Proceedings of the 2009 IEEE International Conference on Systems, Man, and Cybernetics
San Antonio, TX, USA - October 2009

Interactive Emotion Recognition Using Support Vector Machine for Human-Robot Interaction

Ching-Chih Tsai*, You-Zhu Chen, Ching-Wen Liao

[1]Department of Electrical Engineering, National Chung Hsing University
250, Kuo-Kuang Road, Taichung 40227, Taiwan, R.O.C.
*E-mail: cctsai@dragon.nchu.edu.tw

Abstract: This paper presents an interactive emotion recognition system using support vector machine for human-robot interaction. The proposed emotion recognition algorithm is composed of Harr wavelet transform, principal component analysis (PCA) method, and support vector machine (SVM). This algorithm is shown effective and useful in achieving both face identification and facial expression recognition. The performance and merit of the proposed methods are exemplified by conducting several experiments on face identification, emotion recognition and interactive scenarios.

Keywords: emotion, facial expression, human-robot interaction, Harr wavelet transform, principal component analysis (PCA), recognition, support vector machine (SVM).

1. INTRODUCTION

Much research for face detection, face tracking, face identification and interactive facial expressions has been widely studied over the past decades in the field of computer vision and mobile service robots. Such service robots have already gained widespread applications, such as home-caring services, medical robots, military tasks, entertainment, manufacturing servicing, security, and so on, and they are now becoming a part of our life. In addition, there has been a great amount of theoretical and practical research developed for vision-based robot applications in the recent years. For example, home-service robots have been extensively investigated not only for task execution at home, but also for interacting with people [1]

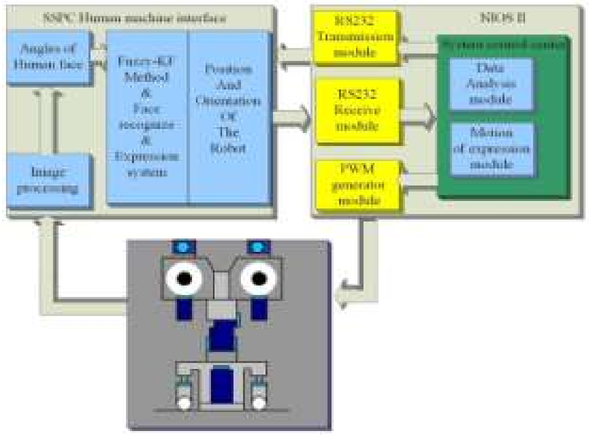
In recent decades, many researchers have paid significant efforts on face tracking and facial expression recognition based on the skin color method, because the skin-tone color is one of obvious features of human faces. Skin detection can be considered as the process of selecting which pixels of a given image correspond to human skin. Skin detection has been shown particularly useful in, for example, face detection and face tracking for security and video indexing applications, model-based video coding and etc. There are many color spaces to represent color images; especially, YCbCr has been widely employed since the skin pixels form a compact cluster in the CbCr plane [2]. In face recognition, Pentland *et al.* [3] proposed PCA, and then used low dimensional characteristics to do face recognition. Jeng *et al.* [4] used the geometry model to find out the position of an middleman's face in a complicated background. However, in the above-mentioned research, all techniques were based on the gray-level imaging processing approaches. Moreover, Sobottka and Pitas [5] and Chen *et al.* [6] offered methods to separate the complexion area to have and search for human faces in the colored images. Aide from that, there have been many methods proposed in recent years; they are fuzzy logics, SVM (support vector machine) [7-9], and ANNs (artificial neural networks) [10]. SVM and ANNs have been successfully applied to face detection, face tracking and even face identification. Since PCA is a well idea to find the best sub-space to represent the training data, and SVM (support vector machine) is a classifier derived from statistical learning theory [11-13], this paper will combine these two techniques for developing new algorithms of face identification and facial expressions recognition. Both algorithms will be shown to work well under different lighting conditions through an equalization operation.

The rest of the paper is outlined as follows. The interactive emotion recognition system is constructed in Section II. Section III describes using the Harr wavelet transform to decrease the image dimension, and Section IV introduces PCA method. The eigenspace based method is adopted to find some face candidates in lower resolution, and then verify these candidates in higher resolution with eigen components and some knowledge-based rules about human faces. In Section V the basic theory of SVM and the Euclidean distance are used for face identification and expression recognition. Experimental results are presented in Section VI. The conclusions are stated in Section VII.

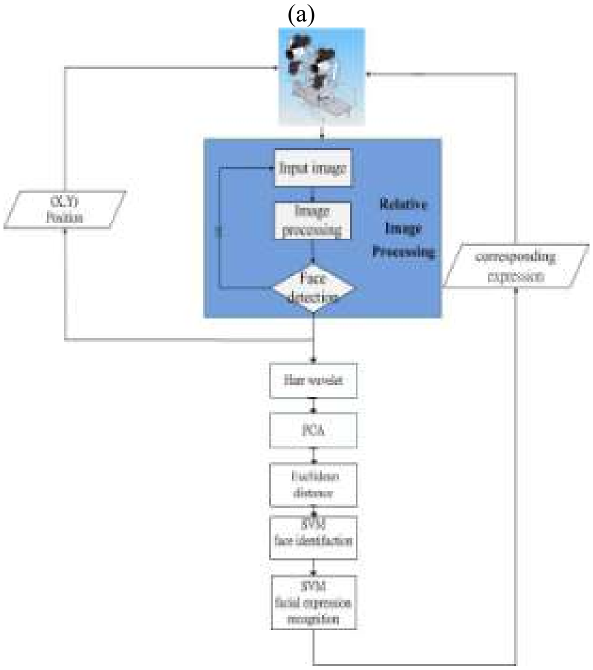
2. INTERACTIVE EMOTION RECOGNITION SYSTEM DESIGN

Fig. 1(a) shows system structure of the experimental interactive emotion recognition system. This system is constructed using a five-degrees-of-freedom robotic head and a FPGA-based motion controller and a small-scale personal computer used for performing main techniques, which are face detection, face tracking, face identification and emotion recognition. From hardware architecture point of view, this type of system is composed of three parts: face detection and identification by the presented image

978-1-4244-2794-9/09/$25.00 ©2009 IEEE



(a)



(b)

Fig. 1. The interactive emotion recognition system. (a) System Structure. (b) emotion recognition procedure.

processing algorithm and the fuzzy EKF-based face tracking method [13], and user emotion recognition that is responsible for recognizing interactive human emotions and then giving simple talks to the user. When the face position (x, y) is found using (1), the FPGA-based motion controller will compute PWM control signals, thus controlling the five servomotors via RS232.

$$\begin{cases} \theta_X(F) = a\tan\left(\dfrac{X\_center(F) - 160}{320}\right) & \text{(degree)} \\ \theta_Y(F) = a\tan\left(\dfrac{Y\_center(F) - 120}{240}\right) & \text{(degree)} \end{cases} \quad (1)$$

Fig. 1(b) depicts the flow chart of the interactive emotion recognition procedure. Once the interactive is activated, the system will start the welcome mode in which the robot will talk to the user. Since then, the system proceeds with algorithms of face identification, tracking and emotion recognition. If the system recognizes the user's emotion expressions, such as natural, happy, anger, sad and joy, and then the robotic system will reply the user an emotion expression and a talk for interactive dialog purpose.

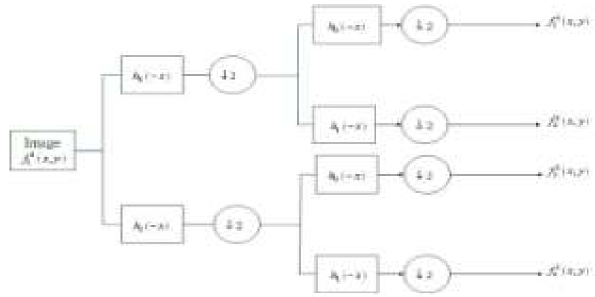
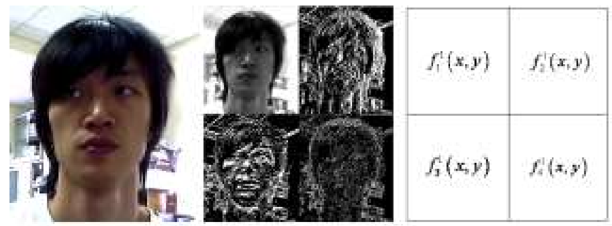


Fig. 2. Diagram of two-dimensional Harr wavelet transforms.



(a) (b) (c)

Fig. 3. The results of the Harr wavelet transform; (a) the original image; (b) the result of the Harr wavelet transform; (c) the position relationship picture of sub-image.

3. HARR WAVELET TRANSFORM

This section aims to reduce image dimension for purposes of face identification and emotion recognition. This is because the training phase for recognition of face and facial expression is often time consuming due to large dimension of the images. Given an original image, the Harr wavelet transform method separates high frequency and low frequency bands of the image by high-pass and low-pass filters from the horizontal direction, and so does the vertical direction of the image. In Fig. 2, the functions $h_0(-x)$ and $h_1(-x)$ denote respectively the high pass and low pass filters, the notation $\downarrow 2$ represents the sampling time and $f_1^0(x, y)$ stands for the original image. After the Harr wavelet transform, the original image is decomposed into four parts, $f_1^1(x, y)$, $f_2^1(x, y)$, $f_3^1(x, y)$ and $f_4^1(x, y)$. The low-frequency part $f_1^1(x, y)$ is the most important part for further training because it is able to reduce the image size without loss of crucial information for face identification. Fig. 3 shows the transformed results in the grey level.

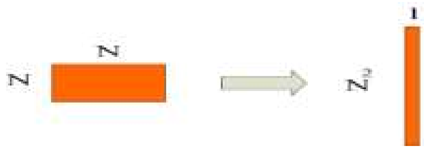


Fig. 4. Transformation of an $N \times N$ image to a $N \times 1$ vector.

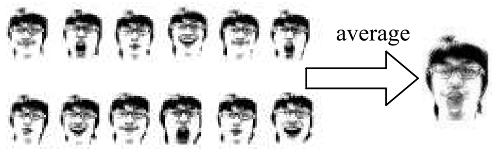


Fig. 5. Training Face images and average face ψ .

4. PRINCIPAL COMPONENT ANALYSIS

PCA called Karhunen-Loeve Transform (KLT) or Hotelling Transform, is a classical technique for multivariate analysis. The method can be traced back to Pearson [14] in 1901 and developed by Hotelling [15]. Its detailed instructions have been given by Jolloffe [16]. PCA can be employed to find out the most important abstract features of faces collected in their database. The main concept behind the method is to calculate the eigenvectors of matrix A which represents a collection of all images of interest, and then to use these eigenvectors to span the column space of matrix A. Hence, if matrix A is constructed from all the faces in the database, then a set of eigenvectors can be calculated in order to span the face space sufficiently and efficiently. Suppose that there are M different human face images and each image has N by N pixels. Fig. 4 illustrates the process of transforming the $N \times N$ image to a $N^2 \times 1$ vector. Accordingly, M different human faces can be represented by m $N^2 \times 1$ vectors.

The working algorithm of PCA method is explained as follows. Consider the training set of M face images $x=\{x_1,x_2,...,x_m\}$ of N-dimensional vectors. Thus, the average face of the training set is defined by $\psi=\frac{1}{m}\sum_{i=1}^{m}x_i$. Since every face differs from the average face ψ, the error vector is defined by $\Phi_i = x_i-\psi$, i = *1,..,m.* Fig. 5 respectively depicts the set of training images and their average face ψ .

After obtaining matrix A, the covariance matrix, C, of the training data is given by (2)

$$C=\sum_{i=1}^{m}\Phi_i\Phi_i^T = AA^T \tag{2}$$

and

$$A=\left[\Phi_1 \; \Phi_2 \; ... \; \Phi_m\right] \tag{3}$$

The eigenvector of the covariance matrix C is found by

$$Cu_k = \lambda_k u_k \tag{4}$$

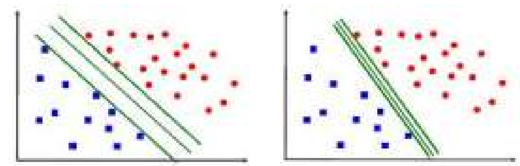


Fig. 6. The chart of the Support Vector Machines.

With the eigenvector u_k , the face image vector w_k , called eigenface, can be formed via (5) and then projected onto the face space F which is spanned by the eigenfaces. The face space F in (5) can further be employed to obtain the abstract features of all the training samples.

$$w_i = u_k\left(x_i-\psi\right) \quad i=1,...,m \tag{5}$$

$$F=\left[w_1 \; w_2 \; ... \; w_{\mathrm{m}}\right] \tag{6}$$

5. RECOGNITION METHODS

5.1 Linear Support Vector Machines

For the two-class classification problem, the goal is to separate the two classes by a function which is induced from available examples. In Fig. 6 there are many possible linear classifiers that can separate the data, but there is only one that maximizes the margin.

Consider the problem of separating the set of training vectors falling into two separate classes $(x_1,y_1),...,(x_l,y_l)$, where $x_i\in R_n, y_i\in\{-1,+1\}$ with a canonical hyperplane [16] $wx+b=0$. The hyperplane has the constraint for parameters $w \; and \; b$. i.e., a separating hyperplane in canonical form must satisfy the following constraints,

$$y_i(w\cdot x_i+b)\geq 1, \; \forall x_i \tag{7}$$

The distance of a point x from the hyperplane is given by

$$d\left(w,b;x\right)=\left(\left|b-\left(b-1\right)\right|\right)/\left\|w\right\|=2/\left\|w\right\| \tag{8}$$

The margin is $2/\left\|w\right\|$ according to its definition. Hence the hyperplane that optimally separates the data is the one that minimizes

$$\Phi\left(w\right)=\frac{1}{2}\left\|w\right\|^2 \tag{9}$$

The solution to the optimization problem of (8) under the constraints of (6) is given by the saddle point of the Lagrange functional,

$$L\left(w,b,a\right)=\frac{1}{2}\left\|w\right\|^2-\sum_{i=1}^{l}\alpha_i\left\{y_i\left[\left(w\cdot x_i\right)+b\right]-1\right\} \tag{10}$$

where α_i are the Lagrange multipliers. The Lagrangian must be minimized with respect to w,b and maximized with respect to $\alpha_i\geq 0$. Classical Lagrangian duality enables the primal problem (9) to be transformed to its dual problem, which is easier to solve. The solution to the dual problem is described by,

$$W\left(\bar{\alpha}\right)=\sum_{i=1}^{l}\alpha_i-\frac{1}{2}\sum_{i=1}^{l}\sum_{j=1}^{l}\alpha_i\alpha_j y_i y_j x_i\cdot x_j \tag{11}$$

The minimum with respect to w and b of the Lagrangian, w *and* b, is given by

$$\begin{cases} \frac{\partial L}{\partial b} = 0 \Rightarrow \sum_{i=1}^{l} \alpha_i y_i \\ \frac{\partial L}{\partial w} = 0 \Rightarrow w = \sum_{i=1}^{l} \alpha_i y_i x_i \end{cases} \tag{12}$$

the constraints is given by,

$$\sum_{i=1}^{l} \alpha_i y_i = 0 \text{ and } \alpha_i \geq 0,\ i = 1,...,l \tag{13}$$

Solving (10) with the constraints (13) determines the Lagrange multipliers; afterwards, the optimal separating hyperplane is given by,

$$w = \sum_{i=1}^{l} \alpha_i y_i x_i \tag{14}$$

For a new data point x, the classification is then done by

$$f(x) = \operatorname{sgn}\left(\sum_{i=1}^{l} y_i \alpha_i \cdot (x \cdot x_i) + b \right) \tag{15}$$

5.2 Liner Support Vector Machines: Non-Separate Case

To generalize the optimal separating hyperplane to the non-separable case shown in Fig. 7, it is necessary to introduce slack variables ξ_i. In doing so, the constraints in (7) are modified by

$$y_i (w \cdot x_i + b) \geq 1 - \xi_i,\ \xi_i \geq 0 \quad \forall x_i \tag{16}$$

The generalized optimal separating hyperplane is determined by minimizing,

$$\Phi(w, \xi) = \frac{1}{2} \|w\|^2 + C \sum_{i=1}^{l} \xi_i \tag{17}$$

subject to the constraints of (16). This optimization problem can also be transformed to its dual problem, and the solution is given by

$$W(\bar{\alpha}) = \sum_{i=1}^{l} \alpha_i - \frac{1}{2} \sum_{i=1}^{l} \sum_{j=1}^{l} \alpha_i \alpha_j y_i y_j x_i \cdot x_j \tag{18}$$

with the constraints,

$$\begin{cases} 0 \leq \alpha i \leq C \quad, \forall i \\ \sum_{i=1}^{l} \alpha_i y_i = 0 \end{cases} \tag{19}$$

The solution to this minimization problem is identical to the separable case except for a modification of the bounds of the Lagrange multipliers.

5.3 Nonlinear Support Vector Machine

Another merit of SVM is to map the input vector into a higher dimensional feature and thus can solve for the nonlinear case. By choosing a nonlinear mapping function $\varphi(x) \in R^M$, where $M > N$, the SVM can construct an optimal hyperplane in this new feature space. $k(x, x_i)$ is the inner-product kernel performing the

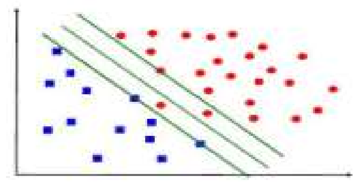


Fig. 7. Non-separate case for linear support vector machines.

nonlinear mapping into feature space. $k(x, x_i) = k(x_i, x) = \varphi(x)^T \varphi(x_i)$. Hence, the dual optimization problem becomes

$$MAX\ \mathrm{W}(\alpha) = \sum_{i=1}^{l} \alpha_i - \frac{1}{2} \sum_{i=1}^{l} \sum_{j=1}^{l} \alpha_i \alpha_j y_i y_j k(x_i \cdot x_j) \tag{20}$$

subject to the same constraints as in Eq. (19). The only requirement on the kernel $k(x, x_i)$ is to satisfy the Mercer's theorem [6]. Using Kernel functions, without treating the high dimensional data explicitly, unseen data can be classified as follows

$$x \in \begin{cases} \textit{positive class}, \textit{if } f(x) > 0 \\ \textit{negative class}, \textit{if } f(x) < 0 \end{cases} \tag{21}$$

where the decision function is

$$f(x) = \operatorname{sgn}\left(\sum_{i=1}^{l} y_i \alpha_i \cdot k(x \cdot x_i) \right) \tag{22}$$

where the kernel function used in the thesis is given by

$$k(x, x_i) = e\left(-\frac{1}{2\sigma^2} \|x - x_i\|^2 \right) \tag{23}$$

5.4 SVM: Multi-Class Case

Since the previous subsection had mentioned the basic theory of SVM for classification, a multi-class pattern recognition system can be obtained by combining two classes SVMs. The one-against-one strategy [10] will be used to classify between each pair. One-against-one will generate a support vector machine for two classes. Denote the number of classes as k, the SVM learns $k(k+1)/2$ discrimination functions in the training stage, and carry out comparisons of $k+1$ times under the fixed binary tree structure.

5.5 Euclidean distance method

The most common recognition rule is the use of the Euclidean distance. Euclidean distance calculates the shortest Euclidean distance from the test samples and training samples. Let y be the feature parameter of the test image, y_i^k be the parameter i of class k in training database. Every class has m training samples and the Euclidean distance is computed by $d(y, y_i^k) = \|y - y_i^k\|$. Therefore, the shortest Euclidean distance is calculated by (24).

Fig. 8. Experimental pictures of face identification.

$$ID_y = \min\left(d\left(y, y_1^k\right), d\left(y, y_2^k\right) ... d\left(y, y_n^{nk}\right)\right) \quad (24)$$

The Euclidean distance is a direct method that can be very simple and fast to reduce the processing time in implementation, but its recognition error may become big.

6. EXPERIMENTAL RESULTS AND DISCUSSION

This section is dedicated to conduct both experiments to examine the efficacy of the proposed face identification and facial expressions recognition. In both experiments, twelve people were invited to join the experiments. For speeding up the experimental procedure, the well-known LIBSVM [11] was adopted to do both experiments. The procedure of the experiments is shown described in following steps:

Step1:Use the Harr wavelet transform to reduce the training samples dimension.

Step2:Find out the face feature's data of the training samples by PCA.

Step3:Use the Euclidean distance to filter out unnecessary training samples.

Step4:Classify the face feature's data to recognize the human face and facial expressions by the support vector machine.

Step5:Shows the interactive expressions like happy smiling, anger, sadness and joy base on information of facial expression recognition.

6.1 Face Identification Experiments and Discussion

This subsection will examine the effectiveness of the proposed face identification method by performing two experiments as shown in Fig. 8. Both experiments respectively took ten and twelve images for each attendee. Table 1 depicts the training data information for the face identification experiment and Tables 2 and 3 presents the experiment results. Clearly, the result in Table 3 has a higher recognition rate than that in Table 2. The result indicates that the more the training samples, the higher the recognition rate; however, this was done at the cost of taking more processing time. To circumvent the difficulty, the Euclidean distance method is used to filter out unnecessary training samples so as to reduce or shorten the processing time. In Tables 2 and 3, the combination of the Euclidean distance method and the SVM always gave the best recognition rate and the shortest processing time.

Table 1. Data used for the face identification experiment

Experiment 1	Training samples	120
	Test samples	120
Experiment 2	Training samples	240
	Test samples	120

Table 2. Experimental results of the proposed face identification experiment 1

Method	Item	Experiment rate
Euclidean distance	Recognition Rate	89%
	Processing time	7 p/sec
SVM	Recognition Rate	90%
	Processing time	5.73 p/sec
Euclidean distance and SVM	Recognition Rate	94%
	Processing time	5.9 p/sec

Table3. Experimental results of the proposed face identification experiment 2

Method	Item	Experiment Results
Euclidean distance	Recognition Rate	90%
	Processing time	6.9 p/sec
Nonlinear SVM	Recognition Rate	92%
	Processing time	5.5 p/sec
Euclidean distance and nonlinear SVM	Recognition Rate	95%
	Processing time	5.8 p/sec

Table 4. Data information of the facial expressions recognition experiment

Experiment 1	Training samples	120
	Test samples	120
Experiment 2	Training samples	240
	Test samples	120

Table 5. Experimental results of the proposed facial expressions recognition experiment 1

Method	Item	Experiment rate
Euclidean distance	Recognition Rate	73%
	Processing time	7.2 p/sec
Nonlinear SVM	Recognition Rate	84%
	Processing time	6.9 p/sec
Euclidean distance and nonlinear SVM	Recognition Rate	78%
	Processing time	7.1 pictures/sec

Table 6. Experimental results of the proposed facial expressions recognition experiment 2

Method	Item	Experiment rate
Euclidean distance	Recognition Rate	77%
	Processing time	7 p/sec
Nonlinear SVM	Recognition Rate	89%
	Processing time	6.5 p/sec
Euclidean distance and nonlinear SVM	Recognition Rate	82%
	Processing time	6.6 pictures/sec

Fig. 9. Experimental results of facial expression recognition; smiling; anger; sadness; happy smiling.

6.2 Emotion Recognition Experiments and Discussion

Similar to the previous face identification experiments as shown in Fig. 9, 12 people were again invited to attend the following two experiments. Both experiments respectively acquired ten and twenty pictures of each attendee for training. Table 4 shows the number of images used for each facial expression recognition experiment. Tables 5 and 6 depict the experiment results. The results clearly reveal that the recognition rate for facial expressions is not as high as that for face identification. This is because the face features, such as eyes, mouth, hair and etc., of each attendee are different, thereby reducing the recognition rate. In other words, the similarity of the training samples and the testing samples affect the result drastically. The recognition rate could be made higher if more training samples are used.

6.3 Facial Expression Recognition Experiments and Discussion

Similar to the previous face identification experiments, 12 people were again invited to attend the following two experiments. Both experiments respectively acquired ten and twenty pictures of each attendee for training. Table 4 shows the number of images used for each facial expression recognition experiment. Tables 5 and 6 depict the experiment results. The results clearly reveal that the recognition rate for facial expressions is not as high as that for face identification. This is because the face features, such as eyes, mouth, hair and etc., of each attendee are different, thereby reducing the recognition rate. In other words, the similarity of the training samples and the testing samples affect the result drastically. The recognition rate could be made higher if more training samples are used.

7. CONCLUSIONS

This paper has developed techniques for face identification, facial expressions recognition and interactive facial expression system design of the intelligent robotic head system. Face identification and facial expressions recognition have been accomplished using PCA and SVM. These two algorithms together with interactive expressions to user's emotions have been integrated for constructing the interactive facial expression system. From the experimental results, the recognition rates of the Euclidean distance method, the support vector machine and the combination of the Euclidean distance and the support vector machine are almost very close in the face identification experiment. However, the recognition rates of the three methods in facial expression recognition experiments are lower than those for the face identification experiment, but the SVM method still has a superior recognition rate than the Euclidean distance method.

ACKNOWLEDGEMENTS

The authors deeply acknowledge final support in part from National Science Council, Taiwan, ROC, under contract 96-2221-E-005-106-MY2, and in part from the ministry of education, Taiwan, ROC, under ATU plan.

REFERENCES

[1] C. C. Tsai, Y.. Z. Chen, Y.-Y. Li, "*Face Tracking Using Fuzzy Kalman Filter*," Proceedings of the 2008 National Symposium on Fuzzy theory and Applications, 2008.

[2] A.Albiol, L. Torres, E.J. Delp, "*Optimum color spaces for skin detection*," Politechnic University of Valencia Spain, vol.1, , pp.122-124, 2001.

[3] M. Turk and A. Pentland, "*Eigenfaces for recognition*," Journal of Cognitive Neuro-science, vol.3, no.1, pp.71-86, 1991.

[4] S. H. Jeng, H. Y. Mark Liao, C. C. Han, M. Y. Chern, and Y. T. Liu, "*An efficient approach for facial feature detection using geometrical face model*," in Proc. of the 13th International Conference on Pattern Recognition, Vienna, Austria, 1996, vol.3, pp.426-430.

[5] K. Sobottka and I. Pitas, "*Extraction of facial regions and features using color and shape information*," in Proc. 13th International Conference on Pattern Recognition, , Vienna, Austria, Aug. 1996, pp.421-425.

[6] H. Wu, Q. Chen, and M. Yachida, "*A fuzzy-theory-based face detector*," in Proc. 13th International Conference on Pattern Recognition, Vienna, Austria, Aug. 1996 pp.406-410.

[7] C. Cortes, V. Vapnik, "*Support vector networks*," Machine Learning, vol. 20, pp. 273-297, 1995.

[8] E. Osuna, Support Vector Machines: Training and Applications, Ph.D. thesis, Dept. of EECS, Massachusetts Institute of Technology, 1998.

[9] E. Osuna, R. Freund, and F. Girosi, "*Training support vector machines: an application to face detection,*" in Proc. Computer Vision and Pattern Recognition, 1997, pp.17-19.

[10] S. Lawrence, C. L. Giles, A. C. Tsoi, and A. D. Back. "*Face recognition: A convolutional neural network approach*," in Proc. IEEE Trans. Neural Networks, vol.8, pp.98–113, 1997.

[11] U. KreBel, "*Pairwise classification and support vector machines,*" in proc Advances in Kernel Methods — Support Vector Learning, pp.255-268, Cambridge, MA, 1999, MIT Press.

[12] C.-C. Chang and C.-J. Lin. LIBSVM: a library for support vector machines, 2001.

[13] J. C. Burges, "*A tutorial on support vector machines for pattern recognition*," Data Mining and Knowledge Discovery, Vol. 2, pp. 121-167,1998.

[14]K. Pearson, "*On line and planes of closest fit to systems of points in space*," Philosophy Magazine, vol.2, pp.559 - 572, 1901.

[15]H. Hotelling, "*Of a complex of statical variables into principal components*," Journal of Educational Psychology, vol.24, pp.417 - 441, 1933.

[16] I. T. Jolliffe, *Principal Component Analysis*, Springer-Verlag, 1986.

[17] V. N. Vapnik. *Statistical learning theory*. John Wiley & Sons, New York, 1998.