A New Facial Expression Recognition Method Based on Local Gabor Filter Bank and PCA plus LDA *

Hong-Bo Deng[1], Lian-Wen Jin[1], Li-Xin Zhen[2], Jian-Cheng Huang[2]

[1] School of Electronic and Information Engineering, South China University of Technology, Guangzhou, 510640, P.R.China
[2] Motorola China Research Center, Shanghai, 210000, P.R.China

{hbdeng, eelwjin}@scut.edu.cn
{Li-Xin.Zhen, Jian-Cheng.Huang}@motorola.com

Abstract

This paper proposes a facial expression recognition system based on Gabor feature using a novel local Gabor filter bank. Traditionally, a global Gabor filter bank with 5 frequencies and 8 orientations is often used to extract the Gabor feature. A lot of time will be involved to extract feature and the dimensions of such Gabor feature vector are prohibitively high. A novel local Gabor filter bank with part of frequency and orientation parameters is proposed. In order to evaluate the performance of the local Gabor filter bank, we first employed a two-stage feature compression method PCA plus LDA to select and compress the Gabor feature, then adopted minimum distance classifier to recognize facial expression. Experimental results show that the method is effective for both dimension reduction and good recognition performance in comparison with traditional entire Gabor filter bank. The best average recognition rate achieves 97.33% for JAFFE facial expression database.

Keyword: Local Gabor filter bank, feature extraction, PCA, LDA, facial expression recognition.

I. Introduction

Facial expressions deliver rich information about human emotion and play an essential role in human communications. In order to facilitate a more intelligent and natural human machine interface of new multimedia products, automatic facial expression recognition [1][18][20] had been studied world wide in the last ten years, which has become a very active research area in computer vision and pattern recognition. There are many approaches have been proposed for facial expression analysis from both static images and image sequences [12][18] in the literature.

In this paper we focus on the recognition of facial expression from single digital images with emphasis on the feature extraction. A number of approaches have been developed for extracting

* Project sponsored by: Motorola Labs Research Foundation (No.303D804372), NSFC (No.60275005), GDNSF (No.2003C50101, 04105938).

features from still images. Turk and Pentland [7] proposed Eigenfaces employed principal component analysis (PCA). PCA [8][9] is an unsupervised learning method, which treats samples of the different classes in the same way. Fisherfaces proposed by Belhumeour and Hespanha [6] is a supervised learning method using the category information associated with each sample to extract the most discriminatory features. It has been shown that Fisherfaces performs well in many applications. Lyons [10] used Gabor wavelet [3][4][10][14] to code facial expressions. Nowadays various researchers reported the model-based methods [12][18][20] for feature extraction, such as active appearance model [19], point distribution model and labeled graphs. But those methods require heavy computation or manually detected feature nodes to construct the model, which can hardly be implemented in real-time automatic facial expression recognition (FER). Donato and Bartlett [1] compared various methods of feature extraction for automatically recognizing facial expression, including PCA, LDA, Gabor wavelet, etc. Best performances were obtained using the Gabor wavelet presentation. But the computation and memory requirement of such Gabor feature are very large and the dimension is very high.

The novelty of our method is to select a local Gabor filter bank, with part of the entire *m*-frequency, *n*-orientation set of Gabor filters, instead of using the entire global filter bank to extract feature. In contrast to the entire global Gabor filter bank, the use of the local Gabor filter bank can effectively decrease the computation and reduce the dimension, even improve the recognition capability in some situations. For further dimensionality reduction and good recognition performance, we adopt a two-phase framework PCA plus LDA for feature compression and selection in our facial expression recognition system.

The remainder of the paper is organized as follows: Section 2 describes the preprocessing procedure to get the pure expression image. Section 3 presents the Gabor feature extraction and our novel local Gabor filter bank. Feature compression based on PCA and LDA is discussed in Section 4. In Section 5, experiments are performed on the JAFFE facial expression database [10] with different experimental conditions. Finally, conclusion is given in Section 6.

II. Preprocessing Procedure

Preprocessing procedure is very important step for facial expression recognition. The ideal output of processing is to obtain pure facial expression images, which have normalized intensity, uniform size and shape. It also should eliminate the effect of illumination and lighting. The preprocessing procedure of our FER system performs the following five steps in converting a TIFF JAFFE image to a normalized pure expression image for feature extraction: 1). detecting facial feature points manually including eyes, nose and mouth; 2). rotating to line up the eye coordinates; 3) locating and cropping the face region using a rectangle according to face model [5] as shown in Fig.1. Suppose

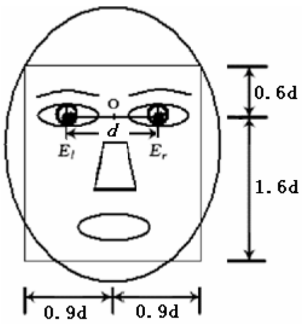

Fig. 1. Facial model

Fig. 2. Example of pure facial expression images after preprocessing from JAFFE database.

the distance between two eyes is *d*, the rectangle will be 2.2d×1.8d; 4). scaling the image to fixed size of 128×96, locating the center position of the two eyes to a fixed position; 5). using a histogram equalization method to eliminate illumination effect. Fig.2 illustrates some examples of pure facial expression images after preprocessing from JAFFE database.

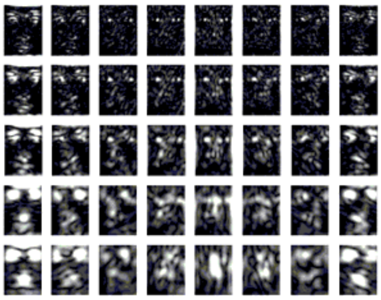

Fig. 4. The magnitudes of the Gabor feature representation of the first face image in Fig.2

III. Gabor Feature Extraction

The Gabor filters, whose kernels are similar to the 2D receptive field profiles of the mammalian cortical simple cells [3][4], have been considered as a very useful tool in computer vision and image analysis due to its optimal localization properties in both spatial analysis and frequency domain [2][14][15][16][17].

A. *Gabor Filters*

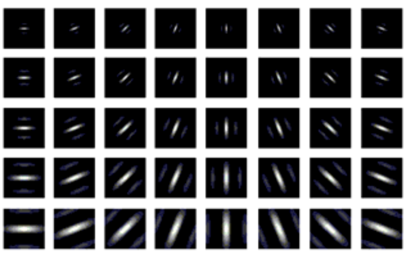

Fig. 3. The real part of the Gabor filters with five frequencies and eight orientations for $\omega_{\max} = \pi/2$, the row corresponds to different frequency ω_m, the column corresponds to different orientation θ_n

In the spatial domain, a Gabor filter is a complex exponential modulated by a Gaussian function [4]. The Gabor filter can be defined as follows,

$$\psi(x, y, \varpi, \theta) = \frac{1}{2\pi\sigma^2} e^{-(\frac{x'^2 + y'^2}{2\sigma^2})} [e^{i\varpi x'} - e^{-\frac{\varpi^2\sigma^2}{2}}]$$
$$x' = x\cos\theta + y\sin\theta, y' = -x\sin\theta + y\cos\theta \tag{1}$$

where (x, y) is the pixel position in the spatial domain, ω the radial center frequency, θ the orientation of Gabor filter, and σ the standard deviation of the round Gaussian function along the *x*- and *y*-axes. In addition, the second term of the Gabor filter, $e^{-\varpi^2\sigma^2/2}$, compensates for the DC value because the cosine component has nonzero mean while the sine component has zero mean. According to [4], we set $\sigma \approx \pi/\omega$ to define the relationship between σ and ω in our experiments.

In most cases a Gabor filter bank with five frequencies and eight orientations [1][2][18] is used to extract the Gabor feature for face representation. Selecting the maximum frequency $\omega_{max}=\pi/2$, $\omega_m=\omega_{(max)}\times\lambda^{-(m-1)}$, m={1,2,3,4,5}, $\lambda=\sqrt{2}$, $\theta_n=(n-1)\pi/8$, n={1,2,…,8}, the real part of the Gabor filters with five frequencies and eight orientations is shown in Fig.3. From Fig.3 it can be seen that the Gabor filters exhibit strong characteristics of spatial locality and orientation selectivity.

B. Gabor Feature Representation

The Gabor feature representation of an image I(x, y) is the convolution of the image with the Gabor filter bank $\psi(x, y, \omega_m, \theta_n)$ as given by:

$$O_{\mathrm{m,n}}(x,y) = I(x,y) * \psi(x,y,\varpi_{\mathrm{m}},\theta_{\mathrm{n}}) \qquad \textbf{(2)}$$

where * denotes the convolution operator. The magnitude of the convolution outputs of a sample image (the first image in Fig.2) corresponding to the filter bank in Fig.3 is shown in

Fig.4. In practice, the time for performing Gabor feature extraction is very long and the dimension of Gabor feature vector is prohibitively large. For example, if the size of normalized

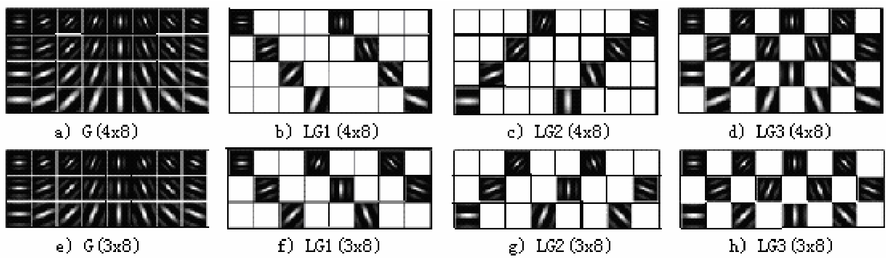

Fig. 5. Examples of several global and local Gabor filter bank, the black blocks are the selected filter for LG(*m*x***n***)

image is 128×96, the dimension of the Gabor feature vector with 40 filters will result in 491520 (128×96×5×8).

C. Local Gabor Filter Bank

It can be seen that the Gabor representations are very similar using the filters with the same orientation, especially using the filters with the two neighboring frequencies, such as the first column in Fig.4. It is found that the Gabor feature vector with all the 40 filters becomes very redundant and correlative. For the global Gabor filter bank with all the *m* frequencies and *n* orientations, we denoted it as G(*m*x*n*). In this paper we proposed a novel local filter bank with part of the entire *m* frequencies and *n* orientations, and we denoted it as LG(*m*x*n*). In order to select few Gabor filters to reduce the dimension and computation without degrading the recognition performance, it should cover all the frequencies and orientations, but only select one frequency for each orientation or increase the interval between the neighboring frequencies with the same orientation. Several global and local Gabor filter banks are shown in Fig.5. The method of selecting the LG1(*m*x*n*) is that the parameter *m* of frequency increases repeatedly from min to max, and the parameter n of orientation adds one for each time. The difference of LG2(mxn) is that the parameter m of frequency decreases from max to min. For LG3(mxn), it is selected with an interval between any two filters.

The computation and memory required by different global and local Gabor filter bank are given in Table1. By comparing the performance of G(mxn) with LG(mxn), LG(mxn) has the advantages of shortening the time for feature extraction, reducing the dimension, decreasing the computation and storage. The recognition performance will be depicted in Section 5. Our experiments were conducted using a Pentium IV 2.8G PC with 512MB memory.

Table 1. Computation and memory required by different Gabor filter bank

Gabor Filter Bank	Number of Filters	Original Dimension (D)	Feature Extraction Time(ms)	Sampling Feature dimension	PCA Matrix
G(5x8)	40	491520	2167	7680	7680×128
G(4x8)	32	393216	1775	6144	6144×128
G(3x8)	24	294912	1357	4608	4608×128
LG1(3x8)	8	98304	435	1536	1536×128
LG3(3x8)	12	147456	681	2304	2304×128

To encompass the properties of spatial locality and orientation selectivity, it should concatenate [2][13] all the outputs of Gabor filter bank and derive the Gabor feature vector. Before the concatenation, we first downsample each output of Gabor filter, and then normalize it to zero mean and unit variance. One simple scheme is to sample the facial feature in a regular grid [13], which sample over the whole face regions with regular interval. The last 2 columns of Table1 show the dimension and PCA matrix of sampling feature when the sampling interval is 8 pixels.

IV. Feature Compression

One approach to coping with the problem of excessive dimensionality is to reduce the dimensionality by linear combining features [9]. In effect, linear methods project the high-dimensional data onto a lower dimensional space, we call it feature compression. There are two classical approaches to finding effective linear transformations, which are Principal Component Analysis (PCA) [1][6][7][8][9] and Linear Discriminant Analysis (LDA) [1][6][8][9]. PCA seeks a projection that best represents the original data in a least-squares sense, and LDA seeks a projection that best separates the data in a least-squares sense.

A. ***PCA***

Let us consider a set of N sample images $\{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_N\}$ represented by t-dimensional Gabor feature vector. The PCA [8][9] can be used to find a linear transformation mapping the original

t-dimensional feature space into an f-dimensional feature subspace, where normally $f << t$. The new feature vector $y_i \in \Re^f$ are defined by

$$y_i = W_{pca}^T x_i \quad (i = 1,2,\ldots,N) \qquad \textbf{(3)}$$

where W_{pca} is the linear transformations matrix, i is the number of sample images.

The columns of W_{pca} are the f eigenvectors associated with the f largest eigenvalues of the scatter matrix S_T, which is defined as

$$S_T = \sum_{i=1}^{N} (x_i - \mu)(x_i - \mu)^T \qquad \textbf{(4)}$$

where $\mu \in \Re^t$ is the mean image of all samples. The disadvantage of PCA is that it may lose important information for discrimination between different classes.

B. ***LDA***

LDA [8][9] is a supervised learning method, which utilizes the category information associated with each sample. The goal of LDA is to maximize the between-class scatter while minimizing the within-class scatter. Mathematically speaking, the within-class scatter matrix S_w and between-class scatter matrix S_b are defined as

$$\begin{aligned} S_w &= \sum_{j=1}^{c} \sum_{i=1}^{N_j} (x_i^j - \mu_j)(x_i^j - \mu_j)^T \\ S_b &= \sum_{j=1}^{c} (\mu_j - \mu)(\mu_j - \mu)^T \end{aligned} \qquad \textbf{(5)}$$

where x_i^j is the ith sample of class j, μ_j is the mean of class j, μ is the mean image of all classes, c is the number of classes, and N_j is the number of samples of class j .

One way to select W_{lda} is to maximize the ratio det$|S_b|$/det$|S_w|$. If S_w is nonsingular matrix then this ratio is maximized, when the transformation matrix W_{lda} consists of g generalized eigenvectors corresponding to the g largest eigenvalues of $S_w^{-1}S_b$ [1][6][8][9]. Note that there are at most c-1 nonzero generalized eigenvalues, and so an upper bound on g is c-1. In this paper, we consider seven kinds of facial expressions, so the dimension of LDA is up to 6.

C. ***PCA+LDA***

In order to guarantee S_w not to become singular, we require at least t+c samples. In practice it is difficult to obtain so many samples when the dimension of feature is very high. To solve this problem, a two-phase framework PCA plus LDA is proposed by [1][6][8][18]. It can be described that PCA maps the original t-dimensional feature x_i to the f-dimensional feature y_i as an intermediate space, and then LDA projects the PCA output to a new g-dimensional feature vectors $\mathbf{z}_i$. More formally, it is given by

$$z_i = W_{lda}^T W_{pca}^T x_i \ \ (i = 1,2,\ldots,\mathrm{N}) \tag{6}$$

To compare the performance of PCA with PCA+LDA, recognition results using PCA feature and PCA+LDA feature respectively will be reported in Section 5.

V. Experiments and Results

The proposed method is evaluated in terms of its recognition performance using the JAFFE female facial expression database [10][11], which includes 213 facial expression images corresponding to 10 persons. Every person posed 3 or 4 examples of each of the seven facial expressions (happiness, sadness, surprise, anger, disgust, fear, neural). Two facial expression images of each expression of each subject were randomly selected as training samples, while the remaining samples were used as test data. We have 138 training images and 75 testing images for each trial. Since the size of the JAFFE database is limited, we perform the trial over 3 times to get the average recognition rate. In our experiments, the nearest neighbor rule is then used to classify the facial expression images. Experiments were performed with the PCA feature and PCA+LDA feature respectively. Two distance measures Euclidean (L2) and cityblock (L1) are employed by the classifier.

A. *Experiment performance of different Gabor Filter Bank*

The first experiment was designed to compare the recognition performance using different Gabor filter bank. Table2 shows the recognition results.

The classification performance shown in Table2 suggested the following conclusions: 1). The recognition rate of PCA feature is from 76.89% to 90.67%, L1 performs better than L2 with PCA feature. When using PCA+LDA feature to for FER, the recognition rate increases several percent and get up to 97.33%. L2 is slightly better than L1 with PCA+LDA feature. 2). In comparison with the performance of different Gabor filter bank, it is found that G(5x8) is the best while G(4x8) and G(3x8) decrease a little along with fewer Gabor filters used. However local Gabor filter banks such as LG1(3x8), LG2(3x8) and LG3(3x8) have almost the same classification performance as global Gabor filter bank G(3x8). LG3(3x8) even outperforms than G(3x8) when using PCA+LDA.

Table 2. Recognition rates using PCA feature and PCA+LDA feature separately corresponding to different Gabor filter bank

Gabor Filter Bank	Classification Methods			
	PCA		PCA+LDA	
	L2	L1	L2	L1

G(5x8)	80.00	89.33	**97.33**	97.33
G(4x8)	79.56	88.44	**96.89**	96.89
G(3x8)	80.00	87.56	**95.56**	95.11
LG1(3x8)	79.11	84.89	95.11	**95.56**
LG2(3x8)	76.89	83.56	**96.00**	95.56
LG3(3x8)	79.56	87.11	**96.44**	95.56
LG3(4x8)	78.22	90.67	**96.44**	96.00

B. Recognition performance against illumination normalization

This experiment was designed to test the effect of illumination normalization which utilized a histogram equalization method. Selecting LG3(3x8) to extract feature, using L1 for PCA feature and L2 for PCA+LDA feature, the recognition results against illumination normalization are given by Table3.

Table 3. Recognition rates with and without illumination normalization

Illumination Normalization	Classification Methods	
	PCA	PCA+LDA
No (without)	83.56	95.11
Yes (with)	87.11	96.44

From Table3 it is obvious to see that illumination normalization is effective for PCA feature to achieve high performance, whose recognition rate is improved 4% or so. As to PCA+LDA feature, illumination normalization also works better, but it is not as clear as PCA feature. We can draw the conclusion that PCA feature is sensitive to various illumination, but PCA+LDA feature may be less sensitive to different illumination.

C. Eliminating the first several principal components

Previous studies in the field of face recognition [6][7][8] reported that discarding the first one to three principal components (PCs) improved performance. In this paper, we discarded the first one to nine PCs to analyze the recognition effect for FER. Selecting G(3x8) and LG3(3x8) to extract feature, using L1 for PCA feature and L2 for PCA+LDA feature, Table4 depicts the results from this experiment.

Table 4. Recognition results by eliminating the first one to nine principal components

Gabor Filter Bank	Input Feature	Number of Eliminated Principal Component							
		0	1	2	3	4	5	7	9

G(3x8)	PCA	87.56	88.44	88.44	**91.11**	89.78	86.67	82.22	84.00
	PCA+LDA	95.56	96.44	96.44	**96.44**	96.44	96.00	94.22	93.78
LG3(3x8)	PCA	87.11	87.11	88.89	**91.11**	89.33	87.56	81.78	84.89
	PCA+LDA	96.44	95.56	**97.33**	95.56	94.67	94.67	94.22	93.33

Table4 demonstrated that eliminating the first one to three PCs will reach to the best performance rates. Best performance for PCA feature, 91.11%, was obtained by eliminating the first three PCs. Best performance for PCA+LDA feature, 97.33%, was obtained by removing the first two PCs. If discarding more than three PCs, in general, the results will become worse.

VI. Conclusions

In this paper, we introduced our FER system based on Gabor feature and PCA+LDA. We proposed a novel local Gabor filter bank for feature extraction. A minimum distance classifier was employed to evaluate the recognition performance in different experiment conditions. The experiments suggest the following conclusions:

1). Local Gabor filter bank outperforms global Gabor filter bank in the aspects of shortening the time for feature extraction, reducing the high dimensional feature, decreasing the required computation and storage, even achieving better performance in some situations.

2). PCA can significantly reduce the dimensionality of the original feature without loss of much information in the sense of representation, but it may lose important information for discrimination between different classes. When using PCA feature to classify, the L1 distance measure performs better than L2. Illumination normalization is effective for PCA feature to achieve high performance.

3). When using PCA+LDA method, the dimensionality drastically reduced to 6 dimensions and the recognition performance is improved several percent compared with PCA. Experiments show that PCA+LDA feature may partly eliminate the sensitivity of illumination.

4). Discarding the first one to three PCs will reach to the best performance rates. If removing more than three PCs, it will commonly worsen the results. The best performance, using LG3(3x8) and PCA+LDA feature by eliminating the first two PCs, was 97.33% in our experiments.

References

[1] G. Donato, M. S. Bartlett, J. C. Hager, “Classifying Facial Actions”, IEEE Trans. Pattern Analysis and Machine Intelligence, Vol. 21, 1999, pp.974-989

[2] C. Liu, H. Wechsler, “Independent Component Analysis of Gabor Features for Face recognition”, IEEE Trans. Neural Networks, Vol. 14, 2003, pp.919-928

[3] J. G. Daugman, “Complete Discrete 2-D Gabor Transforms by Neural Networks for Image Analysis and Compression”, IEEE Trans. Acoustic, speech and signal processing, Vol. 36, 1988, pp.1169-1179

[4] T. S. Lee, “Image Representation Using 2D Gabor Wavelets”, IEEE Trans. Pattern Analysis and Machine Intelligence, Vol. 18, 1996, pp.959-971

[5] F. Y. Shih, C. Chuang, "Automatic extraction of head and face boundaries and facial features", Information Sciences, Vol. 158, 2004, pp.117-130
[6] P. N. Belhumeour, J. P. Hespanha, D. J. Kriegman, "Eigenfaces vs. Fisherfaces: Recognition Using Class Specific Linear Projection", IEEE Trans. Pattern Analysis and Machine Intelligence, Vol. 19, 1997, pp.711-720
[7] M. Turk, A. Pentland, "Eigenfaces for Recognition", Journal Cognitive Neuro-science, Vol. 3, 1991, pp.71-86
[8] A. M. Martinez, A. C. Kak, "PCA versus LDA", IEEE Trans. Pattern Analysis and Machine Intelligence, Vol. 23, 2001, pp.228-233
[9] R. O. Duda, P. E. Hart, D. G. Stork, *Pattern Classification*. Wiley, New York (2001)
[10] M. J. Lyons, S. Akamatsu, M. Kamachi, J. Gyoba, "Coding Facial Expressions with Gabor Wavelets", In: Proceedings of the 3th IEEE International Conference on Automatic Face and Gesture Recognition, Nara, Japan, 1998, pp.200-205
[11] Z. Zhang, M. Lyons, M. Schuster, S. Akamatsu, "Comparison Between Geometry-Based and Gabor-Wavelets-Based Facial Expression Recognition Using Multi-Layer Perceptron", In: Proceedings of the Third IEEE International Conference on Automatic Face and Gesture Recognition, Nara, Japan, 1998, pp.454-459
[12] I. A. Essa, A. P. Pentland, "Coding, Analysis, Interpretation, and Recognition of Facial Expressions", IEEE Trans. Pattern Analysis and Machine Intelligence, Vol. 19, 1997, pp.757-763
[13] D. H. Liu, K. M. Lam, L. S. Shen, "Optimal sampling of Gabor features for face recognition", Pattern Recognition Letters, Vol. 25, 2004, pp.267-276
[14] C. Liu, H. Wechsler, "Gabor Feature Based Classification Using the Enhanced Fisher Linear Discriminant Model for Face Recognition", IEEE Trans. Image Processing, Vol. 11, 2002, pp.467-476
[15] I. R. Fasel, M. S. Bartlett, J. R. Movellan, "A Comparison of Gabor Filter Methods for Automatic Detection of Facial Landmarks", In: Proceedings of Fifth IEEE International Conf. Automatic Face and Gesture Recognition, Washington, USA, 2002, pp.231-235
[16] L. L. Huang, A. Shimizu, H. Kobatake, "Classification-Based Face Detection Using Gabor Filter Features", In: Proceedings of Fifth IEEE International Conference on Automatic Face and Gesture Recognition, Seoul, Korea, 2004, pp.397-402
[17] V. Kyrki, J. K. Kamarainen, H. Kalviainen, "Simple Gabor feature space for invariant object recognition", Pattern Recognition Letters, Vol. 25, 2004, pp.311-318
[18] B. Fasel, J. Luettin, "Automatic facial expression analysis: a survey", Pattern Recognition, Vol. 36, 2003, pp.259-275
[19] T. Cootes, G. Edwards, C. Taylor, "Active appearance models", IEEE Trans. Pattern Analysis and Machine Intelligence, Vol. 23, 2001, pp.681-685
[20] M. Pantic, J. M. Rothkrantz, "Automatic Analysis of Facial Expressions: The State of the Art", IEEE Trans. Pattern Analysis and Machine Intelligence, Vol. 22, 2000, pp.1424-1444.

Hong-Bo Deng received his B.S. degree in Electronical Engineering from South China University of Technology, China, in 2003. He is currently a M.S. student in the School of Electronic and Information Engineering, South China University of Technology, China. His current interests include Face Expression Recognition, Image Processing, Computer Vision, Human-Computer Interface.

Lian-Wen Jin, an IEEE member, received his Ph.D degrees from South China University of Technology, China, in 1996. He is currently a professor at School of Electronic and Information Engineering, South China University of Technology, China. He has research interests in various aspects of Character Recognition, Human-Computer Interface, Image Processing, Neural Network, and Computer Vision.

Dr. Li-Xin Zhen is currently a senior research manager of Motorola Labs China Research Center.

Dr. Jian-Cheng Huang is currently the director of Motorola Labs China Research Center.