Neurocomputing 72 (2009) 3384–3390

Contents lists available at ScienceDirect

Neurocomputing

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

Letters

Feature extraction using fuzzy inverse FDA

Wankou Yang [a,c,*], Jianguo Wang [a,b], Mingwu Ren [a], Lei Zhang [c], Jingyu Yang [a]

[a] School of Computer Science and Technology, Nanjing University of Science and Technology, Nanjing 210094, China
[b] Tangshan College, Network and Education Center, Tangshan 063000, China
[c] Department of Computing, The Hong Kong Polytechnic University, Hong Kong, China

ARTICLE INFO

Article history:
Received 17 February 2008
Received in revised form
8 March 2009
Accepted 8 March 2009
Communicated by T. Heskes
Available online 9 April 2009

Keywords:
FDA
Inverse FDA
Fuzzy IFDA
Feature extraction
Face recognition
Pulse signal recognition

ABSTRACT

This paper proposes a new method of feature extraction and recognition, namely, the fuzzy inverse Fisher discriminant analysis (FIFDA) based on the inverse Fisher discriminant criterion and fuzzy set theory. In the proposed method, a membership degree matrix is calculated using FKNN, then the membership degree is incorporated into the definition of the between-class scatter matrix and within-class scatter matrix to get the fuzzy between-class scatter matrix and fuzzy within-class scatter matrix. Experimental results on the ORL, FERET face databases and pulse signal database show that the new method outperforms Fisherface, fuzzy Fisherface and inverse Fisher discriminant analysis.

1. Introduction

It is well known that FDA is an effective feature extraction method. But, unfortunately, it cannot be applied directly to small size sample (SSS) problems [1], for in these cases, the within-class scatter matrix is singular. As we know, face recognition is a typical small size problem. In order to utilize LDA for face recognition, a number of research works have been done [2–9]. The most popular method, called Fisherface, was build by Swets et al. [2] and Belhumeur et al. [3]. In their methods, PCA is first used to reduce the dimension of the original features space to $N-c$, and the classical FLD is next applied to reduce the dimension to d ($d \leqslant c$). Obviously, in the PCA transform, the small $c-1$ projection components have been thrown away. So some effective discriminatory information may be lost. And PCA step cannot guarantee the transformed within-class scatter matrix still be not singular. Yang and Yang [4] have proven the theoretical foundation for Fisherface and proposed a linear feature extraction method. First, PCA is used to reduce the dimension of feature space to $N-1$ (N denotes the number of training samples), and then the OFLD [5] method is used for the second feature extraction. However, no procedure has been shown to determine the optimal dimensions in the OFLD.

To handle the singularity problem, it is also popular to add a singular value perturbation to within-class scatter matrix to make it nonsingular [6]. A similar but more systematic method is regularized discriminant analysis (RDA) [7]. In RDA, one tries to obtain more reliable estimates of the eigenvalues by correcting the eigenvalue distortion in the sample covariance matrix with a ridge-type regularization. Besides, RDA is also a compromise between LDA and QDA (quadratic discriminant analysis), which allows one to shrink the separate covariances of QDA towards a common covariance as in LDA. Penalized discriminant analysis (PDA) is another regularized version of LDA [8,9]. The goals of PDA are not only to overcome the small sample size problem but also to smooth the coefficients of discriminant vectors for better interpretation. The main problem with RDA and PDA is that they do not scale well. In application such as face recognition, the dimension of covariance matrices is often more than 10 000. It is not practical for RDA and PDA to process such large covariance matrices, especially, when the computing platform is made of PCs. A well-known null subspace method is the LDA+PCA method [10]. When within-class scatter matrix is of full rank, the LDA+PCA method just 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 null space V_0 of S_w. Second, it tries to maximize the between-class scatter in V_0. Although this method solves small sample size problem, it could be sub-optimal because it maximizes the between-class

* Corresponding author at: School of Computer Science and Technology, Nanjing University of Science and Technology, Nanjing 210094, China.
Tel.: (+)86 (0) 25 84317297x403.
E-mail addresses: wankou_yang@yahoo.com.cn, yangwankou@njust.edu.cn (W. Yang).

0925-2312/$ - see front matter Crown Copyright © 2009 Published by Elsevier B.V. All rights reserved.
doi:10.1016/j.neucom.2009.03.011

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 [11]. This is achieved by diagonalizing first S_b and then diagonalizing S_w, which is in the reverse order of conventional simultaneous diagonalization procedure. In Direct LDA, one may also employ S_t instead of S_w. In this way, Direct LDA is actually equivalent to the PCA+LDA. Li et al. [12] further proposed an efficient and robust linear feature extraction method which aims to maximize the following criterion $J = \mathrm{tr}(W^T(S_b - S_w)W)$, which was called maximum margin criterion (MMC). In order to avoid both the singularity and instability critical issues of the within-class scatter matrix S_w when LDA is used in limited sample and high dimensional problems, Thomaz and Gillies [13] proposed a LDA-based approach based on a straightforward covariance selection method for the S_w matrix [14].

Recently, Dai et al. [15,16] develop an inverse Fisher discriminant analysis (IFDA) method. The algorithm modifies the procedure of PCA and derives the regular and irregular information from the within-class scatter matrix by a new criterion, which is called inverse Fisher discriminant criterion. Unfortunately, the IFDA assumes that the same level of typicality of each face to the corresponding class. Inspired by fuzzy Fisherface [19], we propose to incorporate a gradual level of assignment to class being regarded as a membership grade with anticipation that such discrimination helps improve 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 [17], a number of studies have been carried out for fuzzy pattern recognition [18–21]. This paper is the improved version of [22].

The organization of this paper is as follows. In Section 2, we review briefly the related works. In Section 3, we propose the idea and describe the new method in detail. In Section 4, experiments with face images data are presented to demonstrate the effectiveness of the new method. Conclusions are summarized in Section 5.

2. Related works

2.1. FDA

Linear (Fisher) discriminant analysis (FDA) is a well-known and popular statistical method in pattern recognition and classification [23]. Suppose there are c known pattern classes $w_1, w_2, \ldots, w_c$, N training samples. $X = \{x_j^i\}$ $(i = 1, 2, \ldots, l_c,\quad j = 1, 2, \ldots, c)$ is a set of samples with d dimension. l_j is the number of training samples of class j and satisfies $\sum_{i=1}^{c} l_i = N$. Set $\overline{X_j} = (1/l_j)\sum_{i=1}^{l_j} x_j^i$, $\bar{X} = (1/N)\sum_{j=1}^{c}\sum_{i=1}^{l_j} x_j^i$. Let the between-class scatter matrix and the within-class scatter matrix be defined as

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

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

It is easy to verify that $S_t = S_w + S_b$. Now, the projection matrix W_{FDA} of FDA is chosen as a matrix with orthonormal columns maximizing the following quotient, called Fisher discriminant criterion:

$$W_{FDA} = \arg\max_W \frac{|W^T S_b W|}{|W^T S_w W|} = [w_1, w_2, \ldots, w_m] \tag{3}$$

where $\{w_i | i = 1, 2, \ldots, m\}$ is the set of generalized eigenvectors of $(S_w)^{-1} S_b$ corresponding to the m largest generalized eigenvalues $\{\lambda_i | i = 1, 2, \ldots, m\}$.

2.2. Fuzzy Fisherface

Kwak [19] proposed the fuzzy Fisherface for face recognition via fuzzy set. Given a set of feature vectors, $X = \{x_1, x_2, \ldots, x_{N,}\}$, a fuzzy "c"-class partition of these vectors specifies the degree of membership of each vector to the classes. The membership matrix $[u_{ij}](i = 1, 2, \ldots, c, j = 1, 2, \ldots, N)$ can be obtained by FKNN [18]; it will be discussed in Section 3.1 in detail. Taking into account the membership grades, the mean vector of each class $\overline{m_i}$ is calculated as follows:

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

The between-class fuzzy scatter matrix SF_b and within-class fuzzy scatter matrix SF_w incorporate the membership values in their calculations

$$SF_b = \sum_{i=1}^{c} N_i(\overline{m_i} - \overline{X})(\overline{m_i} - \overline{X})^T \tag{5}$$

$$SF_w = \sum_{i=1}^{c}\sum_{x_k \in w_i}(x_k - \overline{m_i})(x_k - \overline{m_i})^T \tag{6}$$

The optimal fuzzy projection $W_{F\text{-}FDA}$ follows the expression:

$$W_{F\text{-}FDA} = \arg\max_W \frac{|W^T SF_b W|}{|W^T SF_w W|} \tag{7}$$

Finally, Kwak gave the strategy: PCA plus fuzzy FDA in small sample size case.

2.3. Inverse FDA

From the inverse relationship

$$\arg\max_W \frac{|W^T S_b W|}{|W^T S_w W|} \Leftrightarrow \arg\min_W \frac{|W^T S_w W|}{|W^T S_b W|} \tag{8}$$

Without considering the singularity, the inverse Fisher discriminant criterion, is

$$W_{IFDA} = \arg\min_W \frac{|W^T S_w W|}{|W^T S_b W|} \tag{9}$$

In contrast with LDA, the program using the above inverse Fisher discriminant criterion is called inverse Fisher discriminant analysis [15,16]. Obviously, the Fisher criterion (3) and inverse Fisher criterion (9) are equivalent, provided that the within-class scatter matrix S_w and the between-class scatter S_b are not singular. However, $rank(S_b) \leqslant c-1$. Thus, the difficulty of SSS problem still exists for IFDA. So the PCA can be exploited to reduce the dimension of the original feature space. First, we apply PCA procedure to lower the dimension from d to d' and get a projection matrix $W_{PCA\_S} \in R^{d \times d'}$ $(W_{PCA\_S} = [u_{i1}, u_{i2}, \ldots, u_{id'}]$, s.t. $u_{ij}^T S_t u_{ij} > 0$, $u_{ij}^T S_b u_{ij} > u_{ij}^T S_w u_{ij})$. Moreover, we project onto the range space of the matrix S_b' and get a projection matrix $W_{proj} \in R^{d' \times d''}$. Second, we use IFDA to find out the feature representation in the lower dimensionality feature space $R^{d''}$ and obtain a transformation matrix W_{IFDA}. Consequently, we have the transformation matrix W_{opt} of IFDA approach as follows:

$$W_{opt} = W_{IFDA}^T \cdot W_{proj}^T \cdot W_{PCA\_S}^T \tag{10}$$

$$W_{IFDA} = \arg\min_W \frac{|W^T W_{proj}^T W_{PCA\_S}^T S_w W_{PCA\_S} W_{proj} W|}{|W^T W_{proj}^T W_{PCA\_S}^T S_b W_{PCA\_S} W_{proj} W|} = \arg\min_W \frac{|W^T W_{proj}^T S_w' W_{proj} W|}{|W^T W_{proj}^T S_b' W_{proj} W|} = \arg\min_W \frac{|W^T S_w'' W|}{|W^T S_b'' W|} \tag{11}$$

Those eigenvectors with respect to the smaller eigenvalues of S_t are taken into account in IFDA, and IFDA can extract discriminant vectors in the null space of S_w rather than just throw them away. Unfortunately, the IFDA assumes that the same level of typicality of each face to the corresponding class.

3. Fuzzy inverse FDA (FIFDA)

IFDA is considered to solve binary classification problems. While overlapping samples existed, performance of IFDA may degenerate. How can we represent the distribution of these samples and improve classification performance through extracting discriminative information from these samples? Obviously, fuzzy set theory is a good choice. In this paper, we proposed a new fuzzy inverse FDA algorithm, which makes full distribution of samples. Samples distribution information is represented by fuzzy membership degree corresponding to every class.

3.1. Fuzzy K-nearest neighbor (FKNN)

In our method, fuzzy membership degree and each class center are obtained through FKNN [18] algorithm. With FKNN algorithm, the computations of the membership degree can be realized through a sequence of steps:

Step 1: Compute the Euclidean distance matrix between pairs of feature vectors in training set.

Step 2: Set diagonal elements of this Euclidean distance matrix to infinity.

Step 3: 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).

Step 4: Compute the membership degree to class "i" for jth pattern using the expression proposed in the literature [18].

$$u_{ij} = \begin{cases} 0.51 + 0.49 \times (n_{ij}/k) & \text{if } i = \text{the same as the } j\text{th label of the pattern} \\ 0.49 \times (n_{ij}/k) & \text{if } i \neq \text{the same as the } j\text{th label of the pattern} \end{cases} \tag{12}$$

In the above expression n_{ij} stands for the number of the neighbors of the jth data (pattern) that belong to the ith class. As usual, u_{ij} satisfies two obvious properties:

$$\sum_{i=1}^{c} u_{ij} = 1 \tag{13}$$

$$0 < \sum_{j=1}^{N} u_{ij} < N \tag{14}$$

Therefore, the fuzzy membership matrix U can be achieved with the result of FKNN.

$$U = [u_{ij}], \quad i = 1, 2, \ldots, c, \quad j = 1, 2, \ldots, N \tag{15}$$

3.2. The idea of fuzzy inverse FDA

The key step of FIFDA is how to incorporate the contribution of each training sample into the redefine of scatter matrices. With the conception of fuzzy set theory, every sample can be classified into multi classes under fuzzy membership degree, which is different to binary classification problem. 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, in contrast to the redefinition of the within-class scatter matrix, class which is far from total center will have more contribution to classification. Taking into account the fuzzy membership degree, the mean vector of each class is

$$\widetilde{m}_i = \frac{\sum_{j=1}^{N} u_{ij}^p x_j}{\sum_{j=1}^{N} u_{ij}^p} \tag{16}$$

Then, the membership degree of each sample (contribution to each class) should be considered and the corresponding fuzzy within-class scatter matrix, fuzzy between-class scatter matrix and fuzzy total scatter matrix can be redefined as follow:

$$FS_w = \sum_{i=1}^{c} \sum_{x_j \in w_i} u_{ij}^p (x_j - \widetilde{m}_i)(x_j - \widetilde{m}_i)^T \tag{17}$$

$$FS_b = \sum_{i=1}^{c} \sum_{j=1}^{N} u_{ij}^p (\widetilde{m}_i - \overline{X})(\widetilde{m}_i - \overline{X})^T \tag{18}$$

$$FS_t = FS_b + FS_w \tag{19}$$

where p is a constant which controls the influence of fuzzy membership degree, $\overline{X}$ is the mean of all samples.

So, all scatter matrices with fuzzy set theory are redefined and the contribution of each sample is incorporated. From Eqs. (4)–(6) and Eqs. (16)–(18), we can find that our proposed method can more fully distribute the samples.

Finally the fuzzy inverse Fisher criterion function can be defined as follows:

$$W_{F\text{-}IFDA} = \arg\min_W \frac{|W^T FS_w W|}{|W^T FS_b W|} \tag{20}$$

This criterion is a Rayleigh quotient in form. If FS_w is nonsingular, it is easy to find its optimal solutions by solving a generalized eigen-equation.

But in many practical face recognition tasks, there are not enough samples to make the FS_w nonsingular. There is still a small sample size problem. Here we adopt the same strategy in IFDA. We can exploit PCA transformation to reduce the dimension of the original feature space. First, we apply the same PCA transformation as in IFDA to lower the dimension from d to d'. Second, in the PCA_S subspace, we calculate membership degree matrix by FKNN and calculate the fuzzy between-class scatter matrix and fuzzy within-class scatter matrix using the membership degree matrix. Moreover, we project fuzzy between-class scatter matrix onto its range space. Last, we use IFDA to find out the fuzzy IFDA feature representation in the range space of the fuzzy between-class scatter matrix. Since FKNN is used to calculate the membership in the proposed method FIFDA and FKNN involves a sort program, it needs more computational cost than PCA and LDA.

3.3. The algorithm of fuzzy inverse FDA

Based on the above descriptions, fuzzy inverse FDA algorithm can be described as follows:

Step 1: (PCA): PCA transformation is implemented on original image spaces, $S_t = U^T \Lambda U$, $\Lambda = diag(\lambda_1, \lambda_2, \ldots, \lambda_g, 0, \ldots, 0)$, $g = rank(S_t)$, $U = (u_1, u_2, \ldots, u_d)$.

Step 2: (PCA_S): Applying selection rule (for $i = 1, 2, \ldots, g$, if $u_i^T S_b u_i > u_i^T S_w u_i$ then u_i is selected) to the set of $\{u_1, u_2, \ldots, u_d\}$, we get $W_{PCA\_S} = [u_{i1}, u_{i2}, \ldots, u_{id'}]$, where $d' \leqslant \min\{g, c-1\}$). We have the projection matrix W_{PCA_S}: $R^d \to R^{d'}$. Applying it to the sampling matrix X.

Step 3: (FKNN): Class center matrix m and fuzzy membership degree matrix U of training samples can be achieved with FKNN algorithm in PCA_S transformed space $R^{d'}$.

Step 4: (Dimension reduction): According to m and U work out fuzzy within-class scatter matrix FS_w and fuzzy between-class scatter matrix FS_b and fuzzy total scatter matrix FS_t. Project onto the range of FS_b, W_{proj}:$R^{d'} \to R^{d''}$. Calculate the between-class scatter matrix $S'_b = W_{proj}^T \cdot FS_b \cdot W_{proj}$ and the within-class scatter matrix $S'_w = W_{proj}^T \cdot FS_w \cdot W_{proj}$ in the reduced space $R^{d''}$. We get $Y = W_{pro}^T \cdot W_{PCA\_S}^T \cdot X = (Y1, Y2, \ldots, Y_N)$.

Step 5: (FIFDA): The optimization problem (20) using inverse Fisher criterion is solved by $(S'_b)^{-1} S'_w v = \lambda v$ with eigenvalues $0 \leqslant \lambda_1 \leqslant \lambda_2 \leqslant \cdots \leqslant \lambda_q$, $q \leqslant d''$ and corresponding normalized eigenvectors $v_1, v_2, \ldots, v_q$. We have $W_{F-IFDA} = [v_1, v_2, \ldots, v_q]$. We get the optimal projection matrix $W_{opt} = W_{F-IFDA}^T \cdot W_{proj}^T \cdot W_{PCA\_S}^T$.

Step 6: (Recognition): Project all samples into the obtained optimal discriminant vectors and classify.

4. Experiments

Two face image databases, namely the ORL database and FERET database, and a wrist pulse signal database are used to compare the proposed fuzzy inverse FDA approach with the following algorithms: PCA [3], inverse FDA [16], LDA [3], fuzzy Fisherface [19]. The experiments are implemented on AMD Athlon(tm) 64 Processor 3000+ Lenovo computer with 512 M RAM and programmed in the MATLAB language (version 7.01).

4.1. Experiments on 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 56×46. Fig. 1 shows 10 images of one person in ORL.

The first experiment was performed using the first three images per class for training, and the remaining seven images for testing. For feature extraction, we used, respectively, PCA, LDA, fuzzy Fisherface, inverse FDA, and the proposed fuzzy inverse FDA. Note that Fisherface and fuzzy Fisherface involve a PCA phase. In this phase, we keep nearly 99 percent image energy and select the number of principal components, m, as 61. The FKNN parameter K is set as $K = l - 1 = 2$, 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.

In the second experiment, three images of each class are randomly selected to form the training images. The remaining seven images are used for testing. We run the system 10 times. PCA, LDA, inverse FDA, fuzzy Fisherface, and proposed method are, respectively, used for feature extraction. In the PCA stage of LDA and fuzzy Fisherface, the number of principal components, m, is set as 61. The FKNN parameter K is set as $K = 2$. Table 2 shows the maximal average recognition rates across 10 runs of each method under nearest neighbor classifier with cosine distance metrics and their corresponding standard deviations (std.). From Table 2, it can be seen that fuzzy inverse FDA outperforms other methods.

Why can fuzzy inverse FDA outperform other methods? In our opinion, first the overlapping sample's distribution information is completely incorporated in the redefinition of corresponding scatter matrices by fuzzy set theory, which is important for classification. Evidently, as the faces are significantly affected by numerous environmental conditions (including illumination, poses, etc.), it is advantageous to investigate these factors and quantify their impact on their "internal" (viz. algorithm-driven) class assignment. We find that FKNN can be used effectively to enhance the performance of Inverse FDA. We envision that FKNN can capture the underlying clustering of samples. Second, the small principal components that might be essential for classification are considered in the PCA with selection step. In the PCA with selection, we select the vectors that are more effective for classification and small eigenvectors corresponding to the small eigenvalues are considered and might be selected according to the selection rule. Third, the discriminative information in the null space of fuzzy within-class scatter matrix is considered in the feature extraction. When the fuzzy within-class scatter matrix is singular, we still get the discriminative projection vectors because of the PCA with selection. So the final projection vectors consider the discriminative information in the null space of fuzzy within-class scatter matrix.

Table 1
Recognition rate on the ORL face database.

Method	l	Class	Recognition
PCA	3	40	0.8429
LDA	3	40	0.7393
Inverse FDA	3	40	0.8643
Fuzzy Fisherface	3	40	0.7321
Proposed	3	40	0.8857

Table 2
Recognition rate on the ORL face database.

Method	l	Class	Mean/std.
PCA	3	40	0.8918/0.0189
LDA	3	40	0.8236/0.0275
Inverse FDA	3	40	0.8961/0.0193
Fuzzy Fisherface	3	40	0.8232/0.0283
Proposed	3	40	0.8993/0.0203

Fig. 1. Ten images of one person in ORL.

4.2. Experiments on 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 [24,25]. 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. 2.

In our 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, l face images from each person are randomly selected as training samples, and the rest is for testing. The classical PCA, LDA, inverse FDA, fuzzy Fisherface, and the proposed method are, respectively, used for feature extraction. In the PCA stage of LDA and fuzzy Fisherface, we keep nearly 99 percent image energy and select the number of principal components, m, as 299. The FKNN parameter K is set as $K = 2$. Finally a nearest neighbor classifier with cosine distance is employed. The recognition results are shown in Table 3. From Table 3, we can find that our method outperforms all the other methods.

In all the experiments, $p = 2$. Fuzzy inverse FDA makes full distribution of the samples by virtue of fuzzy set theory when it calculates the between-class scatter matrix and within-class scatter matrix, so fuzzy inverse FDA can extract more discriminative feature. So the proposed fuzzy inverse FDA method outperformed other method. The face images in FERET database have significant variants in illuminations and expressions. Using FKNN to get the membership degree matrix, inverse FDA with the redefined fuzzy within-class scatter matrix and fuzzy between-class scatter matrix more efficiently captures the distribution of samples than inverse FDA and fuzzy Fisherface.

Fig. 2. Images of one person in FERET.

Table 3
Recognition rate on the FERET face database.

Method	l	Class	Mean/std.
PCA	3	200	0.4566/0.0957
LDA	3	200	0.2875/0.0420
Inverse FDA	3	200	0.4193/0.0563
Fuzzy Fisherface	3	200	0.2337/0.0308
Proposed	3	200	0.4887/0.0532

4.3. Experiments on pulse signal database

Traditional Chinese pulse diagnosis (TCPD), one of the most important diagnostic techniques in Traditional Chinese Medicine (TCM), is to diagonise disease by means of fingers palpating patients. By analyzing the pressure fluctuation signal of pulse, doctors can detect and predict dome symptoms. TCPD can not only deduce the positions and degree of pathological changes, but also it is a convenient, inexpensive, painless, bloodless, noninvasive, and side-effect free method promoted by U.N. [26]. Computerized analysis of wrist-pulse signals is a crucial step in objectifying and standardizing TCPD [27].

The pulse signals are not a random event and cyclic with regularly occurring percussion, Tidal, and Dicrotic waves. The acquisition process of the pulse signal data used here can be found in Zhang's work [28]. There are three steps in the experiments: pre-process, feature extraction, and classification. First, we use the Gaussian filter to remove the noise and apply the wavelet transform to remove the baseline drift. We choose the discrete Meyer wavelet for this purpose because it is infinitely differentiable and can decrease to zero faster than any inverse polynomial [29]. After the preprocessing, we detect the onset points of wrist-pulse signals. We can segment the wrist-pulse signals according to the onset points. Then, the wrist-signal is normalized to a fixed length of 150 points and the range of [0, 1]. Some normalized pulse signal examples are shown in Fig. 3.

In the experiments, we choose 6 classes of pulse signals (5 classes of disease pulse signals and 1 class of healthy pulse signal) for recognition. Each class includes 5 subjects and each subject has 12 samples. We use the first 6 samples of each subject for training, and use the rest samples for testing. The experimental results are shown in Table 4. From Table 4, we can find that the proposed method has the highest recognition rate, while PCA has better performance than LDA, and fuzzy Fisherface method is better than the inverse FDA. Why can the unsupervised method PCA outperform supervised method LDA and inverse FDA? In our opinion, the possible reason is that PCA is more robust than LDA and inverse FDA to outliers. Actually LDA and inverse FDA extract fewer features in our experiments. Considering the fact that TCM and TCPD depend heavily on the experience of practitioners and

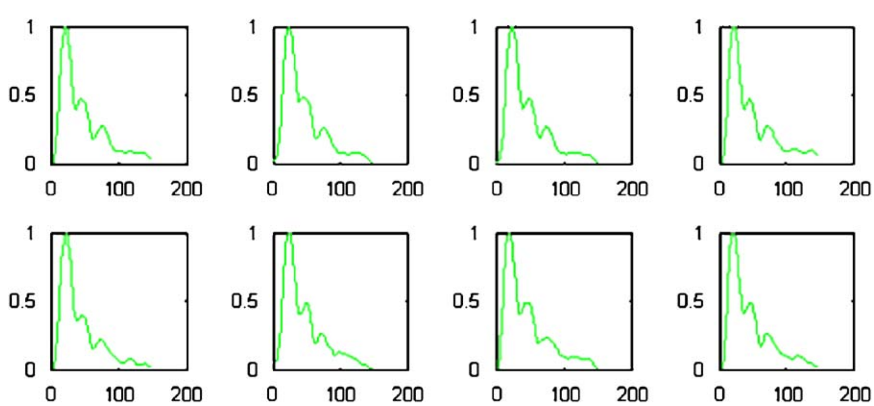

Fig. 3. Some normalized pulse signal samples.

Table 4
Recognition rate on the pulse signal database.

Method	Class	Max recognition rate
PCA	6	0.7556
LDA	6	0.7222
Inverse FDA	6	0.7222
Fuzzy Fisherface	6	0.7444
Proposed	6	0.7778

the diagnosis process is often a fuzzy process, the fuzzy information can be very helpful for TCPD. The experimental results also validate that the methods incorporated with fuzzy information have a better performance.

5. Conclusions

In this paper, we proposed a new method for feature extraction and recognition, namely the fuzzy inverse Fisher discriminant analysis. We re-defined the fuzzy between-class scatter matrix and within-class scatter matrix according to FKNN. This reduces the sensitivity of inverse FDA to substantial variations between face images caused by varying illumination, viewing conditions, and facial expression. By using the fuzzy membership, the influence of the outliers on feature extraction and final classification can be effectively reduced and the more effective discriminative features can be obtained than those by PCA and LDA. The fuzzy between-class scatter matrix and fuzzy within-class scatter matrix in fuzzy inverse FDA can more fully distribute the sample than the fuzzy between-class scatter matrix and fuzzy within-class scatter matrix in fuzzy Fisherface; FIFDA considers the discriminative information in the null space of fuzzy within-class scatter matrix that was ignored in fuzzy Fisherface. Experiments on the ORL, FERET face databases, and pulse signal database show that the proposed method is effective. In the future, we will make more tests on other types of data and extend the fuzzy set theory with kernel methods in the nonlinear case.

Acknowledgments

This work is partially supported by NSFC of China (Grant nos: 60632050, 60503026), Hi-Tech Research and Development Program of China (Grant no: 2006AA01Z119) and the Graduates' Research Innovation Program of Higher Education of Jiangsu Province (Grant no: CX07B_118z). The Hong Kong Polytechnic University Internal Fund under no. YF25. Finally, the authors would like to thank the anonymous reviewers for their constructive advice.

References

[1] 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.
[2] D.L. Swets, J. Weng, Using discriminant eigenfeatures for image retrieval, IEEE Transactions on Pattern Analysis and Machine Intelligence 18 (8) (1996) 831–836.
[3] 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.
[4] J. Yang, J.Y. Yang, Why can LDA be performed in PCA transformed space?, Pattern Recognition 36 (2) (2003) 563–566.
[5] J. Yang, J.Y. Yang, Optimal FLD algorithm for facial feature extraction, SPIE Proceedings of the Intelligent Robots and Computer Vision XX: Algorithm, Techniques, and Active Vision 4572 (2001) 438–444.
[6] 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.
[7] J.H. Friedman, Regularized discriminant analysis, Journal of the American Statistical Association 84 (405) (1989) 165–175.
[8] T. Hastie, R. Tibshirani, Penalized discriminant analysis, The Annals of Statistics 23 (1) (1995) 73–102.
[9] T. Hastie, R. Tibshirani, A. Buja, Flexible discriminant analysis by optimal scoring, Journal of the American Statistical Association 89 (428) (1994) 1255–1270.
[10] 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.
[11] H. Yu, J. Yang, A direct LDA algorithm for high dimensional data—with application to face recognition, Pattern Recognition 34 (10) (2001) 2067–2070.
[12] H. Li, T. Jiang, K. Zhang, Efficient and robust feature extraction by maximum margin criterion, in: NIPS, 2003.
[13] C.E. Thomaz, D.F. Gillies, A maximum uncertainty LDA-based approach for limited sample size problems—with application to face recognition, in: Proceedings of the XVIII Brazilian Symposium on Computer Graphics and Image Processing, pp. 1530–1834.
[14] C.E. Thomaz, D.F. Gillies, R.Q. Feitosa, A new covariance estimate for Bayesian classifier in biometric recognition, IEEE CSVT 14 (2) (2004) 214–223.
[15] 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.
[16] 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.
[17] L.A. Zadeh, Fuzzy sets, Information Control 8 (1965) 338–353.
[18] 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.
[19] K.C. Kwak, W. Pedrycz, Face recognition using a fuzzy Fisherface classifier, Pattern Recognition 38 (10) (2005) 1717–1732.
[20] Y.J. Zheng, J.Y. Yang, et al., Fuzzy Kernel Fisher discriminant algorithm with application to face recognition, in: The Sixth World Congress on Intelligent and Automation (WCICA06), vol. 12(12), pp. 9669–9672.
[21] Y.J. Zheng, J.Y. Yang, X.J. Wu, Y.Z. Li, A new two-dimensional linear discriminant analysis algorithm based on fuzzy set theory, Engineering Science 9 (2) (2007) 49–53.
[22] W.K. Yang, J.G. Wang, M.W. Ren, J.Y. Yang, Fuzzy inverse Fisher discriminant analysis for face recognition, in: International Conference on Wavelet Analysis and Pattern Recognition, vol. 1, Beijing, China, 2007, pp. 107–111.
[23] R. Duda, P. Hart, Pattern Classification and Scene Analysis, Wiley, New York, 1973.
[24] 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.
[25] P.J. Phillps, The facial recognition technology (FERET) database, 〈http://www.itl.nist.gov/iad/humanid/feret/feret_master.html〉, 2004.
[26] L.S. Xu, K.Q. Wang, D. Zhang, Modern researches on traditional Chinese pulse diagnosis, European Journal of Oriental Medicine 4 (5) (2004) 46–54.
[27] Y. Zhou, Q. Xu, W. Kong, Study on several methods of character analysis and classification for pulse-condition based on statistics theory, Journal of Biomedical Engineering 23 (3) (2006) 505–508 (in Chinese).
[28] D.Y. Zhang, L. Zhang, D. Zhang, Y.P. Zheng, Wavelet based analysis of Doppler ultrasonic wrist-pulse signals, in: International Conference on BioMedical Engineering and Informatics, 2008, pp. 539–543.
[29] L.S. Xu, D. Zhang, K.Q. Wang, Wavelet-based cascade adaptive filter for removing the baseline drift in pulse waveforms, IEEE Transactions on Biomedical Engineering 52 (11) (2005) 1973–1975.

Wankou Yang is a Ph.D. candidate at the School of Computer Science and Technology, Nanjing University of Science and Technology, PR China. He received the B.S. and M.S. degrees from the Department of Computer Science, Nanjing University of Science and Technology, PR China, in 2002 and 2004. He is now working as a research assistant in Department of Computing, The Hong Kong Polytechnic University. His research interests include pattern recognition, computer vision, and digital image processing.

Jianguo Wang is an assistant professor in the Network and Education, Center Tangshan College. He received the B.S. degree in printing technology from Xi'an University of technology in 1994 and received the M.S. degree in signal and information processing from Xi'an University of Technology in 1997. He is now working towards his doctor's degree in computer vision and pattern recognition in NUST. His current interests are in the areas of pattern recognition, image processing, face detection and recognition.

Mingwu Ren is an associate professor at the School of Computer Science and Technology, Nanjing University of Science and Technology, PR China. He received M.S. and Ph.D. degrees in Pattern Recognition and Computer Image Processing from Nanjing University of Science and Technology, PR China. His research interests include computer vision, digital image processing, pattern recognition and artificial control.

Lei Zhang received the B.S. degree in 1995 from Shenyang Institute of Aeronautical Engineering, Shenyang, PR China, the M.S. and Ph.D degrees in Automatic Control Theory and Engineering from Northwestern Polytechnical University, Xi'an, PR China, respectively, in 1998 and 2001. From 2001 to 2002, he was a research associate in the Department of Computing, The Hong Kong Polytechnic University. From January 2003 to 2006 he worked as a Postdoctoral Fellow in the Department of Electrical and Computer Engineering, McMaster University, Canada. Since January 2006, he has been an Assistant Professor in the Department 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.

Jingyu Yang received the B.S. degree in computer science from Nanjing University of Science and Technology (NUST), Nanjing, China. From 1982 to 1984, he was a visiting scientist at the Coordinated Science Laboratory, University of Illinois at Urbana-Champaign. From 1993 to 1994, he was a visiting professor at the Department of Computer Science, Missuria University. In 1998, he acted as a visiting professor at Concordia University in Canada. He is currently a professor and chairman in the School of Computer Science and Technology at NUST. He is the author of more than 300 scientific papers in computer vision, pattern recognition, and artificial intelligence. He has won more than 20 provincial awards and national awards. His current research interests are in the areas of pattern recognition, robot vision, image processing, data fusion, and artificial intelligence.