International Journal of Signal Processing, Image Processing and Pattern Recognition
Vol.9, No.1 (2016), pp.33-42
http://dx.doi.org/10.14257/ijsip.2016.9.1.04

Multimodal Recognition Method based on Ear and Profile Face Feature Fusion

Songze Lei[1] and Min Qi[2]

[1]*School of Computer Science and Engineering, Xi'an Technological University, Xi'an 710032, China*

[2] *School of Electronics and Information, Northwestern Polytechnical University, Xi'an 710072, China*

lei_sz@163.com

Abstract

The performance of ear recognition is influenced by pose variation. For the similar position of ear and profile face, a multimodal recognition method is proposed based on the feature fusion of ear and profile face information. A model for ear and profile face feature fusion and recognition is built. The Log-Gabor features of ear and profile face are first extracted separately, and two features are integrated into a combined feature after two Log-Gabor features are standardized. Then combined feature is mapped to kernel space to fuse further, and acquired stronger discriminant feature for classification by kernel Fisher discriminant analysis (KFDA). The minimum distance classifier is finally used in recognition. Experimental results on the profile face database of Notre Dame University show that the fused method improves the recognition rate of pose variation, and the performance of multimodal recognition is better than unimodal recognition using either ear or profile face alone. The method of ear and profile face feature fusion and recognition is effective and robust for the pose variation.

Keywords*: multimodal recognition, feature fusion, Log-Gabor filter, pose*

1. Introduction

Single biometric feature will be influenced by the limitation of every biometric feature. For instance, someone can not use fingerprint recognition system because of fingerprint scar or calluses. On the voice recognition, voice will change caused by cold and tension, and the voice is easy deceived by recording. Handwriting recognition could be easily affected by the situation and emotion of writer. Also handwriting is easy to be imitated. Face recognition would be affected by expression, pose, makeup, ornament, age, *etc*. Although iris is stable and reliable, but iris recognition will not be performed when eye disease happens. The disturbed way of iris recognition is not easy to accept. In order to overcome intrinsic limitation of single biometric feature, the multimodal biometric feature recognition is proposed. It can improve system's performance, and enhance reliability, and expand the scope of application [1].

Multimodal biometric feature recognition includes face and fingerprint, face and iris, fingerprint and iris, face and ear, *etc*. According to the different abstract extent of information, fusion can be divided into four levels: data-level, feature-level, matching-level, and decision-level [2]. The feature fusion is that multimodal biometric features are fused to form fusion feature vector. The most discriminant feature can be found on the feature fusion, and redundancy of data can be eliminate, and the better perform of recognition can be obtained.

Human ear recognition is a new recognition technology on the field of biometric feature recognition. The ear has the characteristic of uniqueness, universality, stability,

ISSN: 2005-4254 IJSIP

etc. The ear does not be affected by expression, and it is not easy to be harmed. The ear image is easy to be collected, and its size is smaller, its amount of data processing is smaller [3]. It does not be disturbed when recognition. The ear recognition becomes research hotspot on the field of biometric feature recognition. Face and ear have the special position relation. The ear image can be obtained on the profile face image on the procedure of image acquisition. Profile face and ear can be handled at the same time, and they can be fused.

On the former research on ear and profile face fusion, Chang [4] used principal component analysis to extract face and ear multimodal feature for recognition. They used CCA and KCCA to extract ear and profile face relevant feature [5-6], and ear and profile face features were fused. The recognition result of method CCA and KCCA are 97.37% and 98.68% on the USTB database. Abate [7] used iterative function for face and ear recognition, which face and ear feature have local characteristic. It is robust for the occlusion. Abate [7] used GMM to evaluate face and ear Gabor features, and used Dempster-Shafer to fuse face and ear features. Its recognition result are 95.53% and 96.16% on the two database of IIT Kanpur and virtual.

Ear and profile face feature fusion is studied in this paper. Log-Gabor filter is used for analyzing image texture, and achieved good result on the field of image analysis and image processing. Log Gabor filter has the advantages of direction and frequency selectivity of Gabor filter, also has good characteristics that the bandwidth is not restricted, and has no DC component. It can not be affected by background illumination, and can extract texture feature of ear image effectively. In addition, research have shown that Log Gabor filter fits more closely with human vision character than Gabor filter[9], and it can reflect frequency response of natural image more actually.

Kernel method has been proven to express nonlinear data effectively [10]. Vapnik first proposed the kernel method and applied it in Support Vector Machine classifier successfully. Then Scholkopf [11] used it to extract features. Kernel Fisher discriminant analysis is an efficient method to extract the nonlinear features which have stronger discriminability than linear algebraic features. Ear and face image are nonlinear data and are suitable for representation by kernel method.

Ear and profile face feature fusion and recognition model is built in this paper. The ear and profile face Log-Gabor features are extracted, and they are standardized for feature fusion. Kernel Fisher discriminant analysis can represent nonlinear ear and face image data effectively. The standardized fusion feature is analyzed and discriminant feature is extracted by kernel Fisher discriminant analysis. The performance of ear and profile face feature fusion recognition method is better than uni-modal recognition method.

2. Recognition Model of Ear and Profile Feature Fusion

The model of ear and profile face feature fusion is shown on Figure 1. The ear image and profile face image are first pre-processed, and then initial feature vectors of ear and profile face are extracted by Log-Gabor filter respectively. The both initial feature vector will be standardized. The standardizing feature vectors of ear and profile face are integrated to combination feature. Then the combination feature is mapped to kernel space to form fusion feature that is fit for the classification. The minimum Euclidean distance is used for recognition.

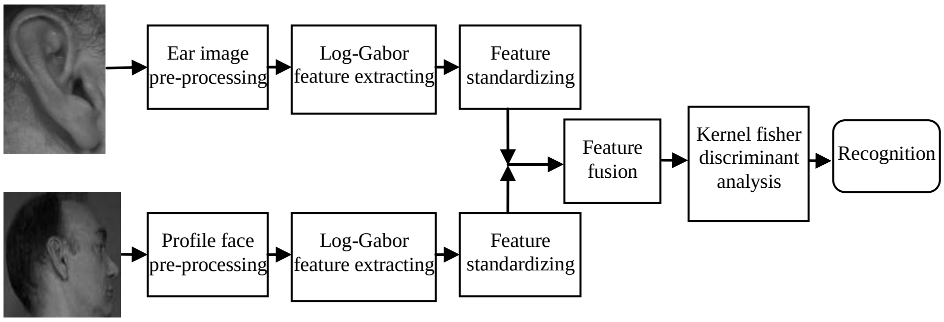

Figure 1. Recognition Model of Ear and Profile Feature Fusion

2.1. Log-Gabor Filters

In [6] Field proposes an alternate method to perform both the DC compensation and to overcome the bandwidth limitation of a traditional Gabor filter. The Log-Gabor filter has a response that is Gaussian when viewed on a logarithmic frequency scale instead of a linear one. This allows more information to be captured in the high frequency areas and also has desirable high pass characteristics.

Log-Gabor filters in frequency domain can be defined in polar coordinates by $H(f, \theta) = H_f \times H_\theta$ where H_f is the radial component and H_θ, the angular one:

$$H(f,\theta) = \exp\left\{\frac{-[\ln(f/f_0)]^2}{2[\ln(\sigma_f/f_0)]^2}\right\} \times \exp\left\{\frac{-(\theta-\theta_0)^2}{2\sigma_\theta^2}\right\}$$

(1)

with f_0, the central frequency, θ_0, the filter direction, σ_f ,which defines the radial bandwidth and σ_θ, which defines the angular bandwidth. In order to maintain constant shape ratio filters, the ratio of σ_f / f_0 should be maintained constant. In the following experiments the shape parameter is chosen such that each filter has a bandwidth of approximately 2 octaves and the filter bank was constructed with a total of 8 orientations and 5 scales.

Amplitude of Log-Gabor shows the local energy spectrum of image, and it should be regarded as the intensity of edge of specific orientation. The amplitude of Log-Gabor filter is used, and it is benefit for ear recognition.

2.2.Ear and Profile Face Images Log-Gabor Filtering and Parameter Selection

Ear and profile face images are filtered by Log-Gabor to extract Log-Gabor feature. This is image convolution operation, as follows:

$$LG_{u,v}(x,y) = E(x,y) * \Phi_{u,v}(x,y)$$

(2)

where * expresses convolution operation, E(x,y) is ear or profile face image, and $\Phi_{u,v}(x,y)$ is Log-Gabor filter with u scales and v orientations. $LG_{u,v}(x,y)$ is the result of convolution operation. Generally amplitude will be selected. Figure 2 shows the convolution result of ear and profile face with 5 scales and 8 orientations.

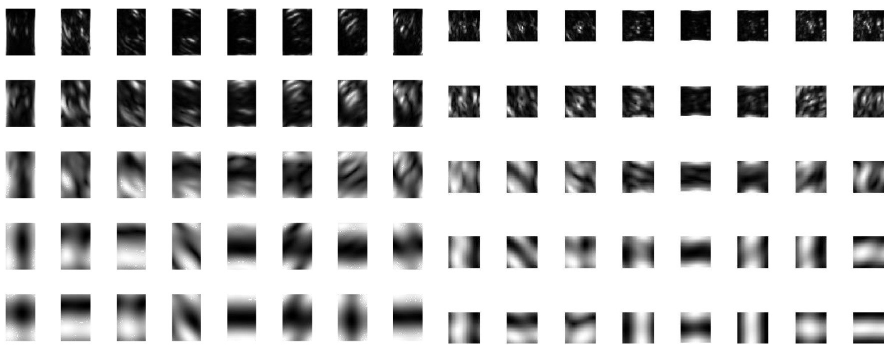

Figure 2. Log-Gabor Filter Amplitude of Ear and Profile Face Image

It is necessary to select proper filter parameters when using Log-Gabor filters to extract ear feature. For the contour of inner and outer ear image has various orientation, the orientation of filters should covering all space orientation. The contour of inner ear image being influenced by position and illumination, filters should be computed within wide frequency range so as to acquire abundant and accurate features.

Log-Gabor filters of 8 orientations are used to extract ear feature. The 8 orientations cover 0°~180° space band of ear image, in which contour feature of every orientations of ear can be obtained. Parameter σ_θ defines orientation bandwidth, which σ_θ's value is bigger, and orientation bandwidth is bigger. The parameter σ_θ is empirically selected as 1.5.

The scales of filters are selected 5. The parameter σ_f defines radial bandwidth, which σ_f's value is bigger, and radial bandwidth is bigger. By the character of Log-Gabor, radial bandwidth has approximately 1 octave as σ_f / f_0 is 0.74, and radial bandwidth has approximately 2 octaves as σ_f / f_0 is 0.55. For ear image, the bigger σ_f's value, the better contour construction would be preserved. But this reduces the denoising ability of filters. The parameter σ_f / f_0 is empirically selected as 0.55, which it could get high performance of two sides.

2.3. Feature Standardization and Combination

The ear feature extraction and profile face feature extraction are two independent processes. They output the ear Log-Gabor feature vector E and the profile face feature vector F. Because the dimension of two biological features is different, it will be quite different on the amount of the ear feature E and profile face feature F[13]. If two original features are combined directly, the data ratio will be imbalance. To eliminate the effect from two original features fusion, the standardization needs to be used. Suppose the total of training sample is c, the feature vector of training sample is expressed as Y_i, i =1,2,…,c. The standardization procedure of feature is as follows:

$$\mu = \frac{1}{c}\sum_{i=1}^{c} Y_i \tag{3}$$

$$\sigma = \frac{1}{c}\sum_{i=1}^{c} \left| Y_i - \mu \right| \tag{4}$$

$$Y_i' = \frac{Y_i - \mu}{\sigma}, \quad i = 1,2,\cdots,c \tag{5}$$

where μ is the mean of feature, σ is the variance of feature, Y_i ' is the feature after standardization. The original ear feature E and profile face F are transformed to matrix $E'$and $F'$by standardization. The vector $E'$and $F'$are combined from begin to end to form a new feature vector, that is EF=[E', F'].

Image filtered by Log-Gabor to form feature vector. The dimension of feature vector is 40 times the number of original image. If the size of image is 80 ×50, the dimension of Log-Gabor feature vector is 80 ×50×40=160000. Classification is very difficult using the high dimension. In order to reduce complexity, down-sampling with 2^k is used to decrease the feature size of each $LG_{u,v}(x,y)$. Then all $LG_{u,v}(x,y)$ column vectors are connected in sequence to one vector. The kernel Fisher discriminant analysis is used to extract feature for classification in next.

2.4. Feature Fusion and Recognition using KFDA

Kernel Fisher discriminant analysis (KFDA) was proposed in [12] to extend the Fishers' linear discriminant analysis nonlinearly. KFDA uses kernel trick to map input data into an implicit feature space. Then FLDA is used to detect the linear relations and discriminate features of input data. KFDA is capable of handling high-dimensional data and extracting most discriminant features for classification.

A image can be described as a point x in a high-dimensional (R^d) space.

The input data is projected into a nonlinear high-dimensional feature space:

$$\Phi : R^d \rightarrow F, x \mapsto \phi(x)$$
(7)

Then LDA can be performed in the F space. The fisher criterion function in the feature space for the same set of M patterns presented in the LDA section, can be defined as

$$J(V) = \frac{V^T S_b^\phi V}{V^T S_w^\phi V}$$
(8)

where S_b^ϕ is the between class scatter matrix and S_w^ϕ is the within class scatter matrix defined in the feature space F. We assume that the projected samples $\phi(x)$ are centered in feature space F, *i.e.*, $\sum_i^M \phi(x) = 0$.

Since all the solutions of V lies in the span of $\{\varphi(x_1),\varphi(x_2),...,\varphi(x_M)\}$, and there exists coefficients α_i $(i=1,2,...,M)$ such that

$$V = \sum_{i=1}^{M} \alpha_i \phi(x_i) = Qa$$
(9)

where $Q = [\varphi(x_1),...,\varphi(x_M)]$ and $a = (\alpha_1,\ldots, \alpha_M)^T$.

Substituting equation (9) into equation (8), the Fisher criterion is converted to:

$$J(\alpha) = \frac{\alpha^T (KWK)\alpha}{\alpha^T (KK)\alpha}$$
(10)

where the matrix K is an $M \times M$ matrix centered in the feature space. The kernel function is corresponding to a given nonlinear mapping φ. Any function satisfying Mercer's condition can be used as K. The Gaussian kernel functions are as follows:

$$k(x_i, x) = \exp(-\|x_i - x\|^2 / \sigma^2) \tag{11}$$

The $M \times M$ matrix W can be defined by

$$W = diag(W_1, \cdots, W_C) \tag{12}$$

where W_i is a ($N_n \times N_n$) matrix with terms all equal to $1/N_n$. In this paper, gaussian kernel is used to construct discriminant feature. The projection of a point x onto W in F is given by: $(W \cdot \phi(x)) = \sum_{i=1}^{M} \alpha_i k(x_i, x)$.

3. Experiments Results

To evaluate proposed method, we have used the Collection E database of University of Notre Dame [14]. The database includes 464 images from 114 individuals, and each individual has 2 to 6 images. Each image has variations of the head position and different light. The 113 individuals and 3 images per individual are selected in this article.

All the images are cropped to fit the profile face and ear. The ear images are normalized to the size of 80×50 pixels according to the ratio of length: width=1.6 by bilinear interpolation algorithm. The profile face images are normalized to the size of 64×64 pixels. The size of ear image is basically the same as the size of profile face image in order to compare result with each other. All the images are preprocessed by means of the histogram equalization. The sample images of profile face and ear are showed on Figure 3.

All the calculation is implemented using Matlab 7 on an Intel i5 2.67 G Hz processor with 4 GB memory. Three images from each subject are used for training, whilst the remainder is used for test. The high dimensional Log-Gabor features are analyzed to extract low dimensional discriminant feature for classification by kernel Fisher discriminant analysis, which kernel function is selected as gaussian kernel. Individual is classified with minimum Euclidean distance.

(a) Samples of profile face image (b) Samples of ear image

Figure 3. Samples of Experiment

Experiments of ear, profile face and both feature fusion recognition is carried out respectively. The experimental results are shown on Table 1. As shown on Table 1, the correct recognition rate of feature fusion is higher than that of single feature for three test sets. For the test sets of group 1 and group 3, the results are similar which recognition rate is high, because the pose of both groups is similar. But for the group 2 with pose variation, the result is different from other groups, the recognition rate of ear is 93.81%, the result of profile face is 91.15% and the result of feature fusion is 99.12%. The recognition rate of feature fusion is higher than single feature. The results shows that proposed method of feature fusion is more effective than single ear or single profile face feature. The recognition time is 0.205 second which satisfy real time requirement.

Table 1. Result of Recognition Rate using Proposed Method

Test set	ear	Profile face	Ear and profile face fusion	Recognition time
Group 1	99.12%	98.23%	99.12%	
Group 2	93.81%	91.15%	99.12%	0.205 s
Group 3	99.12%	97.35%	100%	

The parameters of Log-Gabor should be selected reasonably. Its performance depends on four parameters: orientation bandwidth σ_θ, scale bandwidth σ_f, minimum scale wavelength $1/f_0$, and σ_f/f_0. Parameter σ_θ defines orientation bandwidth, which σ_θ's value is bigger, and orientation bandwidth is bigger. The parameter σ_θ is empirically selected as 1.6. In order to extract effective feature in all directions under the multi-scale, the filters are constructed with a total of 5 scales and 8 orientations. The parameter minimum scale wavelength is set to 10, and radial bandwidth is 3, and σ_f/f_0 is 0.55. The Log-Gabor with these parameter values can extract effective feature of ear's folds.

Figure 4 shows the recognition rate of group 2 test images for varying σ_f/f_0 and f_0. As shown in figure 4, the recognition rate is between 94.8% and 96.1% when σ_f/f_0 is between 0.2 and 0.7, which recognition rate is in a narrow range. It reaches highest the recognition rate when σ_f/f_0 is between 0.6 and 0.7 for one of the f_0, and the recognition rate descends when σ_f/f_0 is outside of range of 0.6 to 0.7. When f_0=1/10 and σ_f/f_0=0.6 ~ 0.65, it has the highest the recognition rate and good overall performance.

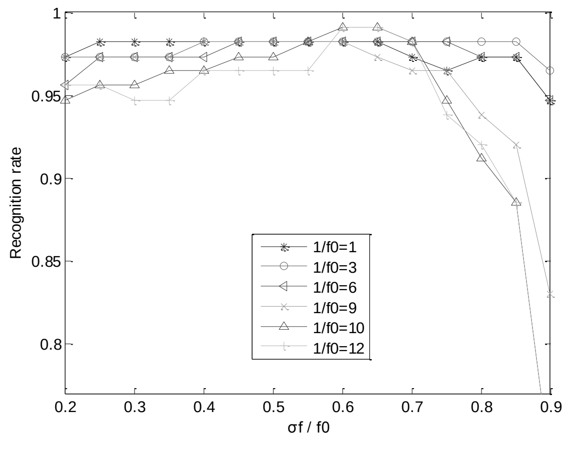

Figure 4. Result of Parameters σf /f0 and f0 Variation

Figure 5 shows the recognition rate for varying parameter σ of gaussian kernel. As shown in Figure 5, recognition rates of group 1 to group 4 are quite stable except some small fluctuation when parameter σ >200. The fluctuation of group 2 is relatively big in comparison to other two groups, but its fluctuation is in the small range of 0.02%. The recognition rate reaches the highest when σ is 900 to 1000 for all three groups.

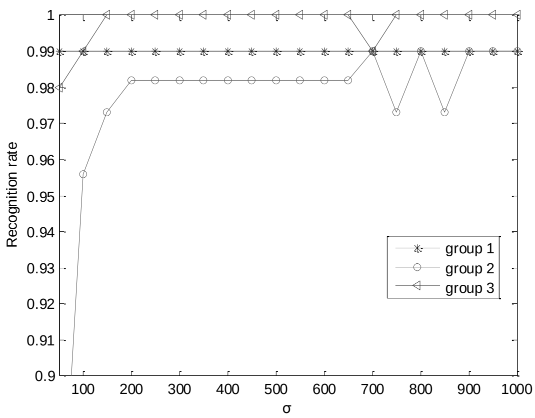

Figure 5. Recognition Rate of Three Test Groups for Varying Parameter σ

Other recognition methods are in Table 2 as comparison. From Table 2 the recognition rate of proposed method is highest in all methods. The results show that profile face and ear feature fusion of Log-Gabor, and classification by kernel Fisherfaces discriminant analysis can achieve good recognition result.

The method of PCA or Fisherfaces is that ear and profile face images are transformed to one column vector directly, and using PCA and Fisherfaces is for recognition after two vectors are combined. The results of PCA and Fisherfaces methods in Table 2 show that PCA and Fisherfaces method are global method, which could not represent details of image enough. They are not suitable for pose recognition. The method of KFDA which is original image plus kernel Fisherfaces discriminant analysis obtains high recognition rate. This shows kernel method is effective for the recognition of variety of ear angle. The different results shows that feature representation of Log-Gabor with multi-scale and multi-orientation plus feature fusion using kernel Fisherfaces discriminant analysis could have strong robustness for multi-pose.

Table 2. Recognition Rate Comparison of Different Methods

Method	Recognition rate		
	Group 1	Group 2	Group 3
PCA[15]	98.23%	90.27%	97.35%
Fisherfaces[16]	98.23%	91.15%	96.46%
KFDA	98.23%	92.04%	98.23%
Proposed method	99.12%	99.12%	100%

4. Conclusions

Single biometric feature has intrinsic limitation for recognition. The multimodal recognition emerges from various practical factors. Because the human ear recognition is influenced by the pose and so on, the correct recognition rate degrades much. For the ear and profile face are near, multimodal fusion of ear and profile face is feasible solution. The feature fusion of ear and profile face is studied, and the model of both feature fusion and multimodal recognition method are proposed. The feature extraction by Log-Gabor filter for ear and profile face is effective method. standardizing each Log-Gabor feature before feature fusion is necessary. With effective nonlinearity data processing of kernel method, the KFDA method is used for the further fusion and extracting discriminative feature. Experimental results show that the proposed method is effective for the pose recognition. Multimodal recognition can improve the pose recognition performance compare with single feature.

Acknowledgements

This work is Scientific Research Program funded by Shaanxi Provincial Education Department (Program No. 2010JK595). Thanks for University of Notre Dame Biometrics Database (Collection E).

References

[1] R. W. Frischholz and U. Dleckemann, “BioID: a multimodal biometric identification system”, Computer, vol. 33, no. 2, **(2000)**, pp. 64-68.
[2] M. Faundez-Zanuy, “Data fusion in biometrics”, Aerospace and Electronic Systems Magazine, vol. 20, no. 1, **(2005)**, pp. 34-38.
[3] M. Burge and W. Burger, “Ear biometrics in computer vision”, Proceedings of the 15th International Conference on Pattern Recognition, Barcelona, Spain, **(2000)** September, pp. 822-826.
[4] K. Chang, K. W. Bowyer, S. Sarkar and Barnabas, “Comparison and combination of ear and face Images in appearance-based biometrics”, IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 25, no. 9, **(2003)**, pp. 1160-1165.
[5] X. N. Xu and Z. C. Mu, “Multimodal recognition using ear and face profile based on CCA”, Application Research of Computer, vol. 24, no.11, **(2007)**, pp. 312-314.
[6] X. N. Xu, Z. C. Mu, X. Q. Pan and Y. Zhao, “Feature fusion method based on KCCA and multi-modality recognition fusing ear and face profile information”, Journal of South China University of Technology (Natural Science Edition), vol. 36, no.9, **(2008)**, pp. 117-121.
[7] A. F. Abate, M. Nappi and D. Riccio, “Face and ear: a bimodal identification system”, Third International Conference on Image Analysis and Recognition, Póvoa de Varzim, Portugal, **(2006)** September, pp. 297-304.
[8] D. R. Kisku, P. Gupta, H. Mehrotra and J. K. Sing, “Multimodal belief fusion for face and ear biometrics”, Intelligent Information Management, vol. 1, no. 2, **(2009)**, pp. 166-171.
[9] D. J. Field, “Relations between the statistics of natural images and the response properties of cortical cells”, Journal of the Optical Society of America, vol. 4, no. 12, **(1987)**, pp. 2379-2394.
[10] H. Peng, B. Li, X. H. Luo, J. Wang and Z. L. Zhang, “A learning-based audio watermarking scheme using kernel Fisher discriminant analysis”, Digital Signal Processing, vol. 23, no. 1, **(2013)**, pp. 382-389.
[11] T. Ishii, M. Ashihara and S. Abe, “Kernel discriminant analysis based feature selection”, Neurocomputing, vol. 71, no. 13-15, **(2008)**, pp. 2544-2552.
[12] S. Mika, G. Ratsch, J. Weston, B. Scholkopf and K. R. Muller, “Fisher discriminant analysis with kernels”, IEEE International Workshop on Neural Networks for Signal Processing IX, Madison, USA, **(1999)** August , pp. 41-48.
[13] C. K. Chen, Z. Q. Wang, J. Y. Yang and J. Yang, “Nonlinear Discriminant Feature Fusion Method for Face Recognition”, Mini-Micro Systems, vol. 26, no. 5, **(2005)**, pp.793-797.
[14] http://cse.nd.edu/~cvrl/CVRL/Data_Sets.html
[15] M. Turk and A. Pentland, “Eigenfaces for recognition”, Journal of Cognitive Neuroscience, vol. 3, no. 1, **(1991)**, pp.71-86.
[16] P. N. Belhumer, J. P. Hespanha and D. J. Kriegman, “Eigenfaecs vs Fisherfaces: recognition using class specific linear projection”, IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 19, no. 7, **(1997)**, pp. 711-720.

Author

Songze Lei, He is an Associate Professor. He received PhD in information and communication engineering from Northwestern Polytechnical University. Now he works at Xi’an Technological University. His main research interests include image processing and pattern recognition.

