Journal of Information Hiding and Multimedia Signal Processing ©2011 ISSN 2073-4212
Ubiquitous International Volume **2**, Number **3**, July **2011**

Applications of Block Linear Discriminant Analysis for Face Recognition

CHUAN-YU CHANG[1], CHUAN-WANG CHANG[2] and CHING-YU HSIEH[1]

[1]Department of Computer Science and Information Engineering
National Yunlin University of Science & Technology
No. 123, Sec. 3, University Rd., Douliou, Yunlin, Taiwan
chuanyu@yuntech.edu.tw
[2]Department of Computer Application Engineering
Far East University
No.49, Zhonghua Rd, Xinshi, Tainan, Taiwan
chuan@cc.feu.edu.tw

Received April 2011; revised July 2011

Abstract. *Face recognition is an important issue in image analysis and understanding. Linear discriminant analysis (LDA) has been widely used in face recognition. However, the LDA-based face recognition methods usually encountered the small sample size (SSS) problem. The SSS problem occurs when the number of samples is far smaller than the dimensionality of the sample space. Therefore, this paper proposed a modified LDA (called block LDA) to divide the gradient image into several non-overlapping subimages of the same size, in order to increase the quantity of samples and reduce the dimensions of the sample space. In addition, to reduce the influence of illumination variations, face images were transferred to gradient image. Experimental results show that the proposed method indeed solves the SSS problem with a good recognition rate.*
Keywords: face recognition; linear discriminant analysis; small sample size problem

1. **Introduction.** In recent years, facial analyses including face detection, face recognition, and facial expression recognition play important rules for human-centered applications [1, 2, 3]. In which, face recognition is a critical issue in the film processing, identification card recognition, security systems, and criminal identification systems. Challenges in face recognition include illumination variation, pose variation, facial expression, aging, hair, and glasses.

Many face recognition algorithms have been proposed to solve these problems in the past few years. Template matching is to detect the positions of specific geometric features on the face (such as the eyes, nose and lips), so that their relative positions can be used as facial features for matching or recognition [4, 5, 6]. The principle component analysis (PCA) is to make the high-dimensional space into a lower-dimensional eigenspace through a proper transformation or projection, while retaining most of the principle components [7, 8]. Independent component analysis (ICA) is the method finding the independent components (ICs) to estimate independent sources only through the information for random variable observed, but without the mixing mechanism and sources [9]. Linear discriminant analysis (LDA), which distances the centroids of different classes and narrows the scatters of the same classes, is used to increase the recognition rate [7, 10, 11, 12].

Among these face recognition methods; PCA and LDA are the most popular ones. However, these methods usually encounter the small size sample (SSS) problem, where the dimensions of image space are more than the sample sizes, making the sample less representative. In addition, the within-class scatter matrix may be singular in LDA, which resulted in fail to calculate the inverse matrix and the transformation matrix.

This paper proposed a modified LDA (called block LDA) to solve the SSS problem, we divided the face image into several non-overlapping subimages of the same size. Accordingly, the number of samples is increased, and the dimension of the sample is decreased. In addition, to reduce the influence of illumination variation, all face images are transformed to gradient images.

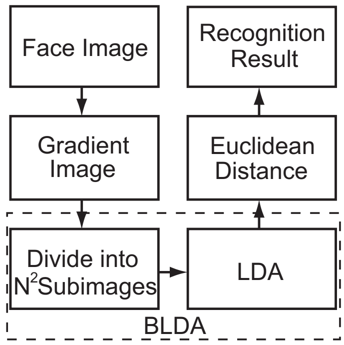

FIGURE 1. Flowchart of the proposed method

Figure 1 shows the flowchart of the proposed method. First, original face image is transformed to gradient image. After that, the gradient image is divided into N^2 subimages. Then, these subimages are projected to feature vectors through LDA. Finally, a Euclidean measurement is applied to determine the recognition result.

The remainder of this paper is organized as follows: Section 2 describes the gradient image; Section 3 describes traditional LDA and BLDA; Section 4 describes the experimental result; and Section 5 gives the conclusions.

2. **Gradient Image.** In general, illumination variations should not highly affect the recognition results of human eyes. The dominant factors of face recognition are facial shape and facial features. Thus, the original facial image is transformed to gradient image for face recognition. Gradient image contains relative changes in gray level [13], which is less sensitive to illumination variations.

The gradient of an image $f(x, y)$ at location (x, y) is defined as

$$\nabla \mathbf{f} = \begin{bmatrix} \mathbf{G}_x \\ \mathbf{G}_y \end{bmatrix} = \begin{bmatrix} \frac{\partial f}{\partial x} \\ \frac{\partial f}{\partial y} \end{bmatrix} \tag{1}$$

The magnitude of the gradient is defined as

$$\nabla f = mag(\nabla \mathbf{f}) = [\mathbf{G}_x^2 + \mathbf{G}_y^2]^{\frac{1}{2}} = \left[(\frac{\partial f}{\partial x})^2 + \frac{\partial f}{\partial y})^2 \right]^{\frac{1}{2}} \tag{2}$$

This quantity gives the maximum rate of $f(x, y)$ increase of per unit distance in the direction of $\nabla \mathbf{f}$. In this paper, the Sobel operators as show in Figure 2(a) and 2(b) are used to obtain the gradient image. Figure 2(c) and (d) show an original facial image and its corresponding gradient image, respectively.

-1	-2	-1
0	0	0
1	2	1

(a)

-1	0	1
-2	0	2
-1	0	1

(b)

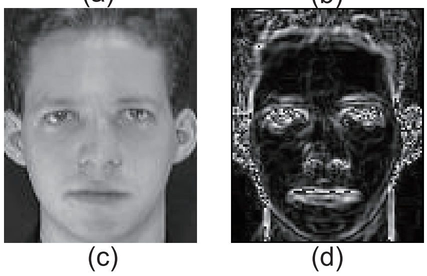

FIGURE 2. Original image and its corresponding gradient image

3. **Linear Discriminant Analysis.** The LDA is able to project an image from a high-dimensional space to a low- dimensional space, thus maximizing the ratio between between-class scatter matrix and within-class scatter matrix. The LDA not only reduces the dimensions, but also distances the distribution of the images. Figure 3(a) shows synthetic two-dimensional data points belong to two classes. The data along with its one-dimensional LDA transformation are shown in Figure 3(b). Notice that, the dimensionality reduction and the important discriminatory properties are fully retained.

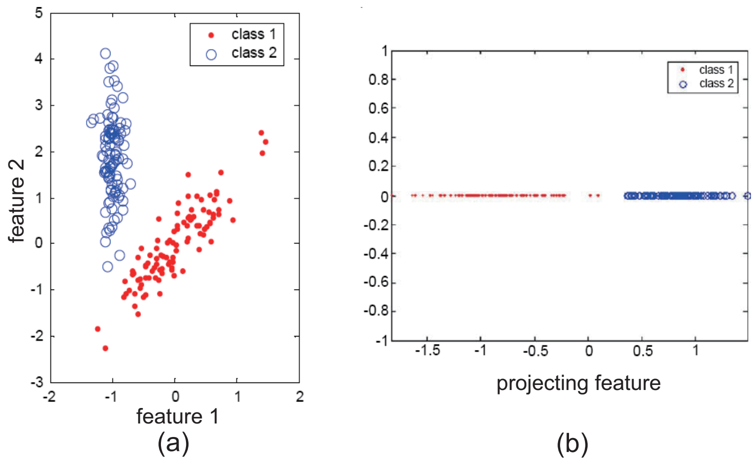

FIGURE 3. (a)Original data set distribution. (b) Feature vectors distribution in (a) through LDA

The detail of the traditional LDA and the proposed BLDA are described as follows.

3.1. **Traditional LDA.** Assumed that there are C classes of images in the database, each class has M images. The feature vector for the m-th image from the c-th category in the original image is set to $\mathbf{X}_m^c$. The distance between the feature vectors and the centroid of this class, can be expressed by the formula for within-class scatter matrix, $\mathbf{S_W}$:

$$\mathbf{S_W} = \sum_{c=1}^{C} \sum_{m=1}^{M} (\mathbf{X}_m^c - \overline{\mathbf{X}}^c)(\mathbf{X}_m^c - \overline{\mathbf{X}}^c)^T \tag{3}$$

where $\overline{\mathbf{X}}^c$ is the average of all the feature vectors in the class c.

The scatter degree among different classes, that is, the distance between the centroids of different classes, can be expressed by the formula for between-class scatter matrix, $\mathbf{S_B}$:

$$\mathbf{S_W} = \sum_{c=1}^{C} (\overline{\mathbf{X}}^c - \overline{\mathbf{X}})(\overline{\mathbf{X}}^c - \overline{\mathbf{X}})^T \tag{4}$$

where $\overline{\mathbf{X}}$ is the average of all the original feature vectors.

If the within-class scatter matrix is nonsingular, then the best projection matrix $\mathbf{W_{opt}}$ can be determined by the ratio of determinants between between-class scatter matrix and within-class scatter matrix:

$$\mathbf{W_{opt}} = \arg \begin{matrix} \max \\ w \end{matrix} \; trace\left\{ \frac{\mathbf{W^T S_B W}}{\mathbf{W^T S_W W}} \right\} = \arg \begin{matrix} \max \\ w \end{matrix} \; trace\left\{ \mathbf{S_W^{-1} S_B} \right\} \tag{5}$$

To find the $\mathbf{W_{opt}}$ which maxim izes $\mathbf{S_w^{-1} S_B}$, the result is derived as $\{w_i | i = 1, 2, ..., p\}$, which is the largest eigenvalue in the first p of $\mathbf{S_w^{-1} S_B}$, $\{\lambda_i | i = 1, 2, ..., p\}$ is the corresponding eigenvector. They meet the equation below:

$$\mathbf{S_W^{-1} S_B w_{opt}} = \lambda_\mathbf{i} \mathbf{w_{opt}} \tag{6}$$

Basically, only the eigenvectors corresponding to the first p eigenvalues are selected. In this paper, p is selected by taking the first p largest eigenvalues whose total sum accounts for 95% of the sum of the general features. Generally, most energy of images can be kept in this way. The rest of eigenvectors corresponding to the trivial eigenvalues can be ignored.

3.2. **Block LDA.** Since the traditional LDA inheres the SSS problem, that is, when the dimensions of image space is larger than the number of training samples, the within-class scatter matrix becomes a singular matrix, and is unable to calculate its inverse matrix. Thus, the transformation matrix $\mathbf{W_{opt}}$ in Eq.(5) cannot be calculated. Therefore, to solve the SSS problem, this paper proposed a method to divide the input images into several non-overlapping subimages of the same size. It could not only increase the sample size, but also reduce the dimension of the training samples.

In this paper, each input image is divided into N^2 subimages. If the original dimension of the image is $W \times H$, where W is the width of an image and H is the height of an image, and then it is divided into subimages of the size of $W/N \times H/N$. The subimage (j, k) is obtained by:

$$\mathbf{X}_{j,k}(\alpha, \beta) = \mathbf{X}\Big(\frac{W}{N}(j-1) + \alpha, \frac{H}{N}(k-1) + \beta\Big) \tag{7}$$

where j and k are between 1 and N; α is between 0 and $(W/N) - 1$; and β is between 0 and $(H/N) - 1$. Figure 4 shows the schematic diagram of image division. After image division, the within-class scatter matrix $\mathbf{S_W}$ is modified as

$$\mathbf{S_W} = \sum_{c=1}^{C} \sum_{m=1}^{M} \sum_{n=1}^{N^2} (\mathbf{X}_{c,m}^n - \overline{\mathbf{X}}_c)(\mathbf{X}_{c,m}^n - \overline{\mathbf{X}}_c)^T \tag{8}$$

where $\overline{\mathbf{X}}_\mathbf{c}$ represents the average image of all subimages of the c-th class, and is defined as:

$$\overline{\mathbf{X}}_c = \frac{1}{MN^2}\sum_{m=1}^{M}\sum_{n=1}^{N^2}\mathbf{X}_{c,m}^{n} \tag{9}$$

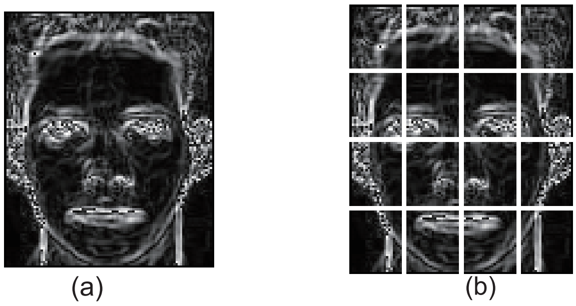

FIGURE 4. The schematic diagram of image division, (a) a gradient image, (b) the gradient image is dividing into 16 subimages

The new between-class scatter matrix $\mathbf{S_B}$ is rewritten as

$$\mathbf{S_B} = \sum_{c=1}^{C}(\overline{X}_c - \overline{X})(\overline{X}_c - \overline{X})^T \tag{10}$$

The average images $\overline{\mathbf{X}}$ of all subimages is defined as

$$\overline{\mathbf{X}} = \frac{1}{CMN^2}\sum_{c=1}^{C}\sum_{m=1}^{M}\sum_{n=1}^{N^2}\mathbf{X}_{c,m}^{n} \tag{11}$$

Similarity, the optimum projection matrix $\mathbf{W_{opt}}$ is calculated by

$$\mathbf{W_{opt}} = \arg\ \begin{matrix}\max\\ w\end{matrix}\ trace\left\{\frac{\mathbf{W^T S_B W}}{\mathbf{W^T S_W} W}\right\} = \arg\ \begin{matrix}\max\\ w\end{matrix}\ trace\left\{\mathbf{S_W^{-1} S_B}\right\} \tag{12}$$

and satisfies the equation below:

$$\mathbf{S_W^{-1} S_B w_{opt}} = \lambda_{\mathbf{i}}\mathbf{w_{opt}} \tag{13}$$

As the traditional LDA, the first p largest eigenvalues are obtained. Since some eigenvalues and eigenvectors may be complex numbers, $\left\|Z\right\|_2$ is calculated, where $Z = v + oi$, v is the real number part and o is the imaginary number part.

Accordingly, the feature vectors of the n-th subimage is obtained by

$$\mathbf{y}^n = \mathbf{W_{opt}}^T\mathbf{X}^n \tag{14}$$

Finally, the feature vectors are combined according to the dividing order as

$$\mathbf{Y} = [\mathbf{y}^1\mathbf{y}^2\cdots\mathbf{y}^n] \tag{15}$$

During testing, the optimal projection matrix $\mathbf{W_{opt}}$ calculated above is retained. The testing image is also divided into N^2 subimages, and every subimage $\mathbf{X_{testn}}$ is projected by the optimal projection matrix $\mathbf{W_{opt}}$ individually. Thus, the n-th feature vector y_{test}^n of the testing subimages is obtained by:

$$\mathbf{y_{test}^n} = \mathbf{W_{opt}^T X_{test}^n} \tag{16}$$

Then combine the feature vectors according to the dividing order:

$$\mathbf{Y_{test}} = [\mathbf{y^1_{test} y^2_{test}} \cdots \mathbf{y^n_{test}}] \tag{17}$$

The testing image and all the training images are compared to find the one with the shortest Euclidean distance, which is the image with the highest similarity. Assume that there are C classes (people) and each one has m facial images in the training images, the feature vector of the mth image in the c class is $\mathbf{Y}^c_m$, and the feature vector of testing face image is $\mathbf{Y_{test}}$, then the category of the testing image is determined by:

$$\tau = \underset{c}{\arg} \min \left\| \mathbf{Y}_{test} - \mathbf{Y}^c_m \right\|_2 \tag{18}$$

4. **Experimental Results.** In order to evaluate the performance of the proposed method, three experiments were performed: (1) determination of the number of dividing subimages, (2) the recognition rates with different preprocessing, and (3) the recognition rate comparison with various databases and approaches. ORL, Yale, Japanese Female Facial Expression database (JAFFE), and MIT-CBCL face databases were used in the experiments.

ORL (Olivetti Research Library) face image database, whose data were collected by AT&T Laboratories Cambridge from April 1992 to April 1994. The database contains 40 people; each has 10 images, which are taken at different times, different facial angles (left, right, look down and up), the different expressions (laugh or not laugh), wear glasses or without glasses. Each image is 92×112, 256 gray-level image with black ground. Samples from the ORL face database are shown in Figure 5.

Yale Database contains 15 people; each one has 11 images with different illumination and facial expressions, and with or without glasses. Each image is 320×243, 256 gray-level image. Samples from the Yale face database are shown in Figure 6.

Figure 5. Partial face images from ORL database.

JAFFE contains 210 images of 10 females with 7 expressions, including neutral, happy, angry, surprised, sad, disgusted, and scared. Each image is 256×256, 256 gray-level image. Samples from the JAFFE face database are shown in Figure 7.

The MIT-CBCL face database contains 10 people; each one has 324 images with different illumination and facial angles. Samples from the MIT-CBCL face database are shown in Figure 8.

A personal computer with a Pentium 4-3.0GHz and 1GB memory was used to perform the experiments. The operating system was Microsoft Windows XP SP3. The recognition rate is defined as

Figure 6. Partial face images from Yale database.

Figure 7. Partial face images from JAFFE database.

FIGURE 8. Partial face images from MIT-CBCL database.

$$\eta = \frac{R}{A} \times 100\% \tag{19}$$

where R is the number of image correctly recognized and A is the total number of the testing images.

Experiment I: Determination of the number of dividing subimages

Assume that there are C classes (people) in the training image set, in which each class consists M images. To obtain the optimal transform matrix of LDA, the $\mathbf{S_W}$ must be a nonsingular matrix. That is the dimension of a subimage must far bigger than the number of training images.

To solve the SSS problem, each image was divided into $N \times N$ subimages. Then the size of each subimage is $W/N \times H/N$, where W and H are the width and height of an image, respectively.

To determine an appropriate dividing number, various dividing numbers with different databases were performed. In ORL and Yale databases, we selected half of the face images (5) as the training set, and the other half (5) as the testing set for each person. Thus, there are 200 training samples and 200 testing samples selected from the ORL database. 75 training samples and 75 testing samples selected from Yale database. The recognition rates under different dividing numbers in ORL and Yale database are shown in Table 1. In

Table 1. The Recognition Rates under Different Dividing Numbers in ORL and Yale Databases.

ORL				Yale			
# of dividing	# of training sample	Size of subimage	Accuracy rate	# of dividing	# of training sample	Size of subimage	Accuracy rate
2 x 2	800	2576	26.5%	6 x 6	2700	2124	85.33%
3 x 3	1800	1178	51%	8 x 8	4800	1240	92%
4 x 4	3200	644	82.5%	10x10	7500	800	93.33%
5 x 5	5000	437	89%	12x12	10800	567	92%
6 x 6	7200	304	93%	14x14	14700	414	94.67%
7 x 7	9800	224	95%	16x16	19200	320	96%
8 x 8	12800	168	94.5%	18x18	24300	252	96%
9 x 9	16200	143	94.5%	20x20	30000	208	94.67%
10x10	20000	120	96.5%	22x22	36300	180	94.67%

Table 2. The Recognition Rates under Different Dividing Numbers in JAFFE and MIT-CBCL Databases.

JAFFE				MIT-CBCL			
# of dividing	# of training sample	Size of subimage	Accuracy rate	# of dividing	# of training sample	Size of subimage	Accuracy rate
6 x 6	108	1849	90.59%	4 x 4	1600	2500	60.99%
8 x 8	192	1024	92.94%	6 x 6	3600	1156	75.80%
10x10	300	676	94.12%	8 x 8	6400	625	95.80%
12x12	432	484	88.82%	10x10	10000	400	92.41%
14x14	588	361	91.18%	12x12	14400	289	92.22%
16x16	768	256	90%	14x14	19600	225	96.91%
18x18	972	225	90.59%	16x16	25600	169	98.58%
20x20	1200	169	90.59%	18x18	32400	144	99.01%
22x22	1452	144	88.24%	20x20	40000	100	99.81%

JAFFE database, for each person, the three neutral images were selected as the training set, and the other 17 facial images were used for testing. In MIT-CBCL database, for each person, we selected 100 images with the same illumination as the training set, and the other 162 images were used for testing. The recognition rates under different dividing numbers in JAFFE and MIT-CBCL database are shown in Table 2.

From Table 1, we obtained the highest recognition rate of 96.5% in ORL database, when the gradient image is divided into 10×10 subimages. The worst recognition rate of 26.5% in ORL database occurred when the gradient image is divided into 2×2 subimages. In Yale database, we achieved the highest recognition rate of 96% when the dividing number is 16×16 and 18×18. From Table 2, we obtained the highest recognition rate of 94.12% in JAFFE database, when the gradient image is divided into 10×10 subimages. In MIT-CBCL database, we achieved the highest recognition rate of 99.81% when the dividing number is 22×22. Observing Table 1 and Table 2, the recognition rate is almost increasing with the increase of the dividing numbers.

Experiment II: The recognition rates with different preprocessing in various databases

Table 3. The Recognition Rates with Different Preprocessing in Various Databases

	ORL	Yale	JAFFE	MIT-CBCL
Original	96.5%	81.33%	82.35%	84.69%
Histogram Equalization	95.5%	89.33%	92.35%	91.15%
Gradient	96.5%	97.33%	94.12%	99.81%

Table 4. Accuracy of Face Recognition with Different Methods

Method \ Database	Accuracy rate	
	ORL	Yale
PCA	76%	79.6%
LDA	80%	91%
Fuzzy fisherface[8]	95.5%	94.8%
ESTM[11]	78.3%	88.7%
Proposed method	96.5%	96%

Illumination variation affects the accuracy of face recognition seriously. In order to reduce the influence of illumination, most of the face recognition methods include preprocessing before face recognition is performed. To demonstrate the benefits of including the preprocessing, the recognition rate of the proposed gradient image was compared with the original face image and the face image after histogram equalization in different databases. The comparison results are shown in Table 3.

Obviously, both histogram equalization and the proposed gradient preprocessing have a great effect to reduce the influence of illumination and thus increase the recognition rates. Compared with the histogram equalization, the proposed gradient image achieved better performance.

Experiment III: Face recognition comparison with different methods and different databases

To show the proposed method has capability of face recognition with good performance. PCA, LDA, Fuzzy fisherface[11], ESTM[14], and the proposed method were compared. The recognition rates are summarized in Table 4.

According to Table 4, the proposed method achieved the highest recognition rate of 96.5% in the ORL and 96% in the Yale databases, respectively. The ESTM[14] method is a template-based classification method, the recognition rate for the ORL database and the Yale database is 78.3% and 88.7%, respectively. Because template-based approaches are based on similarity measurement of two feature sets without consideration of invariant salient feature, the recognition results highly depended on the variations of scale, pose, and shape. Even a slight variation in head pose significantly affected the similarity measurements. The PCA, LDA, and Fuzzy fisher face methods are statistics-based approach that had a higher recognition rate in situations with different head poses and facial expressions (ORL database). However, the statistic-based method is sensitive to illumination variance (Yale database). Among the statistics-based methods, the Fuzzy fisherface obtained the highest recognition rate of 95.5% in ORL database and 94.8% in Yale database, respectively.

5. **Conclusions.** In general, variation in illumination seriously affects the recognition results; hence in this paper, the original facial images were transformed to gradient images to reduce the influence of illumination variation.

To solve the SSS problem of the traditional LDA, a modified LDA (block LDA) is proposed. The gradient images were dividing into several subimages of the same size without overlapping to increase the quantity of training samples, and reduce the dimensions of training sample.

Experimental results show that the proposed method substantially obtained a higher recognition rate than traditional LDA and other methods. In addition, the proposed method can attenuate the influences of illumination, facial expression, and pose variations.

Acknowledgment. This work was supported by the National Science Council, Taiwan, under grant NSC 96-2218-E- 224-007

REFERENCES

[1] S. Krinidis, and I. Pitas, Statistical analysis of human facial expressions, *Journal of Information Hiding and Multimedia Signal Processing*, vol. 1, no. 3, pp. 241-260, 2010.
[2] W. C. Hu, C. Y. Yang, D. Y. Huang, and C. H. Huang, Feature-based face detection against skin-color like backgrounds with varying illumination, *Journal of Information Hiding and Multimedia Signal Processing*, vol. 2, no. 2, pp. 123-132, 2011.
[3] C. Y. Chang, and H. R. Hsu, Application of principal component analysis to a radial basis function committee machine for face recognition, *International Journal of Innovative Computing, Information and Control*, vol. 5, no. 11, pp.4145-4154, 2009.
[4] H. Bae, and S. Kim, Real-time face detection and recognition using hybrid-information extracted from face space and facial features, *Image and Vision Computing*, vol.23, pp.1181-1191, 2005.
[5] R. Bruneli, and T. Poggio, Face recognition: feature versus templates, *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol.15, pp.1042-1052, 1993.
[6] S. Arca, P. Campadelli, and R. Lanzarotti, A face recognition system based on automatically determined facial fiducial points, *Pattern Recognition*, vol.39, pp.432-443, 2006.
[7] R. Gottumukkal, and V. K. Asari, An improved face recognition technique based on modular PCA approach, *Pattern Recognition Letter*, vol.25, pp.429-436, 2004.
[8] Z. Q. Zhao, D. S. Huang, and B. Y. Sun, Human face recognition based on multi-features using neural networks committee, *Pattern Recognition Letter*, vol.25, pp.1351-1358, 2004.
[9] M. S. Barthlett, J. R. Movellan, and T. J. Sejnowski, Face recognition by independent component analysis, *IEEE Trans. Neural Networks*, vol.13, pp.1450-1464, 2002.
[10] D. Zhou, X. Yang, N. Peng, and Y. Wang, Improved-LDA based face recognition using both facial global and local information, *Pattern Recognition Letter*, vol.27, pp.536-543, 2006.
[11] K. C. Kwaka, and W. Pedryczb, Face recognition using a fuzzy Fisherface classifier, *Pattern Recognition*, vol.38, pp.1717-1732, 2005.
[12] W. Zheng, L. Zhao, C. Zou, An efficient algorithm to solve the small sample size problem for LDA, *Pattern Recognition*, vol.37, pp.1077-1079, 2004.
[13] R. C. Gonzalez, and R. E. Woods, *Digital Image Processing* 2nd, Prentice-Hall, Inc, Upper Saddle River, New Jersey, 2002.
[14] X. Xie, and K. M. Lam, Elastic shape-texture matching for human face recognition, *Pattern Recognition*, vol.41, pp.396-405, 2008.