy), we simply used the maximum among the results from the 5 output neurons as the identified facial expression by:

$$\arg_j Y_j = \max(Y_j, j = 1,....5) \tag{5}$$

where $0 \leqq Y_j \leqq 1$, j =1,2,…5, yielded by the artificial neural network.

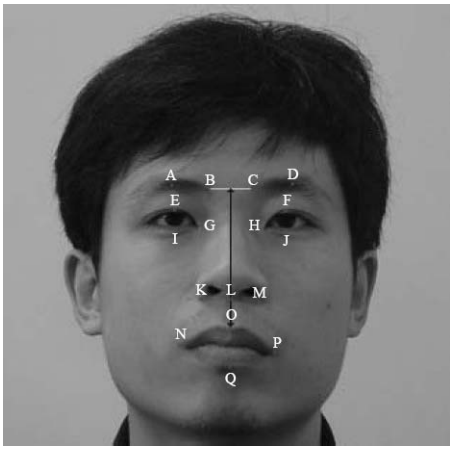

Fig. 6 Location of the central axis used as the reference axis. The central axis was defined as the line from the midpoint between B and C (between eyebrows) extended to the point O (upper lip)

III. RESULTS

Two facial expression databases were collected for the performance evaluation of our system. One is our facial expression database which contains a total of 560 digital images (each with single face), which were acquired from 28 people, including 25 males and 3 females. Among them, 14 people were randomly chosen as the training subjects, and the remaining 14 people were used as the testing subjects. During the image acquisition, each person was asked to repeat 4 times for 5 different facial expressions (i.e., neutral, smile, angry, sad, or happy), resulting in a database of 560 face images. Among these images, 280 face images as obtained from the training subjects were used to train the network, while the remaining 280 face images as obtained from the testing subjects were used to test the network.

The other database for the system evaluation was the Japanese Female Facial Expression (JAFFE) database

which contains 10 subjects with 7 facial expressions [21]. To facilitate our study, a set of 90 digital images containing 4 facial expressions (i.e., neutral, angry, sad, or happy) were collected from the JAFFE database in an attempt to increase our test sample size. Figure 7 shows an example of the 4 facial expressions of a Japanese female as obtained from the JAFFE database. Among the JAFFE database, facial expressions have been previously categorized and were incorporated for system evaluation after our system has been trained using our database.

Table 1 summarizes the detection results of the facial expression recognition system in our facial expression database. As can be seen, our system has achieved the detection rate of 95.7% in the training set, and the detection rate of 87.9% in the testing set. The root-mean-squared error (RMSE) convergence curve of the ANN during the training is given in Figure 8. It obviously shows that the RMSEs converge after about 600 iterations.

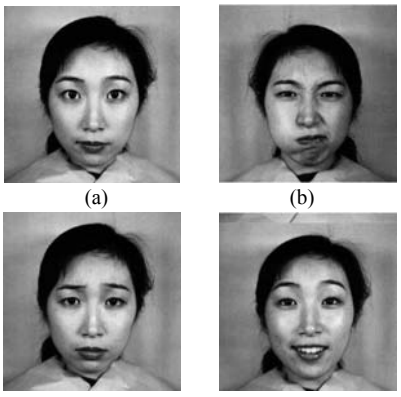

Fig. 7 Various facial expressions of interest from JAFFE: (a) neutral; (b) angry; (c) sad; or (d) happy

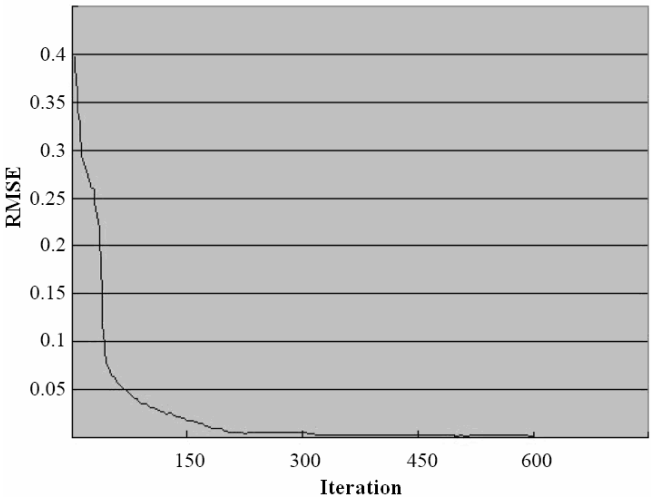

Fig. 8 The RMSE convergence curve

Table 1 Detection results of the facial expression recognition system in our database

Results \ Data Set	Training Set	Testing Set
Number of Images	280	280
Correctly Identified	268	246
Detection Rate (%)	95.7%	87.9%

Table 2 Misclassifications in our database (training set with 280 images)

Result \ Actual	Neutral	Smile	An-gry	Sad	Happy	Error Rate
Neutral	0	1	0	2	0	1.1%
Smile	2	0	0	1	1	1.4%
Angry	0	0	0	1	0	0.4%
Sad	1	1	2	0	0	1.4%
Happy	0	0	0	0	0	0%

Unit: number of images

Table 3 Misclassifications in our database (testing set with 280 images)

Actual \ Result	Neutral	Smile	An-gry	Sad	Happ y	Error Rate
Neutral	0	2	1	2	0	1.8%
Smile	4	0	3	4	2	4.6%
Angry	0	0	0	1	0	0.4%
Sad	4	4	4	0	2	5%
Happy	0	1	0	0	0	0.4%

Unit: number of images

To further understand the misclassifications and their causes, we analyzed the number of images (or misclassifications) for the 5 different facial expressions. The results for the training set or the testing set are summarized in Table 2 or 3, respectively. In both tables, the error rate (%) was defined as the ratio between the number of misclassifications for a specific facial expression and the number oftotal images. Further, examples of the misclassifications are also shown in Figure 9.

As can be seen from both tables, the system has the tendency to misclassify especially for the two facial expressions, i.e., smile and sad. The reason for such misclassifications was because the position changes of the facial features (eyebrows, eyes, and mouth) involved in the two facial expressions were not significant, as compared with other facial expressions (i.e., angry or happy).

Table 4 summarizes the detection results of the system for the JAFFE database. As can be seen, our system has achieved the detection rate of 86.7% with comparable performance with the testing set in our database. Table 5 demonstrated the results of misclassifications in the JAFFE database. Further, several examples of the misclassifications are also shown in Figure 10.

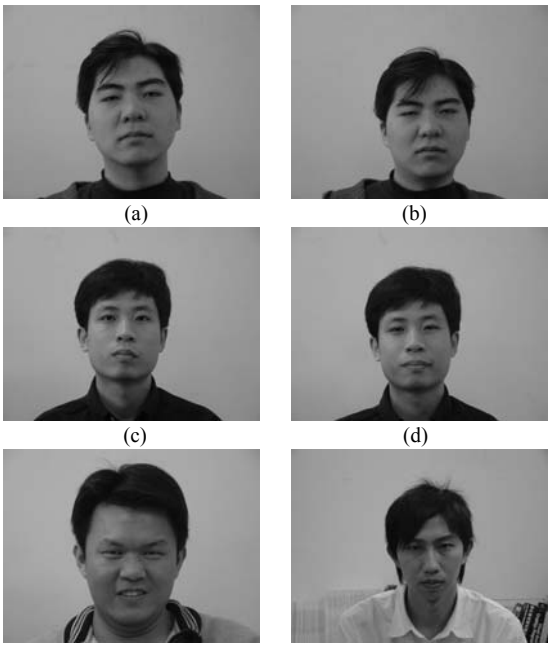

Fig. 9 Examples of misclassifications in our database: (a) "neutral" classified as "smile"; (b) "sad" classified as "angry"; (c) "neutral" classified as "sad"; (d) "sad" classified as "neutral"; (e) "angry" classified as "happy"; (f) "angry" classified as "neutral"

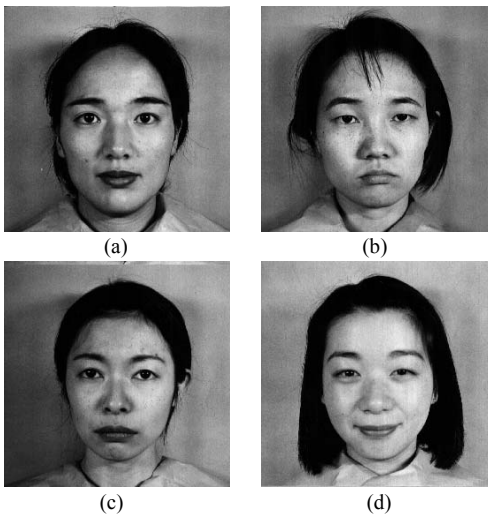

Fig. 10 Examples of misclassifications in the JAFFE database: (a) "neutral" classified as "happy"; (b) "angry" classified as "sad"; (c) "sad" classified as "angry"; (d) "happy" classified as "angry"

Table 4 Detection results of the facial expression recognition system for the JAFFE database

Data Set / Results	Testing Results
Number of Images	90
Correctly Identified	78
Detection Rate (%)	86.7%

Table 5 Misclassifications in the JAFFE database (90 images)

Actual \ Result	Neutral	Angry	Sad	Happy	Error Rate
Neutral	0	1	0	1	2.2%
Angry	0	0	1	0	1.1%
Sad	0	5	0	0	5.5%
Happy	0	4	0	0	4.4%

Unit: number of images

IV. DISCUSSION & CONCLUSION

In this paper, the objective was to develop a facial expression recognition system using facial feature extraction and an artificial neural network. The major tasks included facial expression database collection, automatic facial feature extraction, and facial expression recognition. Our results demonstrated that our system has achieved reasonable detections of approximately 88% in our database and 87% in the JAFFE database. Because the nature of the data is a major factor for the performance of the ANN, a relatively large amount of data was used for the training of the network. In addition, we have also attempted a relatively high-resolution images (640×480 pixels) in extracting the facial features. These high-resolution images allow us to distinguish various facial expressions with preliminary success even with small position changes of facial features. Although the digital images in the JAFFE database are in worse quality, our preliminary results however achieve a comparable performance in the JAFFE databases.

In analyzing the misclassifications in our database, the two specific facial expressions (smile and sad) were found to exhibit larger errors. In comparison, the two specific facial expressions (sad and happy) were found to exhibit larger errors in the JAFFE database. Such classification errors are however not surprising, partly because the position changes of facial features associated were relatively small.

In our study, all the facial features were automatically extracted in conjunction with the automatic ANN facial

expression classification, resulting in a fully-automatic facial expression recognition system once the system was trained in advance. While our system was shown to yield reasonable detections, these results were largely based on the relatively accurate position identifications of facial features that were automatically extracted using image processing techniques.

In recent studies, facial expression classification systems (e.g., FADECS system) have been developed [15, 16]. Among which, promising systems have been proposed. While the systems are also fully automatic, experimental results were however lacking. Xue and Youwei [22] however reported detection rates of approximately 60% using the Principal Component Analysis (PCA) and the difference of statistical features (DSF) in the JAFFE database. In comparison, we have proposed a system using real data (including the JAFFE database) for system evaluation. Our results clearly indicated the feasibility of the methodology being used.

In this study, we have presented a system for the classification of facial expressions. However, facial expressions remain a complex emotion and can vary from person to person which can not be well characterized with mathematical models. The ill-posed nature of the problem generally leads to system misclassifications. Our system is currently designed using static images and can be potentially generalized if dynamic or time-dependent images are available. We anticipate system improvement if such time-dependent information could be acquired and analyzed.

Despite that the processing time associated with the training of the system was rather significant (~5.13 minutes per image on a P4 2.6GHz personal computer), the processing for each face image during the testing was fast (~0.01 seconds per image). This created a potential for the system to be incorporated in a real-time facial expression system. A typical application may include an interactive real-time chatting room on the internet, when personal within the internet chatting room would not reveal his or her identification. Other use will be to incorporate the system in medical care by monitoring patients in a video surveillance system. The system could be used to distinguish if the patient is in pain, despite that there are still some issues for patient care to be resolved. Future investigations will warrant in this regard.

ACKNOWLEDGMENTS

The authors thank all the personals who participated in this research. The authors also thank Professor H. Y. Liao for his support in this research. This work was supported in part by the National Science Council (NSC), Taiwan, R.O.C., under contract NSC95-2221-E-033-042 and the Institute of Information Science, Academia Sinica, Taiwan, R.O.C.

REFERENCES

[1] C. Darwin, *The Expression of the Emotions in Man and Animals.* Chicago: Univ. Chicago Press, Phoenix Books, 1965.

[2] P. Ekman and W. V. Friesen, *Unmasking the face: A guide to recognizing emotions from facial clues.* Englewood Cliffs, New Jersey: Prentice-Hall, 1975.

[3] J. Preece, Y. Rogers and H. Sharp, *Interaction Design Beyond Human-Computer Interaction*. NJ: John Wiley & Sons Inc., 2002.

[4] J. Preece, *Human-Computer Interaction*. Addison Wesley, England., 1994.

[5] I. A. Essa and A. P. Pentland, "Coding, analysis, interpretation, and recognition of facial expressions," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 19, no. 7, pp. 757-763, 1997.

[6] R. Brunelliand and T. Poggio, "Face recognition: features versus templates," *IEEE Transc. on Pattern Analysis and Machine Intelligence*, vol. 15, no. 10, pp. 1042-1052, 1993.

[7] M. Pantie and L. J. M. Rothkrantz, "Automatic analysis of facial expressions: the state of the art," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 22, no. 12, pp. 1424-1445, 2000.

[8] M. Bartlett, J. Hager, P. Ekman and T. Sejnowski, "Measuring facial expressions by computer image analysis," *Psychophysiology*, vol. 36, pp. 253-264, 1999.

[9] M. J. Black and Y. Yacoob, "Recognizing facial expressions in image sequences using local parameterized models of image motion," I*nternational Journal of Computer Vision*, vol. 25, no. 1, pp. 23-48, 1997.

[10] K. Arun and S. Ranganath, "Face recognition using radial basis function networks," *Proc. Second Asian Conf. on Computer Vision*, Singapore, vol. 2, pp. 456-460, 1995.

[11] S. Lawrencs, C. L. Giles, A. C. Tsoi and A. D. Back, "Face recognition: a convolutional neural network approach," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 20, no. 1, pp. 673-686, 1998.

[12] S. H. Lin, S. Y. Kung and L. J. Lin, "Face recognition/detection by probabilistic decision-based neural network," *IEEE Transactions on Neural Network*, vol. 8, no. 1, pp. 114-132, 1997.

[13] J. Xhang, Y. Yan and M. Lades, "Face recognition: eigenface, elastic matching, and neural nets," *Proceedings of the IEEE*, vol. 85, no. 9, 1997.

[14] M. Pantic and L. J. M. Rothkrantz, "Facial action recognition for facial expression analysis from static face images," *IEEE Trans. Systems, Man and Cybernetics,* vol. 34, no. 3, pp. 1449-1461, 2004.

[15] Z. Gengtao, Z. Yongzhao and Z. Jianming, "Facial expression recognition based on selective feature extraction," 6th *International Conference on Intelligent Systems Design and Applications*, vol. 2, pp. 412-417, 2006.

[16] I. -O. Stathopoulou and G. A. Tsihrintzis, "Detection and expression classification system for face images (FADECS)," *IEEE Workshop on Signal Processing Systems*, Athens, Greece, Nov. 2-4, 2005.

[17] 楊靜杰，「運用網路攝影機進行自動化人臉特徵擷」，碩士論文，中原大學資訊工程學系，2007.

[18] J. M. Zurada, *Introduction to Artificial Neural Systems*, PWS Publishing Company, Boston, 1995.

[19] H. D. E. Rumelhart, G. E. Hinton, R. J. Williams, et al. "Learning representations by back-propagation errors," *Nature*, vol. 323, pp. 533-536, 1986.

[20] I. -O. Stathopoulou and G. A. Tsihrintzis, "An improved neural-network-based face detection and facial expression classification system," *IEEE International Conference on Systems, Man, and Cybernetics*, The Hague, The Netherlands, Oct. 10-13, 2004.

[21] The Japanese Female Facial Expression (JAFFE) Database. Available: http://kasrl.org/jaffe.html.

[22] G. Xue and Z. Youwei, "Facial expression recognition based on the difference of statistical features," *Proceedings* of *the 8th International Conference on Signal Processing,* ICSP, Guilin, China, Nov. 16-20, 2006.