Neurocomputing 73 (2010) 1556–1561

Contents lists available at ScienceDirect

Neurocomputing

journal homepage: www.elsevier.com/locate/neucom

Feature extraction based on fuzzy 2DLDA

Wankou Yang [a,*], Xiaoyong Yan [b], Lei Zhang [c], Changyin Sun [a]

[a] School of Automation, Southeast University, Nanjing 210096, People's Republic of China
[b] School of Information Technology, Jinling Institute of Technology, Nanjing 210001, People's Republic of China
[c] Dept. of Computing, The Hong Kong Polytechnic University, Hong Kong

ARTICLE INFO

Available online 12 March 2010

Keywords:
Fisher
LDA
2DLDA
Fuzzy
Feature extraction
Face recognition

ABSTRACT

In the paper, fuzzy fisherface is extended to image matrix, namely, the fuzzy 2DLDA (F2DLDA). In the proposed method, we calculate the membership degree matrix by fuzzy K-nearest neighbor (FKNN), and then incorporate the membership degree into the definition of the between-class scatter matrix and the within-class scatter matrix. Finally, we get the fuzzy between-class scatter matrix and fuzzy within-class scatter matrix. In our definition of the between-class scatter matrix and within-class matrix, the fuzzy information is better used than fuzzy fisherface. Experiments on the Yale, ORL and FERET face databases show that the new method works well.

1. Introduction

Feature extraction by dimensionality reduction is an important research topic in computer vision and pattern recognition. The curse of high dimensionality is a major cause of limitations in many practical technologies, while the large quantities of features may even degrade the performances of the classifiers when the size of the training set is small compared with the number of features [1]. In the past several decades, many feature extraction methods have been proposed, and the most well-known ones are principle component analysis (PCA) and linear discriminant analysis (LDA) [2].

Un-supervised learning cannot properly model the underlying structure and characteristics of different classes. Discriminant features are often obtained by supervised learning. LDA [2] is the traditional approach to learn discriminant subspace. Unfortunately, it cannot be applied directly to small sample size (SSS) problems [3] because the within-class scatter matrix is singular. As we know, face recognition is a typical SSS problem. Many works have been reported to use LDA for face recognition. The most popular method, called fisherface, was proposed by Swets et al. [4] and Belhumeur et al. [5]. In their methods, PCA is first used to reduce the dimension of the original features space to $N-c$ (N is the number of total training samples, c is the class number), and the classical Fisher linear discriminant analysis (FLDA) is then applied to reduce the dimension to d ($d \leq c$). Since the smallest projection components are thrown away in the PCA step, some useful discriminatory information may be lost. On the other hand, the PCA step cannot guarantee the transformed within-class scatter matrix be non-singular. More discussions about PCA and LDA can be found in [6].

To solve the singularity problem, a singular value perturbation can be added to the within-class scatter matrix [7]. A more systematic method is the regularized discriminant analysis (RDA) [8]. In RDA, one tries to obtain more reliable estimates of the eigenvalues by correcting the eigenvalue distortion with a ridge-type regularization. Penalized discriminant analysis (PDA) is another regularized version of LDA [9,10]. The goals of PDA are not only to overcome the SSS problem but also to smooth the coefficients of discriminant vectors for better interpretation. The main problem of RDA and PDA is that they do not scale well. In applications such as face recognition, the dimensionality is often more than ten thousand. It is not practical for RDA and PDA to process such a large covariance matrix.

A well-known null subspace method is the LDA+PCA method [11]. When within-class scatter matrix is of full rank, LDA+PCA only calculates the maximum eigenvectors of $(S_w)^{-1} S_b$ to form the transformation matrix. Otherwise, a two-stage procedure is employed. First, the data are transformed into the null space V_0 of S_w. Second, it maximizes the between-class scatter in V_0. LDA+PCA could be sub-optimal because it maximizes the between-class scatter in the null space of S_w instead of the original input space. Direct LDA is another null space method that discards the null space of S_b [12]. It is achieved by diagonalizing first S_b and then S_w, which is in the reverse order of conventional simultaneous diagonalization procedure. If S_t, instead of S_w, is used in direct LDA, it is actually equivalent to the PCA+LDA. Gao et al. [13] proposed a singular value decomposition (SVD) based LDA approach to solving the single training sample per person problem for face recognition. Dai et al. [14,15] developed an

* Corresponding author.
E-mail address: wankou_yang@yahoo.com.cn (W. Yang).

0925-2312/$ - see front matter © 2010 Elsevier B.V. All rights reserved.
doi:10.1016/j.neucom.2009.12.025

inverse Fisher discriminant analysis (IFDA) method. They modified the procedure of PCA and derived the regular and irregular information from the within-class scatter matrix by a new criterion called inverse Fisher discriminant criterion. Jin et al. [16] proposed the uncorrelated optimal discrimination vectors (UODV) approach which maximizes Fisher criterion simultaneously. Tao et al. [17] proposed to maximize the geometric mean of all pairs of Kullback–Leibler (KL) divergences for subspace selection, which maximizing the geometric mean of KL divergences between different class pairs.

The above-mentioned methods need to transform the 2D images into 1D vectors. This often leads to the so-called "curse of dimensionality" problem, which is often encountered in SSS cases such as face recognition. The matrix-to-vector transform may also cause the loss of some useful structural information embedding in the original images. To overcome the problems, Yang et al. [18] proposed the 2-dimensional principal analysis (2DPCA). 2DPCA is based on 2D image matrices rather than 1D vectors. That is, the image matrix does not need to be transformed into a vector. Instead, the image covariance matrix can be constructed directly from the image matrices, and its eigenvectors are derived for image feature extraction. In contrast to PCA, the size of covariance matrix using 2DPCA is much smaller. As a result, 2DPCA computes the corresponding eigenvectors more quickly than PCA. Inspired by the successful application of 2DPCA to face recognition, 2DLDA was proposed [19–22]. Recently, Zheng et al. [23] investigated the relations between vector-based linear discriminant analysis and matrix-based discriminant analysis. They pointed out that from the bias estimation point of view, 2DLDA might be more stable than 1DLDA. More recently, a method based on the local geometrical structure called tensor subspace analysis (TSA) [24] was proposed, which captures an optimal linear approximation to the face manifold in the sense of local isometry. Tao et al. [25,26] proposed a tensor discriminant analysis method for feature extraction. They proposed a convergent solution to discriminative tensor subspace selection.

Regretfully, 2DLDA assumes the same level of typicality of each face to the corresponding class. In this paper we propose to incorporate a gradual level of assignment to the class being regarded as a membership grade with anticipation that such discrimination helps to improve the classification results. More specifically, when operating on feature vectors resulting from PCA transformation we complete a fuzzy K-nearest neighbor class assignment that produces the corresponding degree of class membership. By taking advantage of the technology of fuzzy sets [27], a number of studies have been carried out for fuzzy pattern recognition [28–31].

The organization of this paper is as follows: In Section 2, we briefly review 2DLDA and Fuzzy fisherface. In Section 3, we propose the idea and describe the new method in detail. In Section 4, experiments on face image databases are presented to demonstrate the effectiveness of the new method. Conclusions are summarized in Section 5.

2. Related works

2.1. 2DLDA

Training is a process of acquiring features from available training images and storing them in a knowledge base for the purpose of recognizing an input image. Given a set of samples of each class, the 2DLDA extracts most informative features which could establish a high degree of similarity between samples of the same class and a high degree of dissimilarity between samples of two classes. Suppose there are c known pattern classes $w_1, w_2, \ldots, w_c$ and N training samples. $X=\{X_j^i\}$ ($i=1,2,\ldots,l_c$, $j=1,2,\ldots,c$) is a set of samples with $m\times n$ dimension. l_j is the number of training samples of class j and satisfies $\sum_{i=1}^{c} l_i = N$. Set $\overline{X_j}=(1/l_j)\sum_{i=1}^{l_j} X_j^i$, $\overline{X}=(1/N)\sum_{j=1}^{c}\sum_{i=1}^{l_j} X_j^i$, $j=1,2,\ldots,c$. The image between-class scatter matrix G_b and the image within-class scatter matrix G_w are computed as

$$G_b=\frac{1}{N}\sum_{j=1}^{c} l_j(\overline{X_j}-\overline{X})(\overline{X_j}-\overline{X})^T \tag{1}$$

$$G_w=\frac{1}{N}\sum_{j=1}^{c}\sum_{i=1}^{l_j}(X_j^i-\overline{X_j})(X_j^i-\overline{X_j})^T \tag{2}$$

Once G_b and G_w are computed, it is recommended to find the optimal projection axis W so that the total scatter of the projected samples of the training images is maximized. The objective function of 2DLDA can be defined as

$$J(W)=\underset{W}{\operatorname{argmax}}\frac{W^TG_bW}{W^TG_wW} \tag{3}$$

It can be proven that the eigenvector corresponding to the maximum eigenvalue of $(G_w)^{-1}G_b$ is the optimal projection vectors which maximizes $J(W)$. Generally, as it is not enough to have only one optimal projection vector, we usually look for d projection axes, say $w_1, w_2, \ldots, w_d$, which are the eigenvectors corresponding to the first d largest eigenvalues of $(G_w)^{-1}G_b$. In 2DLDA, once these projection vectors are computed, each training image X_j^i is then projected onto W to obtain the feature matrix Y_j^i of size $m\times d$ of the training image X_j^i. So, during training, for each training image X_j^i a corresponding feature matrix of size $m\times d$ is constructed and stored for matching at the time of recognition.

2.2. Fuzzy fisherface

K.C-Kwak [28] proposed the fuzzy fisherface for face recognition via fuzzy set. Given a set of feature vectors $X=\{x_1, x_2, \ldots, x_N,\}$, c known pattern classes $w_1, w_2, \ldots, w_c$, l_j, is the number of training samples of class j and satisfies $\sum_{i=1}^{c} l_i=N$, $m=1/N\sum_{i=1}^{N} x_i$. A fuzzy "c"-class partition of these vectors specifies the degree of membership of each vector to the classes. The membership matrix $[u_{ij}](i=1,2,\ldots,c,\quad j=1,2,\ldots,N)$ can be got by FKNN, it will be discussed in section 3.1 in detail. Taking into account the membership grades, the mean vector of each class m_i is calculated as follows:

$$m_i=\frac{\sum_{j=1}^{N} u_{ij}x_j}{\sum_{j=1}^{N} u_{ij}} \tag{4}$$

The between-class fuzzy scatter matrix S_{fb} and within-class fuzzy scatter matrix S_{fw} incorporate the membership values in their calculations

$$S_{fb}=\sum_{i=1}^{c} l_i(m_i-m)(m_i-m)^T \tag{5}$$

$$S_{fw}=\sum_{i=1}^{c}\sum_{j=1}^{l_i}(x_i^j-m_i)(x_i^j-m_i)^T \tag{6}$$

The optimal fuzzy projection matrix W of fuzzy fisherface follows the expression:

$$W=\underset{W}{\operatorname{argmax}}\frac{W^TS_{fb}W}{W^TS_{fw}W} \tag{7}$$

Finally, PCA plus fuzzy FLD is used in SSS cases.

3. Fuzzy 2DLDA

By analyzing 2DLDA, we note that the algorithm dwells on the concept of a binary class assignment. Since the faces are significantly affected by numerous environmental conditions (e.g. illumination, poses, etc.), it is advantageous to investigate these factors and quantify their impact on their "internal" (i.e. algorithm-driven) class assignment. The purpose is to reflect all these factors in a "soft" viz. fuzzy class allocation to the individual face under consideration. Actually, the idea of such "fuzzification" of class assignment has been used by Keller et al. [28] under the notion of fuzzy k-nearest neighbor classifier. Inspired by the fuzzy fisherface, we expect that fuzzy set theory can be used effectively to enhance the performance of 2DLDA. In the following, we formulate the new fuzzy 2DLDA (F2DLDA) algorithm, which makes fully use of the distribution information of samples. Specifically, the sample distribution information is represented by fuzzy membership degree corresponding to each class.

3.1. Fuzzy K-nearest neighbor (FKNN)

In our method, fuzzy membership degree and class centers are obtained through the FKNN [28] algorithm. With FKNN, the computations of the membership degree can be realized through the following steps:

Step1: Compute the Euclidean distance matrix between pairs of feature vectors in the training set.

Step2: Set the diagonal elements of the Euclidean distance matrix to infinity.

Step3: Sort the distance matrix (treat each of its columns separately) in an ascending order. Collect the corresponding class labels of the patterns located in the closest neighborhood of the pattern under consideration (as we are concerned with 'k' neighbors, this returns a list of 'k' integers).

Step4: Compute the membership degree to class 'i' for the jth pattern using the method proposed in the literature [28]

$$u_{ij} = \begin{cases} 0.51 + 0.49 \times (n_{ij}/k) & \text{if } i = \text{the label of the } j\text{th pattern} \\ 0.49 \times (n_{ij}/k) & \text{if } i \neq \text{the label of the } j\text{th pattern} \end{cases} \tag{8}$$

In the above equation, n_{ij} stands for the number of the neighbors of the jth pattern belongs to the ith class.

Finally, the fuzzy membership matrix U can be obtained with the output of FKNN

$$U = [u_{ij}], \quad i = 1,2,\cdots,c, \quad j = 1,2,\ldots,N \tag{9}$$

3.2. The idea of fuzzy 2DLDA

The key step of Fuzzy 2DLDA is how to incorporate the contribution of each training sample into the scatter matrices. Based on the fuzzy set theory, each sample can be classified into multi-classes with fuzzy membership degrees, instead of binary classification. In the redefinition of the fuzzy within-class scatter matrix, samples that are more close to class center have more contribution to classification. In the redefinition of the between-class scatter matrix, class which is far from the total center will have more contribution to classification. Then, the membership degree of each sample (contribution to each class) should be considered and the mean matrix of each class, the corresponding fuzzy within-class scatter matrix G_{fw} and fuzzy between-class scatter matrix G_{fb} can be redefined as follows:

$$\overline{A_i} = \frac{\sum_{j=1}^{N} u_{ij}^p A_j}{\sum_{j=1}^{N} u_{ij}^p} \tag{10}$$

$$G_{fw} = \sum_{i=1}^{c} \sum_{j=1}^{l_i} u_{ij}^p (A_i^j - \overline{A_i})(A_i^j - \overline{A_i})^T \tag{11}$$

$$G_{fb} = \sum_{i=1}^{c} \sum_{j=1}^{N} u_{ij}^p (\overline{A_i} - \overline{A})(\overline{A_i} - \overline{A})^T \tag{12}$$

where p is a constant which controls the influence of fuzzy membership degree,$\overline{A}$ is the mean of all samples. Similarly, the fuzzy total scatter matrix G_{ft} can be achieved as follows:

$$G_{ft} = G_{fw} + G_{fb} \tag{13}$$

Thus the scatter matrices with fuzzy set theory are redefined. From Eqs. (4)–(6) and Eqs. (10)–(13), we can find that our proposed method can make better use of the distribution of samples. The membership degree is fully used in the construction of the fuzzy within-class scatter matrix and the fuzzy between-class matrix, while the membership degree is only used in the calculation of the class mean in fuzzy fisherface.

It is easy to show that G_{fw} and G_{fb} are $m \times m$ matrices and they are in general non-singular. The objective function of Fuzzy 2DLDA can be redefined as

$$J(W) = \underset{W}{\arg\max} \frac{W^T G_{fb} W}{W^T G_{fw} W} \tag{14}$$

This criterion is in a form of Rayleigh quotient, and the optimal solution can be obtained by solving a generalized eigen-equation.

3.3. The algorithm of fuzzy 2DLDA

Based on the above descriptions, the proposed fuzzy 2DLDA (F2DLDA) algorithm can be summarized as follows:

Step 1 (FKNN): The class center matrix $\overline{A_i}$ and the fuzzy membership degree matrix U can be computed with the FKNN algorithm in the original image space.

Step 2 (Fuzzy 2DLDA): According to $\overline{A_i}$ and U, compute the fuzzy within-class scatter matrix G_{fw} and fuzzy between-class scatter matrix S_{fb}. The optimization problem in Eq. (10) can be solved by $(G_{fw})^{-1} G_{fb} u = \lambda u$ with eigenvalues $\lambda_1 > \cdots > \lambda_q > 0$ and normalized eigenvectors $u_1, u_2, \ldots, u_q$. Then the optimal projection matrix can be obtained.

Step 3 (recognition): Project all samples into the obtained optimal discriminant matrix and classify.

4. Experiments

Three face image databases, namely, the Yale database, the ORL database and the FERET database, are used to compare the proposed fuzzy 2DLDA (F2DLDA) approach with the following algorithms: PCA (eigenface) [5], 2DPCA [18], LDA (Fisherface) [5], 2DLDA [23], fuzzy fisherface [29], and LPP [32].

4.1. Experiments on the Yale database

The Yale face database contains 165 images of 15 individuals (each person providing 11 different images) under various facial expressions and lighting conditions. In our experiments, each image was manually cropped and resized to 100×80. Fig. 1 shows seven sample images of one person.

Fig. 1. Eleven images of one person in Yale.

Table 1
Maximal recognition rate on the Yale database.

Method	PCA	2DPCA	LDA	2DLDA	Fuzzy fisherface	LPP	CFLDA	**Proposed**
Result	0.8533	0.9200	0.9333	0.9333	0.8533	0.8267	0.9467	**0.9600**
Dimension	37	100×11	14	100×3	14	37	36	100×3

Table 2
Average recognition comparison on Yale database.

Method	PCA	2DPCA	LDA	2DLDA	Fuzzy fisherface	LPP	CFLDA	**Proposed**
Ex	0.9027	0.9240	0.9520	0.9533	0.9307	0.9187	0.9547	**0.9633**
Std	0.0576	0.0321	0.0566	0.0144	0.0711	0.0481	0.0423	**0.0261**

Fig. 2. Ten images of one person in ORL

The first experiment was performed using the first six images (i.e., center-light, with glasses, happy, left-light, without glasses, and normal) per class for training, and the remaining five images (i.e., right-light, sad, sleepy, surprised, and winking) for testing. For feature extraction, we used, respectively, PCA, 2DPCA, LDA, 2DLDA, fuzzy fisherface, LPP, and the proposed F2DLDA. Note that LDA, LPP ($t=+\infty$), CFLDA, and fuzzy fisherface involve a PCA phase. In this phase, we keep nearly 95 percent image energy and select the number of principal components, m, as 37. In the second phase of LDA and fisherface, the number of discriminant vectors corresponding to the c–1 largest generalized eigenvalues is 14. The FKNN parameter K is set as $K=l$–$1=5$, where l denotes the number of training samples per class. The justification for this choice is that each sample should have l–1 samples of the same class provided that within-class samples are well clustered. Finally, the nearest neighbor (NN) classifier with cosine distance is employed for classification. The maximal recognition rate of each method and the corresponding dimensions are given in Table 1.

From Table 1, we can see that F2DLDA outperforms other methods. The reason is that the overlapping sample's distribution information is incorporated in the redefinition of scatter matrices by fuzzy set theory, which is important for classification, and the image structural information in considered in the 2D methods.

In the second experiment, 10-fold cross-validation tests are performed to re-evaluate the performance of PCA, 2DPCA, LDA, 2DLDA, fuzzy fisherface, LPP ($t=+\infty$), CFLDA, and F2DLDA. In each test, six images of each subject are randomly chosen for training, while the remaining five images are used for testing. In the second experiment, we select the same dimension and parameters as in the first experiment. Table 2 shows the maximal average recognition rates across 10 runs of each method under nearest neighbor classifier with cosine distance metrics. The corresponding standard deviations (std) are also listed. From Table 2, it can be seen that F2DLDA outperforms other methods.

Table 3
Average recognition rate on ORL database.

Method	s	Class	Dim	Mean/Std
PCA	3	40	74	0.8918/0.0189
2DPCA	3	40	112×4	0.9004/0.0261
LDA	3	40	39	0.9061/0.0211
2DLDA	3	40	112×3	0.9161/0.0193
Fuzzy fisherface	3	40-	39	0.8982/0.0269
LPP	3	40	74	0.7575/0.0299
CFLDA	3	40	39	0.8414/0.0281
Proposed	**3**	**40**	$\mathbf{112 \times 3}$	**0.9208/0.0217**
PCA	4	40	93	0.9375/0.00224
2DPCA	4	40	112×4	0.9354/0.0234
LDA	4	40	39	0.9512/0.0195
2DLDA	4	40	112×3	0.9417/0.0127
Fuzzy fisherface	4	40	39	0.9492/0.0220
LPP	4	40	93	0.8350/0.0258
CFLDA	4	40	39	0.9129/0.0256
Proposed	**4**	**40**	$\mathbf{112 \times 3}$	**0.9604/0.0194**

In all the experiments, PCA is used to extract 66 principal component features and 2DPCA is used to extract 112×11 principal component feature vectors. LDA is used to extract 14

Fig. 3. Images of one person in FERET.

Table 4
Recognition rate on FERET.

Method	PCA	2DPCA	LDA	2DLDA	Fuzzy fisherface	LPP	CFLDA	**Proposed**
Ex	0.3662	0.4792	0.4936	0.4388	0.4914	0.3584	0.4515	**0.5288**
Std	0.0688	0.0864	0.1012	0.0503	0.0938	0.0755	0.0835	**0.0605**
Dim	99	40×8	99	40×8	99	99	114	**40×8**

LDA features, 2DLDA, and F2DLDA are used to extract 112×3 feature vectors.

4.2. Experiments on the ORL database

The ORL (http://www.cam-orl.co.uk) database contains 40 persons, each having 10 different images. The images of the same person are taken at different times, under slightly varying lighting conditions and with various facial expressions. Some people are captured with or without glasses. The heads in images are slightly titled or rotated. The images in the database are manually cropped and rescaled to 112×92. Fig. 2 shows ten images of one person in ORL.

In the experiments, we split the whole database into two parts evenly. One part is used for training and the other part is for testing. In order to make full use of the available data and to evaluate the generalization power of algorithms more accurately, we adopt a cross-validation strategy and run the system 10 times. In each time, s face images from each person are randomly selected as training samples, and the rest is for testing. The classical PCA, 2DPCA, LDA, 2DLDA, LPP ($t=+\infty$), CFLDA, and fuzzy fisherface and the proposed F2DLDA are, respectively, used for feature extraction. In the PCA stage of LDA, LPP, and fuzzy fisherface, we keep nearly 95 percent image energy and the number of principal components, m, is set as 61 and 73. The FKNN parameter K is set as $K=S-1$. Finally a nearest neighbor classifier with cosine distance is employed. The recognition results are shown in Table 3. From Table 3, we find that our method all outperforms other methods and our method can work well.

4.3. Experiments on the FERET database

The FERET face database is a result of the FERET program, which was sponsored by the US Department of Defense through the DARPA Program [33,34]. It has become a standard database for testing and evaluating state-of-the-art face recognition algorithms. The proposed method was tested on a subset of the FERET database. This subset includes 1400 images of 200 individuals (each individual has seven images). This subset involves variations in facial expression, illumination, and pose. In our experiment, the facial portion of each original images was automatically cropped based on the location of the eyes and the cropped images was resized to 40×40 pixels. Some example images of one person are shown in Fig. 3.

In order to make full use of the available data and to evaluate the generalization power of algorithms more accurately, we adopt a cross-validation strategy and run the system 10 times. In each time, 2 face images from each person are randomly selected as training samples, and the rest is for testing. The classical PCA, LDA, 2DLDA, fuzzy fisherface, LPP ($t=+\infty$), CFLDA, and the proposed F2DLDA are, respectively, used for feature extraction. The FKNN parameter K is set as $K=2$. Finally a nearest neighbor classifier with cosine distance is employed. In the experiments, in the PCA stage of LDA and fuzzy fisherface, we keep nearly 95 percent image energy and the number of principal components, m, is set as 99. PCA is used to extract 86 principal component features, 2DLDA and F2DLDA are used to extract 40×4 feature vectors. The images in the FERET face database are subject to complex nonlinear changes due to large pose, expression or illumination variations. From Table 4, we can see that our proposed method can work well in the complex circumstance.

5. Conclusions

In this paper, we proposed a new method for feature extraction, namely the fuzzy 2DLDA (F2DLDA). This method is based on LDA, image matrix and the fuzzy set theory. First, the structural information is considered in our method; second, the overlapping sample's distribution information is fully incorporated in the redefinition of corresponding scatter matrices, which is important for classification. Experiments on the Yale, the ORL, and FERET face databases showed that the new method works effectively.

Acknowledgements

This project is supported by NSFC of China (Grant nos.: 90820009, 60632050, 60803049, 60875010) and the Hong Kong RGC General Research Grant (PolyU 5351/08E).

Reference

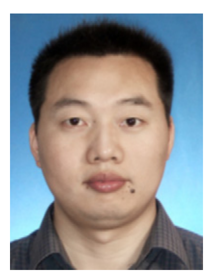
[1] A.K. Jain, B. Chandrasekaran, Dimensionality and sample size considerations in pattern recognition practice, In Handbook of Statistics. Amsterdam, North Holland, 1982.
[2] R.O. Duda, P.E. Hart, Pattern Classification and Scene Analysis, New York, Wiley, 1973.
[3] S.J. Raudys, A.K. Jain, Small sample size effects in statistical pattern recognition: recommendations for practitioners, IEEE Transactions on Pattern Analysis and Machine Intelligence 13 (3) (1991) 252–264.
[4] D.L. Swets, J. Weng, Using discriminant eigenfeatures for image retrieval, IEEE Transactions on Pattern Analysis and Machine Intelligence 18 (8) (1996) 831–836.
[5] V. Belhumeur, J. Hespanha, D. Kriegman, Eigenfaces vs. fisherfaces: recognition using class specific linear projection, IEEE Transactions on Pattern Analysis and Machine Intelligence 19 (7) (1997) 711–720.
[6] J. Yang, J.Y. Yang, Why can LDA be performed in PCA transformed space? Pattern Recognition 36 (2) (2003) 563–566.

[7] Z.Q. Hong, J.Y. Yang, Optimal discriminant plane for a small number of samples and design method of classifier on the plane, Pattern Recognition 24 (4) (1991) 317–324.
[8] J.H. Friedman, Regularized discriminant analysis, Journal of the American Statistical Association 84 (1989) 165–175.
[9] T. Hastie, R. Tibshirani, Penalized discriminant analysis, The Annals of Statistics 23 (1995) 73–102.
[10] T. Hastie, R. Tibshirani, A. Buja, Flexible discriminant analysis by optimal scoring, Journal of the American Statistical Association 89 (1994) 1255–1270.
[11] L.F. Chen, H.Y.M. Liao, M.T. Ko, G.J. Yu, A new LDA-based face recognition system which can solve the small sample size problem, Pattern Recognition 33 (1) (2000) 1713–1726.
[12] H. Yu, J. Yang, A direct LDA algorithm for high dimensional data–with application to face recognition, Pattern Recognition 34 (10) (2001) 2067–2070.
[13] Q.X. Gao, L. Zhang, D. Zhang, Face recognition using FLDA with single training image per-person, Applied Mathematics and Computation 205 (12) (2008) 726–734.
[14] X.S. Zhuang, D.Q. Dai, Inverse fisher discriminant criteria for small sample size problem and its application to face recognition, Pattern Recognition 38 (11) (2005) 2192–2194.
[15] X.S. Zhuang, D.Q. Dai, Improved discriminant analysis for high-dimensional data and its application to face recognition, Pattern Recognition 40 (5) (2007) 1570–1578.
[16] Z. Jin, J.Y. Yang, Z. Hu, Z. Lou, Face recognition based on the uncorrelated discrimination Transformation, Pattern Recognition 34 (7) (2001) 1405–1416.
[17] D. Tao, X. Li, X. Wu, and S. J. Maybank, Geometric mean for subspace selection in multiclass classification, IEEE Transactions on Pattern Analysis and Machine Intelligence.
[18] J. Yang, D. Zhang, A.F. Frangi, J.Y. Yang, Two-dimensional PCA: a new approach to appearance-based face representation and recognition, IEEE Transactions on Pattern Analysis and Machine Intelligence 26 (12) (2004) 131–137.
[19] M. Li, B. Yuan, 2D-LDA: a statistical linear discriminant analysis for image matrix, Pattern Recognition Letters 26 (2005) 527–532.
[20] J. Yang, D. Zhang, Y. Xu, J.J. Yang, Two-dimensional discriminant transform for face recognition, Pattern Recognition 38 (2005) 1125–1129.
[21] H.L. Xiong, M.N.S. Swanmy, M.O. Ahmad, Two-dimensional FLD for face recognition, Pattern Recognition 38 (2005) 1121–1124.
[22] X.Y. Jing, H.S. Wong, D. Zhang, Face recognition based on 2D fisherface approach, Pattern Recognition 39 (2006) 707–710.
[23] W.S. Zheng, J.H. Lai, Stan Z. Li, 1D-LDA vs. 2D-LDA: when is vector-based linear discriminant analysis better than matrix-based, Pattern Recognition 41 (7) (2008) 2156–2172.
[24] X.F. He, D. Cai, P. Niyogi, Tensor Subspace Analysis. Advances in Neural Information Proceeding Systems,, 18, Vancouver, Canada, 2005 December.
[25] D.C. Tao, X.L. Li, X.D. Wu, et al., General tensor discriminant analysis and gabor features for gait recognition, IEEE Transactions on Pattern Analysis and Machine Intelligence 29 (10) (2007) 1700–1715.
[26] D.C. Tao, X.L. Li, W.M. Hu, et al., Supervised tensor learning, Knowledge and Information Systems (Springer: KAIS) 13 (1) (2007) 1–42.
[27] L.A. Zadeh, Fuzzy sets, Information and Control 8 (1965) 338–353.
[28] J.M. Keller, M.R. Gray, J.A. Givern, A fuzzy k-nearest neighbour algorithm, IEEE Transactions on Systems, Man, and Cybernetics 15 (4) (1985) 580–585.
[29] K.C. Kw, W. Pedry, Face recognition using a fuzzy fisher classifier, Pattern Recognition 38 (10) (2005) 1717–1732.
[30] Y.J. Zheng, J.Y. Yang, et al., Fuzzy kernel Fisher discriminant algorithm with application to face recognition. The Sixth World Congress on Intelligent and Automation (WCICA06), 12 (12): 9669-9672.
[31] W. Yang, H. Yan, J. Wang, J. Yang, Face recognition using complete fuzzy LDA, ICPR, 2008.
[32] X. He, S. Yan, Y. Niyogi Hu, H. Zhang, Face recognition using Laplacianfaces, IEEE Transactions on Pattern Analysis and Machine Intelligence 27 (3) (2005) 328–340.
[33] P.J. Phillips, H. Moon, S.A. Rizvi, P.J. Rauss, The FERET evaluation methodology for face-recognition algorithms, IEEE Transactions on Pattern Analysis and Machine Intelligence 22 (10) (2000) 1090–1104.
[34] P.J. Phillips, The facial recognition technology (FERET) database. 〈http://www.itl.nist.gov/iad/humanid/feret/feret_master.html〉, 2004.

Wankou Yang received his B.S., M.S. and Ph.D. degrees at the School of Computer Science and Technology, Nanjing University of Science and Technology (NUST), China, 2002, 2004 and 2009, respectively. Now he is a Postdoctoral Fellow at School of Automation, Southeast University, China. His research interests include pattern recognition, computer vision, digital image processing.

Xiaoyong Yan is a lecture at the School of Information Technology, Jinling Insitute of Technology, China. He received the B.S. degree in mechanical electronic engineering from Nanjing Agricultural University in 2000 and received the MS Degree in computer application technology from Nanjing University of Science and Technology in 2006. He is now working towards his doctor's degree at School of Computer Science and Technology, Nanjing University of Science and Technology (NUST), China. His current interests are in the areas of wireless sensor networks, pattern recognition and image processing.

Lei Zhang received the B.S. degree in 1995 from Shenyang Institute of Aeronautical Engineering, Shenyang, P.R. China, the M.S. and Ph.D degrees in Electrical and Engineering from Northwestern Polytechnical University, Xi'an, P.R. China, respectively, in 1998 and 2001. From 2001 to 2002, he was a research associate in the Dept. of Computing, The Hong Kong Polytechnic University. From January 2003 to January 2006 he worked as a Postdoctoral Fellow in the Dept. of Electrical and Computer Engineering, McMaster University, Canada. Since January 2006, he has been an Assistant Professor in the Dept. of Computing, The Hong Kong Polytechnic University. His research interests include Image and Video Processing, Biometrics, Pattern Recognition, Multisensor Data Fusion and Optimal Estimation Theory, etc.



ChangYin Sun received the B.S. degree in 1996 from Department of Mathematics, Sichuan University, Chengdu, China, the M.S. and Ph.D degrees in Electrical Engineering from Southeast University, Nanjing, China, respectively in 2001 and 2003. From March 2004 to September 2004, he was a Postdoctoral Fellow in the Dept. of Computer Science, The Chinese University of Hong Kong. From April 2006 to March 2007 he worked as a Professor in the School of Electrical Engineering, Hohai University, Nanjing, China; Form July 2007 to September 2009, he worked as a Program Director in Division of Division of Information 3, Department of Information Sciences, National Natural Science Foundation of China, Beijing China; From March 2007, he works as a professor in the School of Automation, Southeast University, China. He is an associate Editor of IEEE Transactions on Neural Networks (2010–), Neural Processing Letters (2008–), International Journal of Swarm Intelligence Research (2010–) and Recent Parents of Computer Science (2008–). His research interests include Intelligent Control, Neural Networks, Support Vector Machine, Data-Driven Control; Theory and Design of Intelligent Control Systems, Optimization Algorithms; Pattern Recognition, etc.