Journal of Southeast University (English Edition) Vol. 21, No. 4, pp. 427 – 431 Dec. 2005 ISSN 1003—7985

Feature fusing in face recognition

Yu Weiwei Teng Xiaolong Liu Chongqing

(School of Electronic Information and Electrical Engineering, Shanghai Jiaotong University, Shanghai 200030, China)

Abstract: With the aim of extracting the features of face images in face recognition, a new method of face recognition by fusing global features and local features is presented. The global features are extracted using principal component analysis (PCA). Active appearance model (AAM) locates 58 facial fiducial points, from which 17 points are characterized as local features using the Gabor wavelet transform (GWT). Normalized global match degree (local match degree) can be obtained by global features (local features) of the probe image and each gallery image. After the fusion of normalized global match degree and normalized local match degree, the recognition result is the class that included the gallery image corresponding to the largest fused match degree. The method is evaluated by the recognition rates over two face image databases (AR and SJTU-IPPR). The experimental results show that the method outperforms PCA and elastic bunch graph matching (EBGM). Moreover, it is effective and robust to expression, illumination and pose variation in some degree.

Key words: face recognition; feature fusion; global features; local features

Face recognition technology has been an active research area in the last two decades. It has been expected to play an important role in surveillance, identity verification, man-machine conversation and so on.

Feature extraction is crucial for face recognition. Facial features can be either global or local. Global features are based on the entire face image, while local features are on some fiducial points. Ref. [1] presented that recognition difficulty was statistically linked to the type of feature extraction under different subject covariate factors such as age and gender. Its results show that local and global features supplement each other under some conditions. We also noticed that some recognition results based on global features were correct while those based on local features were wrong, and vice versa. However, few studies have fused the two types of features[2]. So we have presented a method of feature extraction by fusing local and global features.

1 Extraction of Global and Local Features

1.1 Extraction of global features

Many algorithms have been proposed for the extraction of global features[3-5]. In this paper, principal component analysis (PCA) was chosen to extract global facial features because it is basic and useful[3].

Let the random training face images be denoted by n-dimensional vector $\boldsymbol{x}$ formed via lexicographic ordering of array pixels. The averaged face is $\boldsymbol{\mu}_x = E[\boldsymbol{x}]$. We calculate the principal components $\{w_1, w_2, \dots, w_m\}$, $m < n$, corresponding to m largest eigenvalues of $n \times n$ sample covariance matrix:

$$\overline{\boldsymbol{\Sigma}}_x = \boldsymbol{X}\boldsymbol{X}^{\mathrm{T}} \tag{1}$$

with $\boldsymbol{X} = \{\boldsymbol{x}_1 - \boldsymbol{\mu}_x, \dots, \boldsymbol{x}_{\bar{n}} - \boldsymbol{\mu}_x\}$ for $\bar{n}$ training samples. In most cases, the number of training samples is smaller than image dimension. We can find the principal components by realizing the eigensystem of smaller $\bar{n} \times \bar{n}$ matrix $\boldsymbol{X}^{\mathrm{T}}\boldsymbol{X}$. The derived eigenvectors are orthonormal to span the eigenspace. The m-dimensional global features $\boldsymbol{y}$, here $m = 0.6 \times \bar{n}$, is extracted by multiplying the normalized face $\boldsymbol{x} - \boldsymbol{\mu}_x$ using a linear transformation matrix $\boldsymbol{W}_{\mathrm{pca}} = \{w_1, w_2, \dots, w_m\}$, i. e.,

$$\boldsymbol{y} = \boldsymbol{W}_{\mathrm{pca}}^{\mathrm{T}}(\boldsymbol{x} - \boldsymbol{\mu}_x) \tag{2}$$

1.2 Extraction of local features

There also have been some algorithms for extraction of local features[6,7]. In this paper, active appearance models (AAMs)[8] were used to locate the facial fiducial points. According to the result of Ref. [7], the points were characterized with Gabor wavelet transform[9]. The AAM algorithm has proved to be a successful method for face alignment and synthesis[8]. The Gabor wavelet representation captures salient visual properties such as spatial localization, orientation selectivity and spatial frequency characteristics. Refs. [6, 7, 10] applied Gabor wavelets for face recognition.

1.2.1 Active appearance model

In our experiments, 58 fiducial points were located, as shown in Fig. 3(a). An AAM contained a statistical model of the shape and grey-level appearance of the face image that could generalize almost any valid

Received 2005-03-09.

Biographies: Yu Weiwei (1978—), female, graduate; Liu Chongqing (corresponding author), male, professor, liuchqing@263.net.

example. During the training phase we learned the relationship between model parameter displacements and the residual errors induced between a training image and a synthesized model example. To match to an image we measured the current residuals and used the model to predict changes to the current parameters, leading to a better fit. For more details, please refer to Ref. [8]. Two AAM models were separately trained for AR and SJTU-IPPR face databases because of the difference among races.

1.2.2 Gabor wavelet analysis

AAM located 58 facial fiducial points, from which L points were used to extract the local features. To characterize a fiducial point $\boldsymbol{x}$, we convolved the portion of gray image $I(\boldsymbol{x})$ around it,

$$J_j(\boldsymbol{x}) = \int I(\boldsymbol{x}')\psi_j(\boldsymbol{x} - \boldsymbol{x}')\,\mathrm{d}^2\boldsymbol{x}' \tag{3}$$

with a bank of Gabor kernels

$$\psi_j(\boldsymbol{x}) = \frac{\|\boldsymbol{k}_j\|^2}{\sigma^2}\exp\left(-\frac{\|\boldsymbol{k}_j\|^2\|\boldsymbol{x}\|^2}{2\sigma^2}\right)\cdot \left[\exp(\mathrm{i}\,\boldsymbol{k}_j\boldsymbol{x}) - \exp\left(-\frac{\sigma^2}{2}\right)\right] \tag{4}$$

in the shape of plane waves with wave vector $\boldsymbol{k}_j$, restricted by a Gaussian envelope function. $\|\cdot\|$ denotes the norm operators. We employed a discrete set of five different frequencies, index $\nu = 0, 1, \ldots, 4$, and eight orientations, index $\mu = 0, 1, \ldots, 7$.

$$\boldsymbol{k}_j = \begin{Bmatrix} k_{jx} \\ k_{jy} \end{Bmatrix} = \begin{Bmatrix} k_\nu \cos\varphi_\mu \\ k_\nu \sin\varphi_\mu \end{Bmatrix},\ k_\nu = 2^{-\frac{\nu+2}{2}}\pi,\ \varphi_\mu = \mu\frac{\pi}{8} \tag{5}$$

The obtained 40 coefficients are complex numbers. The phase parts are sensitive to translation variation, so the feature vector was obtained considering the magnitude parts only. Applying the Gabor wavelet transform to facial fiducial points, we obtained the face characterization, consisting in a feature vector of $40 \times L$ real coefficients where L is the number of fiducial points to extract local features.

2 Fusion of Global and Local Features

Give a probe image $\boldsymbol{P}$ and a gallery image $\boldsymbol{G}$. Denote the Euclidean distance of global features between $\boldsymbol{P}$ and $\boldsymbol{G}$ as D, and assume that the match degree $m_$global is the inverse of D, i.e., $m\_\text{global} = 1/D$.

For local features, the match degree between $\boldsymbol{P}$ and $\boldsymbol{G}$ is

$$m\_\text{local} = \frac{\boldsymbol{l}_\mathrm{p}^\mathrm{T}\boldsymbol{l}_\mathrm{g}}{\|\boldsymbol{l}_\mathrm{p}\|\ \|\boldsymbol{l}_\mathrm{g}\|} \tag{6}$$

where $\boldsymbol{l}_\mathrm{p}$ is the local feature of probe image, and $\boldsymbol{l}_\mathrm{g}$ is that of gallery image.

2.1 Match degree normalization

There are several normalization methods, for example min-max, z-score, tanh, and adaptive[11]. The result in Ref. [11] shows that MM and adaptive normalization methods outperform other normalization methods. For the sake of simplicity, min-max was applied in this paper. We denoted a raw matching degree ($m_$global or $m_$local) as m_raw from the set M_raw of all match degrees for those types of features (global or local), and the corresponding normalized match degree $m_\mathrm{normalized}$ as

$$m_\mathrm{normalized} = \frac{m - \min(M_\mathrm{raw})}{\max(M_\mathrm{raw}) - \min(M_\mathrm{raw})} \tag{7}$$

This method maps the raw match degrees to the [0, 1] range. The quantities $\max(M_\mathrm{raw})$ and $\min(M_\mathrm{raw})$ specify the end points of the match degree range.

2.2 Feature fusion

We tested with five different fusion methods, namely simple-sum (SS), square, attention fusion function (AFF)[12], feature weighting (FW) and user weighting (UW)[11]. The quantity n_g^i represents the normalized global match degree applied to the i-th gallery image, and n_l^i represents the normalized local match degree applied to the i-th gallery image, $i = 1, 2, \ldots, \bar{n}$, where $\bar{n}$ is the number of gallery images. The fused match degree for the i-th gallery image is denoted as f_i.

Simple-sum (SS)

$$f_i = n_\mathrm{g}^i + n_\mathrm{l}^i \tag{8}$$

Square

$$f_i = \sqrt{(n_\mathrm{g}^i)^2 + (n_\mathrm{l}^i)^2} \tag{9}$$

Attention fusion function (AFF)

$$f_i = \frac{1}{2}\left[(n_\mathrm{g}^i + n_\mathrm{l}^i) + \frac{1}{1+r}\left|n_\mathrm{g}^i + n_\mathrm{l}^i\right|\right] \tag{10}$$

Feature weighting (FW)

Weights are assigned to the individual type of features based on their equal error rates (EER). Denote the EER of global features as e_g and that of local features as e_l. Then the weightings are separately $w_\mathrm{g} = \frac{e_\mathrm{l}}{e_\mathrm{g} + e_\mathrm{l}}$ and $w_\mathrm{l} = \frac{e_\mathrm{g}}{e_\mathrm{g} + e_\mathrm{l}}$. The FW fused degree for the i-th gallery image is calculated as

$$f_i = w_\mathrm{g} n_\mathrm{g}^i + w_\mathrm{l} n_\mathrm{l}^i \tag{11}$$

User weighting (UW)

We assumed that for the i-th gallery image and the type of feature extraction, the mean and standard deviation of the associated genuine and impostor distributions were known. Denote the means of these distributions as $u_\mathrm{g}^i(\mathrm{gen})$, $u_\mathrm{g}^i(\mathrm{imp})$, $u_\mathrm{l}^i(\mathrm{gen})$ and $u_\mathrm{l}^i(\mathrm{imp})$, respectively, and denote the standard deviations as $\sigma_\mathrm{g}^i(\mathrm{gen})$, $\sigma_\mathrm{g}^i(\mathrm{imp})$, $\sigma_\mathrm{l}^i(\mathrm{gen})$ and $\sigma_\mathrm{l}^i(\mathrm{imp})$, respectively. We used the d-prime metric as a measure of the separation of genuine and imposter distributions:

$$d_{\mathrm{g}}^{i}=\frac{u_{\mathrm{g}}^{i}(\mathrm{gen})-u_{\mathrm{g}}^{i}(\mathrm{imp})}{\sqrt{(\sigma_{\mathrm{g}}^{i}(\mathrm{gen}))^{2}+(\sigma_{\mathrm{g}}^{i}(\mathrm{imp}))^{2}}} \tag{12}$$

$$d_{1}^{i}=\frac{u_{1}^{i}(\mathrm{gen})-u_{1}^{i}(\mathrm{imp})}{\sqrt{(\sigma_{1}^{i}(\mathrm{gen}))^{2}+(\sigma_{1}^{i}(\mathrm{imp}))^{2}}} \tag{13}$$

The weightings for gallery image and the two types of feature extraction are $w_{\mathrm{g}}^{i}=\frac{d_{\mathrm{g}}^{i}}{d_{\mathrm{g}}^{i}+d_{1}^{i}}$ and $w_{1}^{i}=\frac{d_{1}^{i}}{d_{\mathrm{g}}^{i}+d_{1}^{i}}$. The UW fused match degree for the i-th gallery image is calculated as

$$f_{i}=w_{\mathrm{g}}^{i}\,n_{\mathrm{g}}^{i}+w_{1}^{i}\,n_{1}^{i} \tag{14}$$

3 Classification

After feature fusion, the fused match degree f between the probe face image $\boldsymbol{P}$ and each gallery image $\boldsymbol{G}$ is obtained. Suppose that the gallery images are $\boldsymbol{G}_1$, $\boldsymbol{G}_2, \ldots, \boldsymbol{G}_{\bar{n}}$, and that each of these images is assigned a given class C_k, $k=1,2,\ldots,c$, and c is the number of individuals in the gallery. If $f(\boldsymbol{P},\boldsymbol{G}_1)=\max_j f(\boldsymbol{P},\boldsymbol{G}_j)$ and $\boldsymbol{G}_1\in C_k$, then the resulting decision is $\boldsymbol{P}\in C_k$.

4 Experiments

4.1 Face database

1 512 images of 126 persons from AR face image database and 1 800 images of 120 persons from SJTU-IPPR face image database were used to test the proposed method.

The AR face image database[13] contains over 4 000 color face images of 126 persons (70 men and 56 women), including frontal views of faces with different facial expressions, lighting conditions and occlusions. The pictures of most persons were taken in two sessions (separated by two weeks). Each section contains 13 color images and 126 individuals participated in both sessions. The images of these 126 individuals were selected and used in our experiment. Only the full facial images were considered here (no attempt was made to handle occluded face recognition and a kind of immoderate expression in each session). One person's sample faces are shown in Fig. 1, where Figs. 1 (a), (b), (c), (d), (e) and (f) are from session 1, and Figs. 1 (o), (p), (q), (r), (s) and (t) are from session 2. The details of the images are as follows: (a) neutral expression, (b) smile, (c) anger, (d) left light on, (e) right light on, (f) all sides light on; and Figs. 1 (o), (p), (q), (r), (s) and (t) were taken under the same conditions as Figs. 1 (a), (b), (c), (d), (e) and (f).

SJTU-IPPR face database includes more than 1 800 images of 120 persons where pose, illumination, expression and background are varied. Each person has 15 images. Fig. 2 is one person's sample images from SJTU-IPPR.

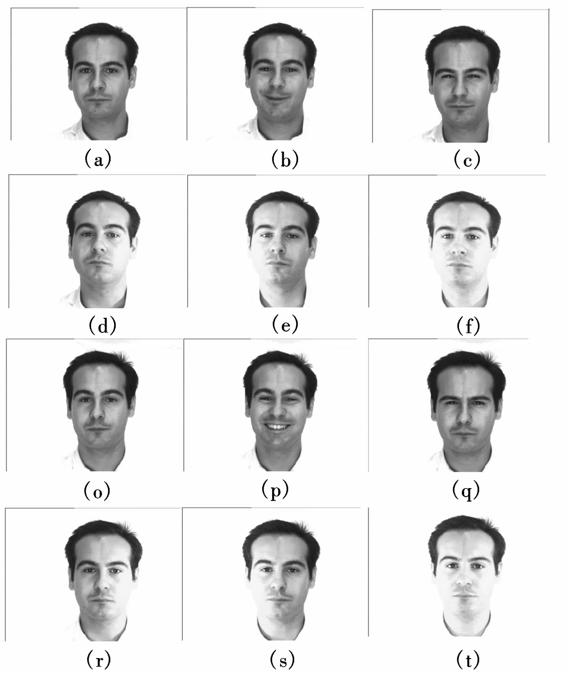

Fig. 1 Sample images for one person from AR face database

Fig. 2 Sample images for one person from SJTU-IPPR face database

4.2 Image normalization

In experiments, color images were transformed to gray ones. Before extraction of global features, firstly, the face region was detected based on Haar-like features and AdaBoost cascade classifier[14]. Secondly, the face image was scaled, translated and rotated, such that the eye centers were in fixed positions. The eye centers were located by fast radial symmetry transform[15]. Thirdly, the face image was cropped using a 128×128 elliptical mask to remove the background. Finally, histogram equalization was applied to the face images for photometric normalization. Before extraction of local features, only the face region was detected and eye centers were located, which were used in scaling, translating and rotating the AAM model.

4.3 Selecting fiducial points for extraction of local features

AAM located 58 fiducial points to outline the main features of a face image, but not all the points were necessary to represent local features. The two types of selecting fiducial points were compared: all 60 fiducial points (58 fiducial points located by AAM and eyes), and 17 fiducial points, as shown in Fig. 3.

(a) (b)

Fig. 3 Two types of selecting fiducial points. (a) 60 fiducial points (58 fiducial points located by AAM and eyes); (b) 17 fiducial points

An experiment was performed using the SJTU-IPPR face database. The first image samples per person was for training, and the remaining images were for testing. The performance of two types of selecting fiducial points was compared. The recognition results are shown in Fig. 4.

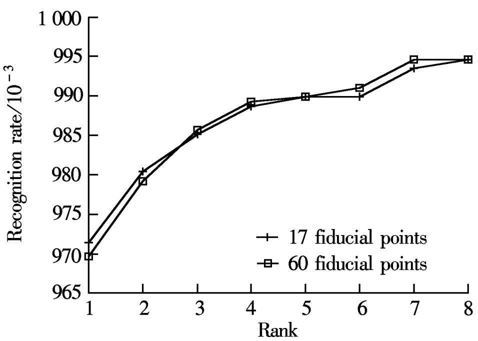

Fig. 4 Comparison of recognition rate of two types of selecting fiducial points

Fig. 4 indicates that one recognition rate of selecting 17 fiducial points is better than that of selecting 60 fiducial points. We did not select fiducial points on outline because the Gabor wavelet analysis may be sensitive to the background on these points. And some fiducial points on the face were rejected to reduce the effect of incorrect location on recognition. Therefore, the type of selecting 17 fiducial points was used in next experiments.

4.4 Comparison of five fusion methods

We tested five fusion methods, SS, square, AFF, FW, and UW. AR and SJTU-IPPR face databases were used to test the recognition rate of the five fusion methods. The recognition rates are shown in Tab. 1.

From Tab. 1, we can see that square and UW outperform other methods. Note that AFF was experimented by choosing $r = 0.5$ in Eq. (10). For the purpose of simplicity, we chose the square fusion method.

Tab. 1 Comparison of recognition rate of five fusion methods %

Face database	SS	Square	AFF($r = 0.5$)	FW	UW
AR	94.08	95.02	94.52	93.15	95.17
SJTU-IPPR	96.43	97.14	96.61	96.01	97.32

4.5 Comparison of three methods

In this experiment, the performance of the proposed method was compared with the other methods, including PCA and EBGM. PCA employed the nearest neighbor classifier for recognition, and EBGM used similarity. In these comparisons, one strategy was "using one image per person for training". The recognition rate of three methods is shown in Tab. 2.

Tab. 2 Comparison of recognition rate of three recognition methods %

Face database	Recognition method		
	PCA	EBGM	Local + Global
AR	90.04	90.48	95.02
SJTU-IPPR	92.20	92.98	97.14

PCA is a method to extract global features, whereas EBGM is a method to extract local features. The recognition rate of the method proposed in this paper is better than the other two methods.

4.6 Performance under variations

The purpose of this experiment was to test the performance of the method under variations over time, in expression and illumination variations. Fig. 1 (a) was used for training, and the others (Figs. 1(b), (c), (d), (e), (f), (o), (p), (q), (r), (s) and (t)) were used for testing. The corresponding recognition rates are illustrated in Tab. 3, where b + p indicates the images corresponding to Fig. 1(b) and Fig. 1(p).

Tab. 3 Recognition rate under variations %

Variation	o	b + p	c + q	d + r	e + s	f + t
Recognition rate	99.21	97.62	92.86	96.83	96.03	89.68

Group (o) indicates that the performance is very high under variations over time. Groups (b + p), (d + r) and (e + s) obtain recognition rates. However, the recognition rates of groups (c + q) and (f + t) are low. Tab. 3 indicates that the performance may be influenced by large variation of expression or illumination

5 Conclusion and Future Work

In this paper, a new method to extract facial features, the fusion of global features and local features, was presented. 1 512 images of 126 persons from an AR face database and 1 800 images of 120 persons from an SJTU-IPPR face database were used to test the method. Compared with PCA and EBGM, the method is more efficient. The method was tested under conditions where time, expression and lighting were varied. When the variation was not large, the performance was very high. Moreover, the computational complexity for recognition was small, and the time for recognizing one probe image was just 0.5 s (CPU: Pentium III 800 MHz; RAM: 256 MB), which was fast enough for application.

There are still some aspects of the location that deserve further study. Though two types of selecting fiducial points were compared, it is unclear how to select fiducial points. Experiments show that the performance is low when the location of 17 fiducial points is not precise. It is necessary to decrease the time consumed for location and feature extraction. It will also be interesting to explore how to fuse global and local features.

References

[1] Givens G, Beveridge J R, Draper B A, et al. How features of the human face affect recognition: a statistical comparison of three face recognition algorithms [A]. In: *IEEE Computer Society Conference on Computer Vision and Pattern Recognition* [C]. Washington, DC, 2004, **2**: 381 − 388.

[2] Zhang Pengfei, Li Desheng, Wang Qi. A novel iris recognition method based on feature fusion [A]. In: *Proceedings of 2004 International Conference on Machine Learning and Cybernetics* [C]. Shanghai, 2004, **6**: 3661 − 3665.

[3] Turk M, Pentland A. Eigenfaces for recognition [J]. *Journal of Cognitive Neuroscience*, 1991, **3**(1): 71 − 86.

[4] Belhumeur P N, Hespanda J, Kriegeman D. Eigenfaces vs. Fisherfaces: recognition using class specific linear projection [J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 1997, **19**(7): 711 − 720.

[5] Moghaddam B. Principle manifolds and probabilistic subspace for visual recognition [J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2002, **24**(6): 780 − 788.

[6] Wiskott L, Fellous J M, Kruger N, et al. Face recognition by elastic bunch graph matching [A]. In: *International Conference on Image Processing* [C]. Washington, DC, 1997, **1**: 129 − 132.

[7] Arca Stefano, Campadelli Paola, Lanzarotti Raffaella. A face recognition system based on local feature analysis [A]. In: 4*th International Conference on Audio-and Video-Based Biometrie Person Authentication* [C]. Guildford, UK, 2003. 182 − 189.

[8] Cootes T F, Edwards G J, Taylor C J. Active appearance models [A]. In: *European Conference on Computer Vision* [C]. Berlin, 1998, **2**: 484 − 498.

[9] Lee Tai Sing. Image representation using 2D Gabor wavelets [J]. *IEEE Transaction on Pattern Analysis and Machine Intelligence*, 1996, **18**(10): 959 − 971.

[10] Liu Chengjun. Gabor-based kernel PCA with fractional power polynomial models for face recognition [J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2004, **26**(5): 572 − 581.

[11] Snelick R, Uludag U, Mink A, et al. Large scale evaluation of multimodal biometric authentication using state-of-the-art systems [J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2005, **27**(3): 450 − 455.

[12] Hua Xiansheng, Zhang Hongjiang. An attention-based decision fusion scheme for multimedia information retrieval [A]. In: *The Fifth Pacific-Rim Conference on Multimedia* [C]. Tokyo, Japan, 2004. 1001 − 1010.

[13] Martinez A M, Benavente R. The AR face database [EB/OL]. http://rvl1.ecn.purdue.edu/~aleix/aleix_face_DB.html, 2003/2005-04-01.

[14] Lienhart Rainer, Kuranov Alexander, Pisarevsky Vadim. Empirical analysis of detection cascades of boosted classifiers for rapid object detection [A]. In: *Proc* 25*th German Pattern Recognition Symposium* [C]. Magdeburg, 2003. 297 − 304.

[15] Loy Gareth, Zelinsky Alexander. A fast radial symmetry transform for detecting points of interest [A]. In: *European Conference on Computer Vision* [C]. Copenhagen, 2002. 358 − 368.

用于人脸识别的特征融合

于威威　滕晓龙　刘重庆

(上海交通大学电子信息与电气工程学院,上海 200030)

摘要:针对人脸识别中人脸图像的特征提取问题,提出了一种将全局特征与局部特征相融合的人脸识别方法.全局特征的提取采用主成分分析算法.主动外观模型定位 58 个特征点,在其中 17 个特征点处进行 Gabor 小波变换则可提取局部特征.归一化的全局匹配度(局部匹配度)可由测试图像和训练图像的全局特征(局部特征)得到.对归一化的全局匹配度和局部匹配度进行融合后,融合匹配度最大的训练图像所属的类即为识别结果.实验利用 2 个人脸图像数据库(AR 和 SJTU-IPPR)测试该方法的识别率,结果表明该方法要优于 PCA 和 EBGM,并且在一定的表情、光照和姿态变化的条件下是有效、稳健的.

关键词:人脸识别;特征融合;全局特征;局部特征

中图分类号:TP391