JOURNAL OF INFORMATION SCIENCE AND ENGINEERING 26, 1525-1537 (2010)

## Short Paper

# Face Recognition Using L-Fisherfaces*

CHENG-YUAN ZHANG[1,2] AND QIU-QI RUAN[1]
[1]*Institute of Information Science*
*Beijing Jiaotong University*
*Beijing, 100044 China*
[2]*College of Information and Electrical Engineering*
*Shandong University of Science and Technology*
*Qingdao, 266510 China*

An appearance-based face recognition approach called the L-Fisherfaces is proposed in this paper, By using Local Fisher Discriminant Embedding (LFDE), the face images are mapped into a face subspace for analysis. Different from Linear Discriminant Analysis (LDA), which effectively sees only the Euclidean structure of face space, LFDE finds an embedding that preserves local information, and obtains a face subspace that best detects the essential face manifold structure. Different from Locality Preserving Projections (LPP) and Unsupervised Discriminant projections (UDP), which ignore the class label information, LFDE searches for the project axes on which the data points of different classes are far from each other while requiring data points of the same class to be close to each other. We compare the proposed L-Fisherfaces approach with PCA, LDA, LPP, and UDP on three different face databases. Experimental results suggest that the proposed L-Fisherfaces provides a better representation and achieves higher accuracy in face recognition.

***Keywords:*** face recognition, local Fisher discriminant embedding, manifold learning, locality preserving projections, unsupervised discriminant projections

## 1. INTRODUCTION

Face recognition has aroused wide concerns over the past few decades due to its potential applications, such as criminal identification, credit card verification, and security system and scene surveillance. In the literature, various algorithms have been proposed for this problem [1, 2]. PCA and LDA are two well-known linear subspace-learning techniques and have become the most popular methods for face recognition [3-5]. Recently, He *et al.* [6, 7] and Yang *et al.* [8, 9] proposed two manifold learning based methods, namely, Locality Preserving Projections (LPP) and unsupervised discriminant projection (UDP), for face recognition. LPP is a linear subspace method derived from Laplacian Eigenmap [10]. It results in a linear map that optimally preserves local neighborhood information and its objective function is to minimize the local scatter of the projected data. Unlike LPP, UDP finds a linear map based on the criterion that seeks to maximize

Received July 29, 2008; revised October 30, 2008; accepted January 8, 2009.
Communicated by H. Y. Mark Liao.
* This work was partially supported by the National Natural Science Foundation of China (NSFC, No. 60672062) and the Major State Basic Research Development Program of China (973 Program No. 2004CB318005).

the ratio of the non-local scatter to the local scatter. UDP is declared more effective than LPP for classification purpose since it utilizes the information of "non-locality". In contrast to most manifold learning algorithms, such as LLE [11] and ISOMAP [12], a remarkable advantage of LPP and UDP is that it can generate a simple and efficiently computable linear map. However, LPP and UDP are unsupervised methods that ignore the class label information of the training samples, so they can hardly perform well if the face images are badly clustering in data space.

In this paper, we present a manifold-learning face recognition approach called the Local Fisher Discriminant Embedding (LFDE) method. Different from Linear Discriminant Analysis (LDA), which effectively sees only the Euclidean structure of face space, LFDE finds an embedding that preserves local information, and obtains a face subspace that best detects the essential face manifold structure. Different from Locality Preserving Projections (LPP) and Unsupervised Discriminant projections (UDP), LFDE searches for the project axes on which the data points of different classes are far from each other while requiring data points of the same class to be close to each other. Considering the fact that samples on the edge of a class cluster are more misclassiable, the weighted value of a sample is chosen according to its distance from the analyzed sample. Thus, samples at different distance contribute differently to the local within-class scatter matrix and local between-class scatter matrix. By LFDE, each face image in the image space is mapped to a low-dimensional subspace, which is characterized by a set of feature images, called Local Fisherfaces (L-Fisherfaces). This subspace preserves local data structure and has discriminanting power for face recognition.

The rest of this paper is organized as follows. Section 2 reviews the PCA and LDA algorithms. Section 3 outlines LPP and UDP algorithms. Section 4 develops the idea of LFDE and the algorithm of L-Fisherfaces. Experiments are carried out on PIE, FERET, and ORL face databases in section 5. Some conclusions are given in section 6.

## 2. PCA AND LDA

Linear subspace methods project the high-dimensional data onto a lower dimensional subspace by linearly combining features. Considering the problem of representing all of the vectors in a set of $M$ $N$-dimensional samples $\boldsymbol{X} = [\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_M]_{N\times M}$ by a single vector $\boldsymbol{y} = [y_1, y_2, \ldots, y_M]$ such that $y_i$ represents $\boldsymbol{x}_i$. The transformation vector is denoted by $\boldsymbol{w}$. That is, $y_i = \boldsymbol{w}^T\boldsymbol{x}_i$.

PCA aims to extract a subspace in which the variance is maximized. Its objective function is as follows:

$$\max_{w} \sum_{i=1}^{M} (y_i - \bar{y})^2, \tag{1}$$

where $\bar{y}$ is the mean of $\boldsymbol{y}$. The output set of principal vectors $\boldsymbol{w}_1, \boldsymbol{w}_2, \ldots, \boldsymbol{w}_n$ is set of eigenvectors of the sample covariance matrix associated with the $n < N$ largest eigenvalues.

While PCA seeks projections that are efficient for representation, LDA seeks directions that are efficient for classification. Suppose that each sample belongs to one of the classes $\{C_i\}_{i=1}^{L}$. The objective function is as follows:

$$\max_{w} \frac{\boldsymbol{w}^T S_b \boldsymbol{w}}{\boldsymbol{w}^T S_w \boldsymbol{w}}, \tag{2}$$

where,

$$\begin{aligned} S_w &= \sum_{j=1}^{L} \sum_{\boldsymbol{x}_i \in C_j} (\boldsymbol{x}_i - m_j)(\boldsymbol{x}_i - m_j)^T \\ S_b &= \sum_{i=1}^{L} M_i (m_i - m)(m_i - m)^T, \end{aligned} \tag{3}$$

where $m_i$ denotes the mean of samples in $C_i$ class, $M_i$ is the number of samples in this class, and $m$ is the mean of all the samples. $S_w$ and $S_b$ are the within-class and between-class scatter matrices of the training samples. So, the aim of LDA is to find a projection axis that maximizes the between-class scatter and at the same time minimizes the within-class scatter.

## 3. LPP AND UDP

PCA and LDA aim to preserve the global structure. However, in many real-world applications, the local structure is more important. LPP [6, 7] and UDP [8, 9] are manifold- learning based algorithms that preserve the "locality" information of data space.

### 3.1 LPP

The objective function of LPP is as follows:

$$\min \sum_{i,j} S_{ij} (y_i - y_j)^2, \tag{4}$$

where $y_i$ is the representation of $\boldsymbol{x}_i$ in the transformed space, *i.e.* $y_i = \boldsymbol{w}^T\boldsymbol{x}_i$. $\boldsymbol{S}$ is a weight matrix, and the choice of $S_{ij}$ incurs a heavy penalty if neighboring points $\boldsymbol{x}_i$ and $\boldsymbol{x}_j$ are mapped far apart. So, the matrix $\boldsymbol{S}$ can be defined as follows:

$$S_{ij} = \begin{cases} \exp(-\|\boldsymbol{x}_i - \boldsymbol{x}_j\|^2 / \sigma), & \boldsymbol{x}_i \text{ is among } k \text{ nearest neighbor of } \boldsymbol{x}_j \text{ or } \boldsymbol{x}_j \\ & \text{is among } k \text{ nearest neighbor of } \boldsymbol{x}_i \\ 0, & \text{otherwise} \end{cases} \tag{5}$$

where $\sigma$ is a parameter that can be determined empirically.

Suppose $w$ is a mapping vector, the objective function (4) can be reduced to:

$$\frac{1}{2}\sum_{i,j} S_{ij}(y_i - y_j)^2 = \frac{1}{2}\sum_{i,j} S_{ij}(\boldsymbol{w}^T\boldsymbol{x}_i - \boldsymbol{w}^T\boldsymbol{x}_j)^2 = \sum_i \boldsymbol{w}^T\boldsymbol{x}_i D_{ii}\boldsymbol{x}_i^T\boldsymbol{w} - \sum_{i,j}\boldsymbol{w}^T\boldsymbol{x}_i S_{ij}\boldsymbol{x}_j^T\boldsymbol{w} \\ = \boldsymbol{w}^T X(D-S)X^T\boldsymbol{w} = \boldsymbol{w}^T XLX^{\mathrm{T}}\boldsymbol{w}, \tag{6}$$

where $\boldsymbol{D}$ is a diagonal matrix, its entries are column (or row, since $\boldsymbol{S}$ is symmetric) sum of $\boldsymbol{S}$, $D_{ii} = \sum_j S_{ij}$. $L = D - S$ is the Laplacian matrix [10]. A constraint is added as follows:

$$y^T D y = 1, \text{ so that } \boldsymbol{w}^T XDX^T\boldsymbol{w} = 1. \tag{7}$$

Finally, the projection vectors are given by the eigenvectors corresponding to the $n$ smallest eigenvalues of the following eigenfunction:

$$XLX^T\boldsymbol{w} = \lambda XDX^T\boldsymbol{w}. \tag{8}$$

### 3.2 UDP

Different from LPP that only considers the local scatter defined by Eq. (6), UDP simultaneously considers the local and nonlocal scatter in the projected space [8, 9]. As LDA is better than PCA when used in pattern recognition, UDP is declared to outperform LPP due to the criterion used in UDP that maximizes the ratio of nonlocal scatter to local scatter. Let a set $U^b$ denote two samples $\boldsymbol{x}_i$ and $\boldsymbol{x}_j$ within a local $\delta-$ neighborhood, that is $U^b = \{(i, j) \| \|\mathbf{x}_i - \mathbf{x}_j\|^2 < \delta\}$. The local scatter in the transformed space is defined by

$$J_L(\boldsymbol{w}) \triangleq \frac{1}{2}\sum_{(i,j)\in U^b}(y_i - y_j)^2, \tag{9}$$

and the nonlocal scatter in the transformed space is defined by

$$J_N(\boldsymbol{w}) \triangleq \frac{1}{2}\sum_{(i,j)\notin U^b}(y_i - y_j)^2, \tag{10}$$

where, $y_i = \boldsymbol{w}^T\boldsymbol{x}_i$ and $y_j = \boldsymbol{w}^T\boldsymbol{x}_j$.

Let us define the adjacency matrix $H$:

$$H_{ij} = \begin{cases} 1, & \boldsymbol{x}_i \text{ is among } k \text{ nearest neighbor of } \boldsymbol{x}_j \text{ or } \boldsymbol{x}_j \text{ is among} \\ & k \text{ nearest neighbor of } \boldsymbol{x}_i \\ 0, & \text{otherwise} \end{cases}. \tag{11}$$

So, Eqs. (9) and (10) can be rewritten by

$$J_L(\boldsymbol{w}) = \frac{1}{2}\sum_{i=1}^{M}\sum_{j=1}^{M} H_{ij}(y_i - y_j)^2, \tag{12}$$

and

$$J_N(\boldsymbol{w}) = \frac{1}{2}\sum_{i=1}^{M}\sum_{j=1}^{M}(1-H_{ij})(y_i - y_j)^2. \quad (13)$$

For the purpose of classification, after being projected onto subspace, the close samples are closer together while mutually distant samples are more distant from each other. From this point of view, we can obtain such a projection by maximizing the following criterion:

$$J(\boldsymbol{w}) = \frac{J_N(\boldsymbol{w})}{J_L(\boldsymbol{w})}. \quad (14)$$

The criterion in Eq. (14) is formally similar to the Fisher criterion in Eq. (2). However, Eq. (14) can be resolving without knowing the class-label of samples. This means Fisher discriminant projection is supervised, while the projection of UDP can be obtained in an unsupervised manner.

## 4. LOCAL FISHER DISCRIMINANT EMBEDDING (LFDE)

LPP and UDP suffer from a limitation: they de-emphasize discriminant information of the training samples, which make them not suitable for recognition task when the face images are badly clustering in data space. On the other hand, although LDA emphasizes discriminant information, it only aims to preserve the global geometric structure of data in a transformed low-dimensional space and the global within-class scatter and between-class scatter of samples are considered. Instead, the proposed LFDE aims to discover the local structure of data as well as to extract the discriminant information for pattern representation and classification; so, the local within-class scatter and between-class scatter of samples are considered in our proposed method.

### 4.1 Local Within-Class Scatter Matrix

Let $\boldsymbol{X} = [\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_M]_{N\times M}$ denote the training set, where $N$ is the dimension and $M$ is number of the samples, respectively. Each sample belongs to one of $L$ classes $\{C_i\}_{i=1}^{L}$.

The local within-class scatter can be characterized by the mean square of the Euclidean distance between any pair of the projected sample points that belong to a same class. After the projection of $\boldsymbol{x}_i$ and $\boldsymbol{x}_j$ onto a direction $\boldsymbol{w}$, we get their images $y_i$ and $y_j$. The local within-class scatter of the projected samples is then defined by

$$J_W(\boldsymbol{w}) \triangleq \frac{1}{2}\sum_{k=1}^{L}\sum_{\boldsymbol{x}_i,\boldsymbol{x}_j\in C_k}(y_i - y_j)^2. \quad (15)$$

Let us denote the within-class matrix $\boldsymbol{H}^w$:

$$H_{ij}^{w}=\begin{cases}1, & \boldsymbol{x}_i \text{ and } \boldsymbol{x}_j \text{ belong to a same class}\\ 0, & \text{otherwise.}\end{cases} \quad (16)$$

So Eq. (15) can be rewritten as:

$$J_w(\boldsymbol{w}) \triangleq \frac{1}{2}\sum_{i=1}^{M}\sum_{j=1}^{M}H_{ij}^{w}(y_i-y_j)^2=\frac{1}{2}\sum_{i=1}^{M}\sum_{j=1}^{M}H_{ij}^{w}(\boldsymbol{w}^T\boldsymbol{x}_i-\boldsymbol{w}^T\boldsymbol{x}_j)^2=\boldsymbol{w}^T S_w^L\boldsymbol{w}, \quad (17)$$

where

$$S_w^L=\tfrac{1}{2}\sum_{i=1}^{M}\sum_{j=1}^{M}H_{ij}^{w}(\mathbf{x}_i-\mathbf{x}_j)(\mathbf{x}_i-\mathbf{x}_j)^T. \quad (18)$$

$S_w^L$ is called the local within-class scatter matrix. Due to the symmetry of $\boldsymbol{H}^w$, it follows that

$$\begin{aligned}S_w^L&=\frac{1}{2}\Big(\sum_{i=1}^{M}\sum_{j=1}^{M}H_{ij}^{w}\boldsymbol{x}_i\boldsymbol{x}_i^T+\sum_{i=1}^{M}\sum_{j=1}^{M}H_{ij}^{w}\boldsymbol{x}_j\boldsymbol{x}_j^T-2\sum_{i=1}^{M}\sum_{j=1}^{M}H_{ij}^{w}\boldsymbol{x}_i\boldsymbol{x}_j^T\Big)\\&=\sum_{i=1}^{M}D_{ii}^{w}\mathbf{x}_i\mathbf{x}_i^T-\sum_{i=1}^{M}\sum_{j=1}^{M}H_{ij}^{w}\boldsymbol{x}_i\boldsymbol{x}_j^T=\boldsymbol{X}\boldsymbol{D}^w\boldsymbol{X}^T-\boldsymbol{X}\boldsymbol{H}^w\boldsymbol{X}^T=\boldsymbol{X}\boldsymbol{L}^w\mathbf{X}^T,\end{aligned} \quad (19)$$

where $\boldsymbol{D}^w$ is a diagonal matrix whose elements on diagonal are column (or row) sum of $\boldsymbol{H}^w$ and $\boldsymbol{L}^w = \boldsymbol{D}^w - \boldsymbol{H}^w$.

### 4.2 Local between-Class Scatter Matrix

The local between-class scatter can be characterized by the mean square of the Euclidean distance between any pair of the projected sample points that belong to different classes. So, the local between-class scatter of the projected samples is defined by

$$J_b(\boldsymbol{w}) \triangleq \frac{1}{2}\sum_{\boldsymbol{x}_i\in C_k,\boldsymbol{x}_j\in C_l,k\neq l}H_{ij}^{b}(y_i-y_j)^2, \quad (20)$$

where, the between-class matrix $\boldsymbol{H}^b$ is defined as

$$H_{ij}^{b}=\begin{cases}0, & \boldsymbol{x}_i,\boldsymbol{x}_j\in C_k,(k=1,\ldots,L),\\ \exp(-\|\boldsymbol{x}_i-\boldsymbol{x}_j\|^2/2\sigma), & \boldsymbol{x}_i\in C_k,\boldsymbol{x}_j\in C_l,k\neq l,\end{cases} \quad (21)$$

where $\sigma=\frac{1}{M(M-1)}\sum_{i,j=1}^{M}\|\boldsymbol{x}_i-\boldsymbol{x}_j\|^2$ is the mean distance of each two training samples in the data space. The motivation of $\boldsymbol{H}^b$'s choice is based on the fact that those samples, which belong to different classes and are closer to each other, are potentially more confusing and should be given more attention during the feature extraction stage. From this point, LFDE is somewhat similar to the weighted linear discriminant analysis (WLDA) methods [13, 14].

So, Eq. (20) can be rewritten as:

$$J_b(\boldsymbol{w}) \triangleq \frac{1}{2}\sum_{i=1}^{M}\sum_{j=1}^{M} H_{ij}^b (y_i - y_j)^2 = \tfrac{1}{2}\sum_{i=1}^{M}\sum_{j=1}^{M} H_{ij}^b (\boldsymbol{w}^T\boldsymbol{x}_i - \boldsymbol{w}^T\boldsymbol{x}_j)^2 = \boldsymbol{w}^T S_b^L \boldsymbol{w}, \tag{22}$$

where

$$S_b^L = \frac{1}{2}\sum_{i=1}^{M}\sum_{j=1}^{M} H_{ij}^b (\boldsymbol{x}_i - \boldsymbol{x}_j)(\boldsymbol{x}_i - \boldsymbol{x}_j)^T. \tag{23}$$

$S_b^L$ is called the local between-class scatter matrix. Due to the symmetry of $\boldsymbol{H}^b$, it follows that

$$\begin{aligned} S_b^L &= \frac{1}{2}\Big(\sum_{i=1}^{M}\sum_{j=1}^{M} H_{ij}^b \boldsymbol{x}_i\boldsymbol{x}_i^T + \sum_{i=1}^{M}\sum_{j=1}^{M} H_{ij}^b \boldsymbol{x}_j\boldsymbol{x}_j^T - 2\sum_{i=1}^{M}\sum_{j=1}^{M} H_{ij}^b \boldsymbol{x}_i\boldsymbol{x}_j^T\Big) \\ &= \sum_{i=1}^{M} D_{ii}^b \boldsymbol{x}_i\boldsymbol{x}_i^T - \sum_{i=1}^{M}\sum_{j=1}^{M} H_{ij}^b \boldsymbol{x}_i\boldsymbol{x}_j^T = \boldsymbol{X}\boldsymbol{D}^b\mathbf{X}^T - \boldsymbol{X}\boldsymbol{H}^b\boldsymbol{X}^T = \boldsymbol{X}\boldsymbol{L}^b\boldsymbol{X}^T, \end{aligned} \tag{24}$$

where $\boldsymbol{D}^b$ is a diagonal matrix whose elements on diagonal are column (or row) sum of $\boldsymbol{H}^b$ and $\boldsymbol{L}^b = \boldsymbol{D}^b - \boldsymbol{H}^b$.

### 4.3 Fisher Criterion

For the purpose of classification, an object is to find a projection, which makes the samples in the same class become closer, and the samples in different classes more distant. From this point of view, the Fisher criterion is used to obtain such a projection by maximizing the local between-class scatter as well as minimizing the local within-class scatter:

$$\boldsymbol{w} = \arg\max_{\boldsymbol{w}} \frac{\boldsymbol{w}^T \boldsymbol{S}_b^L \boldsymbol{w}}{\boldsymbol{w}^T \boldsymbol{S}_w^L \boldsymbol{w}}. \tag{25}$$

If the local within-class scatter matrix $S_w^L$ is non-singular, the solution of Eq. (25) can be calculated directly by solving the following generalized eigen-equation:

$$S_b^L \boldsymbol{w} = \lambda S_w^L \boldsymbol{w}. \tag{26}$$

The projection matrix $\boldsymbol{w}$ of LFDE can be formed by the eigenvectors $\boldsymbol{w}_1, \boldsymbol{w}_2, \ldots, \boldsymbol{w}_n$ corresponding to $n$ largest positive eigenvalues $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_n$.

In face recognition, however, $S_w^L$ is always singular due to the limitation of training samples. In such cases, the classical algorithm cannot be used directly to solve the eigen-equation. To avoid this difficult, we can adopt two-phase strategy, that is, PCA is first used for dimension reduction and then LFDE is performed in the PCA-projected space.

### 4.3 Algorithm of L-Fisherfaces

In summary of the preceding description, the L-Fisherfaces algorithm is given below:

**Step 1:** Use PCA to transform the input space $\Re^N$ into an $n$-dimensional space $\Re^n$, where $n \ll N$. Since $rank(S_w^L) \leq M - L$, where $M$ is the number of training samples, and $L$ is the number of classes. So, we choose $n \leq M - L$ to make $S_w^L$ nonsingular.

**Step 2:** Construct the within-class and between-class matrix: For the given training data projected into PCA subspace, construct the within-class matrix $\boldsymbol{H}^w$ and between-class matrix $\boldsymbol{H}^b$ using Eqs. (16) and (21), respectively.

**Step 3:** Construct the local within-class and between-class scatter matrices $S_w^L$ and $S_b^L$ in PCA subspace using Eqs. (19) and (24).

**Step 4:** Calculate the generalized eigenvectors $\alpha_1, \alpha_2, \ldots, \alpha_d$ $(d \leq n)$ of $S_b^L\alpha = \lambda S_w^L\alpha$ corresponding to the $d$ largest positive eigenvalues $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_d$. Let $W_{PCA}$ denote the transform matrix of PCA and $A = (\alpha_1, \alpha_2, \ldots, \alpha_d)$. $W_{LFDE} = W_{PCA}A$ is the transformation matrix from the image space into the LFDE subspace. The column vectors of $W_{LFDE}$ are the so-called L-Fisherfaces. The feature vector extracted by LFDE is $\boldsymbol{y} = W_{LFDE}^T\boldsymbol{x} = A^TW_{PCA}^T\boldsymbol{x}$.

Using the ORL face database as the training set, we present the first ten L-Fisherfaces in Fig. 1, together with Eigenfaces, Fisherfaces and Laplacianfaces [7]. It is interesting to note that the L-Fisherfaces are somehow similar to the combination of Fisherfaces and Laplacianfaces.

Fig. 1. The first 10 Eigenfaces (1st row), Fisherfaces (2nd row), Laplacianfaces (3rd row), and L-Fisherfaces (4th row) calculated from the face images in the ORL database.

## 5. EXPERIMENTS

In this section, the performance of L-Fisherfaces is evaluated on the PIE, FERET, and ORL face databases and compared with the performances of PCA, LDA, LPP, and UDP.

### 5.1 PIE Database

The CMU PIE face database contains 68 subjects with 41,368 face images as a whole [15]. The face images were captured under varying pose, illumination, and expression. Each image is cropped to 32 × 32 pixels based on the location of eyes and mouth, and with 256 gray levels per pixel. We use 170 face images for each individual in our experiment. Fig. 2 shows all the 170 samples of one individual in this database.

Fig. 2. Samples of one individual in the PIE database.

We choose the first 15 images of each person for training and the images left for testing. In the PCA phase of LDA, LPP, UDP, and LFDE the number of principal components is set as 200. The *K*-nearest neighborhood parameter *K* in LPP and UDP is chosen as $K = l - 1$ ($l$ denotes the number of training samples in a class, so $l = 15$). After feature extraction, a nearest neighbor classifier is employed for classification. Fig. 3 shows a plot of recognition rates of PCA (its dimension is fixed at 200), LPP, UDP, LDA, and L-Fisherfaces versus the variation of dimensions. The maximal recognition rate of each method and the corresponding dimension are given in Table 1.

From Table 1 we can find three main points. First, LPP and UDP outperforming LDA and PCA demonstrates that the local data structure based methods are more suitable than the global data structure based ones in classification. Although LPP and UDP are unsupervised methods, they even outperform the supervised method LDA. Second, the proposed L-Fisherfaces achieves the best recognition accuracy, which demonstrates the superiority of simultaneously considering local data structure and class label information. Third, the dimension corresponding to top recognition accuracy of L-Fisherfaces is smaller than that of other methods.

### 5.2 FERET Database

Our second experiment is carried out on a subset of the FERET database [16], which includes 582 images of 194 individuals (each one has 3 images). It is composed of the images whose names are marked with two-character strings: “ba”, “bj”, “bk”. This subset

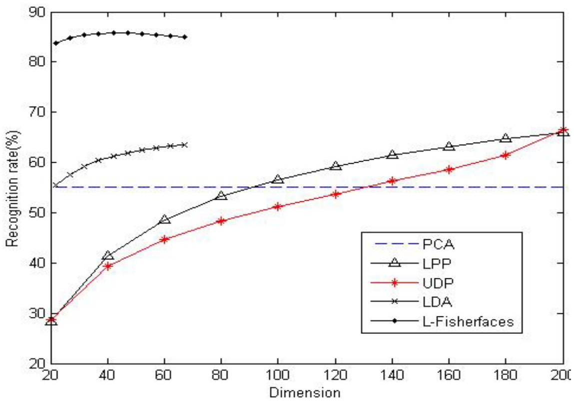


Fig. 3. Recognition rates of LDA, LPP, UDP, and L-Fisherfaces versus dimensions on PIE database.

**Table 1. The maximal recognition rates (%) of PCA, LDA, LPP, UDP, and L-Fisherfaces on PIE database and the corresponding dimensions.**

| Methods | Recognition rates (%) | Dimension |
|---|---|---|
| PCA | 55.01 | 200 |
| LDA | 63.5 | 67 |
| LPP | 65.98 | 200 |
| UDP | 66.51 | 200 |
| L-Fisherfaces | 85.75 | 47 |

involves variations in facial expression and illumination. In our experiment, the facial portion of each original image was cropped based on the location of eyes and mouth, and the cropped image was resized to 64 × 64 pixels.

We use two images per class for training, and the remaining image for test, so there are three different selections and the recognition rate is averaged over the three tests. PCA, LDA, LPP, UDP, and the proposed L-Fisherfaces are test on the same training and testing set, respectively. In the PCA phase of LDA and LFDE the number of principal components is set as 194, and for LPP and UDP, it is set as 387 (2 × 194 − 1). After feature extraction, a nearest neighbor classifier is employed for classification. Fig. 4 shows a plot of recognition rates of PCA (its dimension is fixed at 300), LDA, LPP, UDP, and L-Fisherfaces versus the variation of dimensions.

There are also three main points can be seen from Fig. 4. First, unlike the trends of LDA and LFDE's recognition rates, the recognition rates of LPP and UDP are ascending with the increasing of dimensions. LPP and UDP outperform LDA when the feature dimension is over 160. Second UDP performs better than LPP when the dimension is over 180. Third, the proposed L-Fisherfaces consistently performs better than other methods. The maximal recognition rate of each method is as follows: PCA: 38.14%, LDA: 46.74%, LPP: 53.61%, UDP: 58.76%, and L-Fisherfaces: 80.76%.

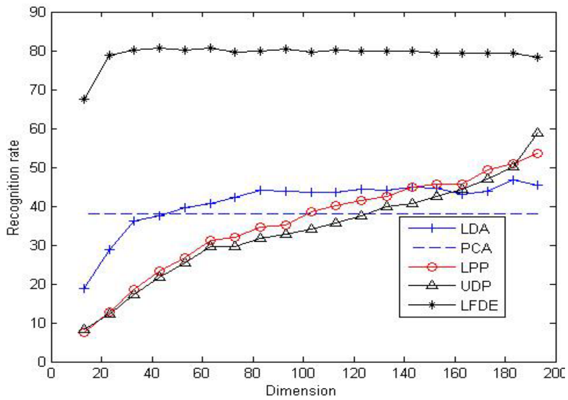


Fig. 4. Recognition rates of LDA, LPP, UDP, and L-Fisherfaces (LFDE) versus dimensions on FERET database.

### 5.3 ORL Database

The ORL database contains 400 images from 40 individuals, each individual providing 10 different images. The facial expressions (open or closed eyes, smiling or non-smiling) and facial details (glasses or no glasses) vary in this database. Some tilting and rotation of the face is up to 20°. Moreover, there is also some variation in the scale of up to about 10%. All images in this database are grayscale images with 256 gray levels and normalized to a resolution of 92 × 112 pixels. For the purpose of computation efficiency, all images are cropped and resized to 48 × 48 pixels based on the location of eyes and mouth. Two persons' sample faces are shown in Fig. 5.

Fig. 5. Samples of two individual in the ORL database.

In our experiments, $l$ images ($l$ varies from 2 to 5) are randomly selected from the image gallery of each individual to form the training set. The remaining images are used for testing. For each fixed $l$, we run the system 10 times and the recognition rates are averaged. In the PCA phase of LDA, LPP, UDP, and LFDE the number of principal components is set as $40 \times (l - 1)$. The $K$-nearest neighborhood parameter $K$ in LPP and UDP is chosen as $K = l - 1$. After feature extraction, a nearest neighbor classifier is employed for classification. The top recognition rate of each method and the corresponding dimension (filled in parentheses) with different $l$ are shown in Table 2.

Table 2 demonstrates again that L-Fisherfaces overall outperforms PCA, LDA, LPP, and UDP no matter how many samples are selected for training.

**Table. 2. The maximal recognition rates (%) of PCA, LDA, LPP, UDP, and L-Fisherfaces and the corresponding dimensions with different training sample sizes on ORL database.**

| Training samples/class | 2 | 3 | 4 | 5 |
|---|---|---|---|---|
| PCA | 72.37 (79) | 78.14 (80) | 80.63 (99) | 82.35 (149) |
| LDA | 77.22 (21) | 83.36 (39) | 90.46 (33) | 93.65 (36) |
| LPP | 72.28 (39) | 78.15 (79) | 81.45 (119) | 83.1 (159) |
| UDP | 72.22 (39) | 79.36 (79) | 83.75 (119) | 85.1 (159) |
| L-Fisherfaces | 79.53 (24) | 88.31(39) | 94.33 (39) | 96.3 (27) |

## 6. CONCULATIONS

A novel feature extraction approach called the Local Fisher Discriminant Embedding (LFDE) method is proposed in this paper. Based on LFDE, an effective face recognition method L-Fisherfaces is developed. LFDE finds an embedding that preserves local information, and obtains a face subspace that best detects the essential face manifold structure. By defining the locality within-class and between-class scatter matrices and utilizing the Fisher criterion, LFDE searches for the project axes on which the data points of different classes are far from each other while requiring data points of the same class to be close to each other. Experimental results suggest that the proposed L-Fisherfaces method provides a better representation and achieves higher accuracy than PCA, LDA, LPP, and UDP in face recognition.

## REFERENCES


1. R. Chellappa, C. L. Wilson, and S. Sirohey, "Human and machine recognition of faces: a survey," in *Proceedings of the IEEE*, Vol. 83, 1995, pp. 705-740.
2. W. Zhao, R. Chellapa, A. Rosenfeld, and P. J. Phillips, "Face recognition: a literature survey," *ACM Computing Surveys*, Vol. 35, 2003, pp. 399-459.
3. M. Turk and A. Pentland, "Eigenfaces for recognition," *Journal of Cognitive Neuroscience*, Vol. 3, 1991, pp. 71-86.
4. D. L. Swets and J. Weng, "Using discriminant eigenfeatures for image retrieval," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 18, 1996, pp. 831-836.
5. P. N. Belhumeur, J. P. Hespanha, and D. J. Kriengman, "Eigenfaces vs. fisherfaces: Recognition using class specific linear projection," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 19, 1997, pp. 711-720.
6. X. He and P. Niyogi, "Locality preserving projections," *Neural Information Processing Systems*, Vol. 16, 2003, pp. 153-160.
7. X. He, S. Yan, Y. Hu, P. Niyogi, and H. J. Zhang, "Face recognition using laplacianfaces," *Transactions on Pattern Analysis and Machine Intelligence*, Vol. 27, 2005, pp. 328-340.
8. J. Yang, D. Zhang, Z. Jin, and J. Y. Yang, "Unsupervised discriminant projection analysis for feature extraction," in *Proceedings of the 18th International Conference on Pattern Recognition*, Vol. 1, 2006, pp. 904-907.

9. J. Yang, D. Zhang, J. Y. Yang, and B. Niu, "Globally maximizing, locally minimizing: Unsupervised discriminant projection with applications to face and palm biometrics," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 29, 2007, pp. 650-664.
10. M. Belkin and P. Niyogi, "Laplacian eigenmaps for dimensionality reduction and data representation," *Neural Computation*, Vol. 15, 2003, pp. 1373-1396.
11. J. B. Tenenbaum, V. de Silva, and J. C. Langford, "A global geometric framework for nonlinear dimensionality reduction," *Science*, Vol. 290, 2000, pp. 2319-2323.
12. L. K. Saul and S. T. Roweis, "Think globally, fit locally: Unsupervised learning of low dimensional manifolds," *Journal of Machine Learning Research*, Vol. 4, 2003, pp. 119-155.
13. M. Loog, R. P. W. Duin, and R. Haeb-Umbach, "Multiclass linear dimension reduction by weighted pairwise fisher criteria," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 23, 2001, pp. 762-766.
14. J. H. Xuan, Y. B. Dong, J. Khan, E. Hoffman, R. Clarke, and Y. Wang, "Robust feature selection by weighted fisher criterion for multiclass prediction in gene expression profiling," in *Proceedings of the 17th International Conference on Pattern Recognition*, Vol. 2, 2004, pp. 291-294.
15. T. Sim, S. Baker, and M. Bsat, "The CMU pose, illumination, and expression (PIE) database," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 25, 2003, pp. 1615-1618.
16. P. J. Phillips, H. Moon, S. A. Rizvi, and P. J. Rauss, "The FERET evaluation methodology for face-recognition algorithms", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 22, 2000, pp. 1090-1104.

**Cheng-Yuan Zhang (張成元)** received the B.S. in Physics from Nankai University, Tianjin, China, and the M.S. in Computer Science from Shandong University, Jinan, China. Now he is pursuing his Ph.D degree at Institute of Information Science, Beijing Jiaotong University, China. He is currently an Associate Professor in the College of Information and Electrical Engineering, Shandong University of Science and Technology, China. His research interests include image processing, pattern recognition, *etc*.

**Qiu-Qi Ruan (阮秋琦)** was born in 1944. He received the B.S. and M.S. degrees from Northern Jiaotong University, China in1969 and 1981 respectively. From January 1987 to May 1990, he was a visiting scholar in the University of Pittsburgh, and the University of Cincinnati. Subsequently, he has been a Visiting Professor in U.S.A. for several times. He has published 2 books and more than 100 papers, and achieved a national patent. Now he is a Professor, Doctorate Supervisor at the Institute of Information Science, Beijing Jiaotong University. He is a senior member of IEEE. His main research interests include digital signal processing, computer vision, pattern recognition, and virtual reality, *etc*.