A Novel Unsupervised Feature Extraction Based on Image Matrix

Li Yong-zhi[1, 2], Li Zuo-yong[2]
1. School of Information Science & Technology
Nanjing Forestry University
Nanjing, China
E-mail: yzli@njfu.com.cn
2. School of Computer Science & Technology
Nanjing University of Science & Technology
Nanjing, China
E-mail: yzli2008@126.com

Wu Song-song, Yang Jing-yu
School of Computer Science & Technology
Nanjing University of Science & Technology
Nanjing, China

Gu Lin-feng
School of Information Science & Technology
Nanjing Forestry University
Nanjing, China

***Abstract*—By the idea of manifolds learning, this paper presents a new method of dimensionality reduction of high dimensional data. The trait of the method is to exploit image matrixes to directly construct the local scatter matrix and the nonlocal scatter matrix. Its discriminant criterion function is characterized by maximizing the difference between the nonlocal scatter and the local scatter after the samples are projected. The new method is called the two-dimensional marginal discriminant projection (2DMDP). The new discriminant criterion is similar to the maximum margin criterion in form. The criterion main purpose is to find a projection direction (i.e. projection axes) that simultaneously maximizes the nonlocal scatter of projected sample, and minimizes the local scatter of projected sample. The experimental results on YALE face database and ORL face database show that the proposed method outperforms LPP and UDP in terms of recognition rate, and even outperforms LDA when the training sample size per class is small.**

***Keywords*—manifold learning; local scatter matrix; nonlocal scatter matrix; feature extraction; face recognition**

I. INTRODUCTION

Feature extraction is the key to face recognition. The aim of feature extraction is to reduce the dimensionality of face image, so that the extracted features are as representative as possible. Among feature extraction methods, PCA [1][2] and FDA [3] is the most popular and relatively effective methods in face recognition. However, both PCA and FDA can only see the Euclidean structure effectively, and they fail to discover the underlying structure, if the face images lie on a nonlinear manifold $\Re$ hidden in the image space. In recent years, some nonlinear techniques have been proposed to discover the nonlinear structure of the manifold, i.e. ISOMAP[4], LLE [5][6] and Laplacian Eigenmap [7]. These nonlinear methods do yield impressive results on some artificial data sets. However, they yield maps that are defined only on the training samples and it is unclear how to evaluate the maps for new test samples. Thus, these nonlinear manifold learning techniques might not be suitable for pattern recognition. Recently, He et al. [8][9] proposed locality preserving projection (LPP), which is a linear subspace learning method derived from Laplacian Eigenmap. In contrast to most manifold learning algorithms, LPP possesses the remarkable advantage that it can generate an explicit map. This map is linear and easily computable, like that of PCA or FDA. It is also effective encouraging results on face recognition tasks.

The PCA aims to preserve the global structure of the image space, the FDA aims to preserve the discriminating information, and the LPP aims to preserve the local structure of the image space. In many real-world classification problems, the local manifold structure is more important than the global Euclidean structure, especially when nearest neighbor classifier is used for classification. LPP shares some similar properties to LLE, such as a locality preserving character. Moreover, LPP is defined everywhere. So LPP may be simply applied to any new test sample to locate it in the reduced feature space. Like most manifold learning algorithms, LPP has the weakness of having no direct connection to classification. The objective function of LPP is to minimize the local scatter of the projected samples. In some cases, this criterion cannot be guaranteed to yield a good projection for classification purposes. For weakness of LPP, Yang et al. proposed an unsupervised discriminant projection (UDP)[10]. The purpose of UDP will draw the close samples closer together while simultaneously making the mutually distant samples even more distant from each other. The criterion function of UDP is to maximize the ratio of the nonlocal scatter to the local scatter, i.e. to simultaneously maximize the nonlocal scatter and minimize the local scatter.

In this paper, by the idea of manifolds learning, we consider two quantities, local scatter and nonlocal scatter at the same time in the modeling process for classification purposes. The method trait is to exploit image matrixes to directly construct local scatter matrix and nonlocal scatter matrix. Its discriminant criterion function is characterized by maximizing the difference between the nonlocal scatter and the local scatter after the samples are projected. The new method is called the two-dimensional marginal discriminant projection (2DMDP). The criterion is similar to the maximum margin criterion (MMC) [11] in form. The purpose of the criterion is to maximize the nonlocal scatter while simultaneously minimizing the local scatter after the projection. Thus, it is not hard to find its optimal solutions by solving a generalized eigen-equation. The experimental results on YALE face database and ORL face database show that the proposed method outperforms LPP and

UDP in terms of recognition rate, and even outperforms LDA when the training sample size per class is small.

II. Global Scatter, Local Scatter and Nonlocal Scatter

A. The Global Scatter based on image matrix

Let $A_1, A_2, \cdots, A_m$ in $R^{r\times n}$ be m training sample images, and vectors $y_1, y_2, \cdots y_m$ are a group of projected sample points in R^r by transformation $y_i = A_i\alpha$ ($\alpha \in R^n$). vector α is called **projection axis (or optimal discriminant vector)**. The global scatter of projected sample can be characterized by the mean square of the Euclidean distance between any pair of the projected sample points, i.e.

$$J_T(\alpha) = \frac{1}{2}\frac{1}{m^2}\sum_{i=1}^{m}\sum_{j=1}^{m}\| y_i - y_j \|^2 \quad (1)$$

It follows that

$$\begin{aligned} J_T(\alpha) &= \frac{1}{2}\frac{1}{m^2}\sum_{i=1}^{m}\sum_{j=1}^{m}(A_i\alpha - A_j\alpha)^T(A_i\alpha - A_j\alpha) \\ &= \frac{1}{2}\frac{1}{m^2}\sum_{i=1}^{m}\sum_{j=1}^{m}\alpha^T(A_i - A_j)^T(A_i - A_j)\alpha \\ &= \alpha^T\left[\frac{1}{2}\frac{1}{m^2}\sum_{i=1}^{m}\sum_{j=1}^{m}(A_i - A_j)^T(A_i - A_j)\right]\alpha \quad (2) \end{aligned}$$

Let us denote $S_T^M = \frac{1}{2}\frac{1}{m^2}\sum_{i=1}^{m}\sum_{j=1}^{m}(A_i - A_j)^T(A_i - A_j)$ and mean matrix $\overline{A} = \frac{1}{m}\sum_{i=1}^{m} A_i$.

Then, it can be easily proven

$$\begin{aligned} S_T^M &= \frac{1}{2}\frac{1}{m^2}\sum_{i=1}^{m}\sum_{j=1}^{m}(A_i - A_j)^T(A_i - A_j) \\ &= \frac{1}{2}\frac{1}{m^2}\sum_{i=1}^{m}\sum_{j=1}^{m}(A_i^T A_i - 2A_i^T A_j + A_j^T A_j) \\ &= \frac{1}{m^2}\left(m\sum_{i=1}^{m} A_i^T A_i - (\sum_{i=1}^{m} A_i^T)(\sum_{j=1}^{m} A_j)\right) \\ &= \frac{1}{m}\sum_{i=1}^{m}(A_i - \overline{A})^T(A_i - \overline{A}) . \quad (3) \end{aligned}$$

So, matrix S_T^M is called **the global scatter matrix based on image matrix.**

Thus, (1) can be rewritten by

$$J_T(\alpha) = \alpha^T S_T^M \alpha \quad (4)$$

The projection axis α that maximizes (4) can be selected as the eigen-vector of S_T^M corresponding to the largest eigen-value. Similarly, we can obtain a set of projection axes of PCA by selecting the d eigenvectors of S_T^M corresponding to the first d largest eigen-values.

B. Local Scatter based on image matrix

In order to preserve the global geometric structure of sample in a transformed low dimensional space, PCA seeks to find a group of projection axes so that the global scatter is maximized after the projection of samples. Correspondingly, if we aim to discover the local structure of data, we should take account of the local scatter of samples.

Given m sample images $A_1, A_2, \cdots, A_m$ sampled from the underlying manifold $\Re$. Let $N_k(A_i) = \{A_i^1, A_i^2, \cdots, A_i^k\}$ denote k-nearest neighbors of A_i . Thus, each sample A_i , we can find its k-nearest neighbors. The local scatter can be characterized by the mean square of the Euclidean distance between any pair of the projected sample points that are within any local k-nearest neighbors.

Let a set $U^k = \{(i,j) \mid A_i \in N_k(A_j) \ \ and \ \ A_j \in N_k(A_i)\}$. Let y_i and y_j denote the projection feature vectors of image samples A_i , A_j by projection transformation $y = A\alpha$. The local scatter is defined by

$$J_L(\alpha) = \frac{1}{2m_L}\sum_{(i,j)\in U^k}\| y_i - y_j \|^2 \propto \frac{1}{2}\frac{1}{m^2}\sum_{(i,j)\in U^k}\| y_i - y_j \|^2 \quad (5)$$

Where m_L is the number of sample pair satisfying $(i,j) \in U^k$. Let us define an adjacency matrix $W = (w_{ij})_{m\times m}$, whose elements are given below

$$w_{ij} = \begin{cases} 1 & x_i \in N_k(A_j) \ \ and \ \ x_j \in N_k(A_i) \\ 0 & \text{otherwise.} \end{cases} \quad (6)$$

It is obvious that the adjacency matrix W is a symmetric matrix. By virtue of the adjacency matrix W , (5) can be rewritten by

$$\begin{aligned} J_L(\alpha) &= \frac{1}{2}\frac{1}{m^2}\sum_{i=1}^{m}\sum_{j=1}^{m} w_{ij}\| y_i - y_j \|^2 \\ &= \frac{1}{2}\frac{1}{m^2}\sum_{i=1}^{m}\sum_{j=1}^{m} w_{ij}(A_i\alpha - A_j\alpha)^T(A_i\alpha - A_j\alpha) \\ &= \alpha^T\left[\frac{1}{2}\frac{1}{m^2}\sum_{i=1}^{m}\sum_{j=1}^{m} w_{ij}(A_i - A_j)^T(A_i - A_j)\right]\alpha = \alpha^T S_L^M \alpha \quad (7) \end{aligned}$$

Where $S_L^M = \frac{1}{2}\frac{1}{m^2}\sum_{i=1}^{m}\sum_{j=1}^{m} w_{ij}(A_i - A_j)^T(A_i - A_j)$, matrix S_L^M is called **the local scatter matrix based on image matrix.**

$$\begin{aligned} S_L^M &= \frac{1}{2}\frac{1}{m^2}\sum_{i=1}^{m}\sum_{j=1}^{m} w_{ij}(A_i - A_j)^T(A_i - A_j) \\ &= \frac{1}{2}\frac{1}{m^2}\sum_{i=1}^{m}\sum_{j=1}^{m} w_{ij}(A_i^T A_i - 2A_i^T A_j + A_j^T A_j) \\ &= \frac{1}{m^2}\left(\sum_{i=1}^{m}(\sum_{j=1}^{m} w_{ij})A_i^T A_i - \sum_{i=1}^{m}\sum_{j=1}^{m} w_{ij}A_i^T A_j\right) \\ &= \frac{1}{m^2}\left(X^T DX - X^T WX\right) = \frac{1}{m^2}X^T LX \quad (8) \end{aligned}$$

Where

$$X=\begin{pmatrix} A_1 \\ A_2 \\ \vdots \\ A_m \end{pmatrix}, \quad D=\begin{pmatrix} d_1 I & 0 & \cdots & 0 \\ 0 & d_2 I & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & \cdots & d_m I \end{pmatrix}_{mr\times mr}, \quad d_i=\sum_{j=1}^{m} w_{ij},$$

And $W=\begin{pmatrix} w_{11}I & w_{12}I & \cdots & w_{1m}I \\ w_{21}I & w_{22}I & \cdots & w_{2m}I \\ \cdots & \cdots & \cdots & \cdots \\ w_{m1}I & w_{m2}I & \cdots & w_{mm}I \end{pmatrix}_{mr\times mr}$, $I\in R^{r\times r}$ is the identity matrix, matrix $L=D-W$ is a Laplacian matrix. It is obvious that S_L^M are both real symmetric matrices. From (8), we know that $\alpha^T S_L^M \alpha \geq 0$ for any nonzero vector α. So, the local scatter matrix S_L^M must be nonnegative definite.

C. Nonlocal Scatter matrix based on image matrix

The nonlocal scatter of projected sample can be characterized by the mean square of the Euclidean distance between any pair of the projected sample points that are outside any local k-nearest neighbors. The nonlocal scatter is defined by

$$J_N(\alpha)=\frac{1}{2m_N}\sum_{(i,j)\notin U^k}\|y_i-y_j\|^2 \propto \frac{1}{2}\frac{1}{m^2}\sum_{(i,j)\notin U^k}\|y_i-y_j\|^2 \quad (9)$$

Where m_N is the number of sample pair satisfying $(i,j)\notin U^k$. By virtue of the adjacency matrix $W=(w_{ij})_{m\times m}$, then (9) can be rewritten by

$$\begin{aligned} J_N(\alpha) &= \frac{1}{2}\frac{1}{m^2}\sum_{i=1}^{m}\sum_{j=1}^{m}(1-w_{ij})\|y_i-y_j\|^2 \\ &= \frac{1}{2}\frac{1}{m^2}\sum_{i=1}^{m}\sum_{j=1}^{m}(1-w_{ij})(A_i\alpha-A_j\alpha)^T(A_i\alpha-A_j\alpha) \\ &= \alpha^T\left[\frac{1}{2}\frac{1}{m^2}\sum_{i=1}^{m}\sum_{j=1}^{m}(1-w_{ij})(A_i-A_j)^T(A_i-A_j)\right]\alpha \\ &= \alpha^T S_N^M \alpha \end{aligned} \quad (10)$$

Where $S_N^M=\frac{1}{2}\frac{1}{m^2}\sum_{i=1}^{m}\sum_{j=1}^{m}(1-w_{ij})(A_i-A_j)^T(A_i-A_j)$, matrix S_N^M is called **the nonlocal scatter matrix based on image matrix.**

$$\begin{aligned} S_N^M &= \frac{1}{2}\frac{1}{m^2}\sum_{i=1}^{m}\sum_{j=1}^{m}(1-w_{ij})(A_i-A_j)^T(A_i-A_j) \\ &= \frac{1}{2}\frac{1}{m^2}\sum_{i=1}^{m}\sum_{j=1}^{m}(A_i-A_j)^T(A_i-A_j)- \\ &\frac{1}{2}\frac{1}{m^2}\sum_{i=1}^{m}\sum_{j=1}^{m}w_{ij}(A_i-A_j)^T(A_i-A_j)=S_T^M-S_L^M \end{aligned} \quad (11)$$

Thus

$$S_T^M=S_L^M+S_N^M,\ J_T(\alpha)=J_L(\alpha)+J_N(\alpha) \quad (12)$$

III. TWO-DIMENSIONAL MARGINAL DISCRIMINANT PROJECTION (2DMDP)

The objective function of LPP is actually to minimize the local scatter, i.e.

$$\begin{aligned} &\min_{\alpha} J_L(\alpha)=\alpha^T S_L \alpha=\alpha^T XLX^T\alpha \\ &s.t.\ \ \alpha^T XDX^T\alpha=1 \end{aligned} \quad (13)$$

Given x_i and x_j are sample vectors, obviously, the projection transformation determined by LPP can ensure that, if samples x_i and x_j are close, their projections y_i and y_j are close as well. But, LPP cannot guarantee that, if samples x_i and x_j are not close, their projections y_i and y_j are not either. This means that it may happen that two mutually distant samples belonging to different classes may result in close feature vectors after the projection of LPP.

Yang et al.[10] proposed an unsupervised discriminant projection (UDP), which can be seen as a linear approximation of a manifold learning framework that takes into account both the local and nonlocal quantities. Purpose of UDP will draw the close samples closer together while simultaneously making the mutually distant samples even more distant from each other. The criterion function of UDP is to maximize the ratio of the nonlocal scatter to the local scatter, i.e.

$$\max_{\alpha} J_U(\alpha)=\frac{\alpha^T S_N \alpha}{\alpha^T S_L \alpha} \quad (14)$$

But, one is confronted with the difficulty that the local scatter matrix S_L is sometimes singular in the face recognition problem. Because sometimes the number of the training samples is much smaller than the dimension of training sample vectors. To avoid the complication of a singular local scatter matrix, we are to exploit image matrixes to directly construct local scatter matrix and nonlocal scatter matrix and propose a discriminant criterion, which is to maximize the difference between the nonlocal scatter and the local scatter based on image matrix, this criterion function is similar to the maximum margin criterion in form [11]. So, the method is called **the two-dimensional marginal discriminant projection (2DMDP)**. The purpose of the criterion is to maximize the nonlocal scatter while simultaneously minimizing the local scatter after the projection. We can obtain just such a projection axis α by maximizing the following criterion:

$$\begin{aligned} &\max_{\alpha} J_M(\alpha)=J_N(\alpha)-J_L(\alpha)=\alpha^T(S_N^M-S_L^M)\alpha \\ &s.t.\ \ \alpha^T\alpha=1,\ i=1,2,\cdots,d \end{aligned} \quad (15)$$

Theorem1 Optimal projection axes $\alpha_1,\alpha_2,\cdots,\alpha_d$ based on 2DMDP are d unit eigenvectors of matrix $S_N^M-S_L^M$, corresponding to the first d largest positive eigenvalues.

From theorem1, we can denote that 2DMDP overcomes the “small sample size” problem by using a new discriminant criterion to replace Yang Jian et.al. proposed UDP.

IV. 2DMDP WITH KERNEL WEIGHING

In this section, we will build a kernel-weighted version of 2DMDP. We know that Laplacian Eigenmap and LPP use kernel coefficients to weight the edges of the adjacency graph, where a heat kernel (Gaussian kernel) is defined by

$$k(A_i, A_j) = \exp(-\gamma \| A_i - A_j \|_F^2) \tag{16}$$

Obviously, for any A_i , A_j and parameter γ , $0 < k(A_i, A_j) \le 1$ always holds. Further, the kernel function is a strictly monotone decreasing function with respect to the distance between two variables A_i and A_j. The purpose of the kernel weighting is to indicate the degree of A_i and A_j belonging to a local k-neighborhood. If the smaller the distance were, the larger the degree would be. Otherwise, the degree is zero. The kernel weighting, like other similar weightings, may be helpful in alleviating the effect of the outliers on the projection directions of the linear models and thus, makes these models more robust to outliers. If we redefine the adjacency matrix $W = (w_{ij})_{m \times m}$ as

$$w_{ij} = \begin{cases} k_{ij} & A_i \in N_k(A_j) \ \ and \ \ A_j \in N_k(A_i) \\ 0 & ortherwise \end{cases} \tag{17}$$

Let $k_{ij} = k(A_i, A_j)$, the kernel-weighted global scatter can be characterized by

$$\begin{aligned} J_T(\alpha) &= \frac{1}{2}\frac{1}{m^2}\sum_{i=1}^{m}\sum_{j=1}^{m} k_{ij} \| y_i - y_j \|^2 \\ &= \frac{1}{2}\frac{1}{m^2}\sum_{i=1}^{m}\sum_{j=1}^{m} k_{ij} (A_i\alpha - A_j\alpha)^T (A_i\alpha - A_j\alpha) \\ &= \frac{1}{2}\frac{1}{m^2}\sum_{i=1}^{m}\sum_{j=1}^{m} k_{ij} \alpha^T (A_i - A_j)^T (A_i - A_j)\alpha \\ &= \alpha^T \left[\frac{1}{2}\frac{1}{m^2}\sum_{i=1}^{m}\sum_{j=1}^{m} k_{ij} (A_i - A_j)^T (A_i - A_j) \right] \alpha \ = \alpha^T S_T^M \alpha \end{aligned} \tag{18}$$

Where $S_T^M = \frac{1}{2}\frac{1}{m^2}\sum_{i=1}^{m}\sum_{j=1}^{m} k_{ij} (A_i - A_j)^T (A_i - A_j)$, matrix S_T^M is called **the kernel weighted global scatter matrix based on image matrix.**

$$\begin{aligned} S_L^M &= \frac{1}{2}\frac{1}{m^2}\sum_{i=1}^{m}\sum_{j=1}^{m} w_{ij} (A_i - A_j)^T (A_i - A_j) \\ &= \frac{1}{m^2}\left(X^T D X - X^T W X \right) = \frac{1}{m^2} X^T L X \end{aligned} \tag{19}$$

Matrix S_L^M is called **the kernel weighted local scatter matrix based on image matrix.**

The kernel-weighted nonlocal scatter is characterizerized by

$$\begin{aligned} J_N(\alpha) &= \frac{1}{2}\frac{1}{m^2}\sum_{i=1}^{m}\sum_{j=1}^{m} (k_{ij} - w_{ij}) \| y_i - y_j \|^2 \\ &= \frac{1}{2}\frac{1}{m^2}\sum_{i=1}^{m}\sum_{j=1}^{m} (k_{ij} - w_{ij})(A_i\alpha - A_j\alpha)^T (A_i\alpha - A_j\alpha) \\ &= \alpha^T \left[\frac{1}{2}\frac{1}{m^2}\sum_{i=1}^{m}\sum_{j=1}^{m} (k_{ij} - w_{ij})(A_i - A_j)^T (A_i - A_j) \right] \alpha \\ &= \alpha^T S_N^M \alpha \end{aligned} \tag{20}$$

Where $S_N^M = \frac{1}{2}\frac{1}{m^2}\sum_{i=1}^{m}\sum_{j=1}^{m} (k_{ij} - w_{ij})(A_i - A_j)^T (A_i - A_j)$, matrix S_N^M is called **the kernel weighted nonlocal scatter matrix based on image matrix.**

$$S_N^M = \frac{1}{2}\frac{1}{m^2}\sum_{i=1}^{m}\sum_{j=1}^{m} (k_{ij} - w_{ij})(A_i - A_j)^T (A_i - A_j) \qquad = S_T^M - S_L^M \tag{21}$$

Thus

$$S_T^M = S_L^M + S_N^M \tag{22}$$

Then criterion function of 2DMDP with kernel weighing is defined by

$$\begin{aligned} & \max_{\alpha} J_M(\alpha) = \alpha^T (S_N^M - S_L^M)\alpha \\ & s.t. \ \ \alpha^T \alpha = 1, \ \ i = 1, 2, \cdots, d \end{aligned} \tag{23}$$

V. EXPERIMENT AND ANALYSIS

In this section, the performance of 2DMDP is evaluated on the Yale face image databases and ORL face image databases and compared with the performances of LDA, LPP, UDP and 2DMDP.

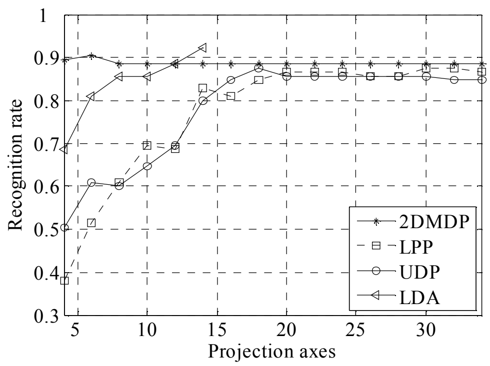
A. Experiment 1

The Yale face database contains 165 images of 15 individuals (each person providing 11 different images) under various facial expressions and lighting conditions. The size of each image is 92×112 pixels, with 256 gray-levels. Figure.1 shows some images of one person in the YALE face database.

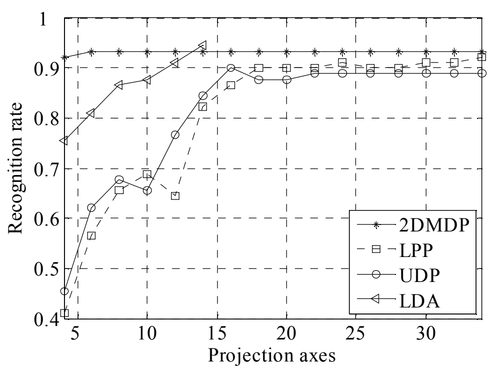
Figure 1. Some images of one person in YALE face database

The experiment was performed using the first four (or first five) images per class for training samples, and the remaining seven (or six) images for testing samples. In experiments, we used, respectively, LDA, LPP, UDP and 2DMDP for feature extraction. The k-nearest neighborhood parameter k in LPP, UDP and 2DMDP can be chosen as $k = l - 1$, where l denotes the number of training samples per class. The justification for this choice is that each sample should connect with the remaining $l-1$ samples of the same class provided that within-class samples are well clustered in the observation. In experiments, LDA, LPP and UDP all involve a PCA phase. In this phase, we keep nearly 98 percent image energy and select the number of principal components, and then we used, respectively, LDA, LPP and UDP for feature extraction. We use the nearest neighbor strategy for classification based on

Euclidean measure. Figure 2 shows recognition rate of LDA, LPP, UDP and 2DMDP respectively versus the number of projection axes. Figure 2(a) chooses four training samples in each class; figure 2(b) chooses five training samples in each class. From figure 2, we can see: no matter we chose four or five training samples in each class, we can see that 2DMDP almost always achieves the highest recognition rate and more stable as the number of projection axes is varying from 4 to 34 (i.e. the number of projection axes is varying among 4, 6, 8, … 32, 34). Recognition rate of LDA is higher than recognition rate of LPP and UDP. From figure 2, it is demonstrated that 2DMDP is the most efficient among four methods.



(a) The first four samples per class are used for training



(b) The first five samples per class are used for training

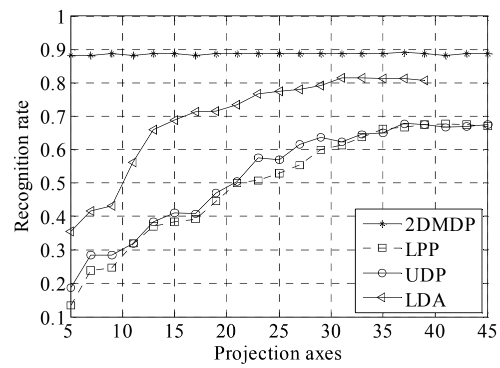
Figure 2. On the YALE database, recognition rate of LDA, LPP, UDP and 2DMDP versus the number of projection axis when Euclidean measure is used

B. *Experiment 2*

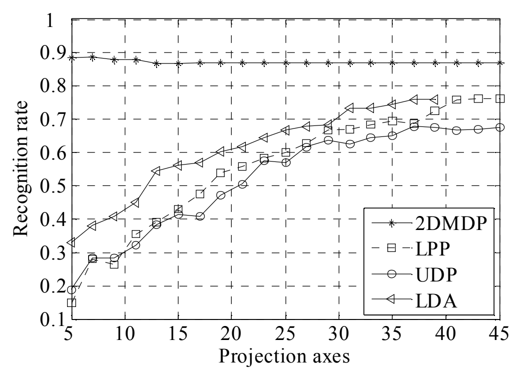
The proposed method is tested on the ORL face database. In the ORL face database, there are 10 different images of 40 distinct subjects; there are variations in facial expression (open/closed eyes, smiling/non-smiling) and facial details (glasses/no glasses). All the images are taken against a dark homogeneous background with the subjects in an up-right, frontal position, with tolerance for some tilting and rotation of up to about 20^0. There is some variation in scale of up to about 10%. The size of each image is 92×112, with 256 gray-levels. Some images of one person are shown in Figure3. In this experiment, the experiment was performed using the first four (or first five) images per class for training samples, and the remaining seven (or six) images for testing samples. For feature extraction, we used, respectively, LDA, LPP, UDP and

Figure 3. shows sample images of one person in ORL face database

the proposed 2DMDP. The k-nearest neighborhood parameter k in LPP, UDP and 2DMDP can be chosen as $k = l - 1$, where l denotes the number of training samples per class. In experiments, LDA, LPP and UDP all involve a PCA phase. In this phase, we keep nearly 98 percent image energy and select the number of principal components, and then we used, respectively, LDA, LPP and UDP for feature extraction. We use the nearest neighbor strategy for classification based on Euclidean measure and Cosine measure respectively. Figure 4 and figure5 shows the recognition rate of LDA, LPP, UDP and



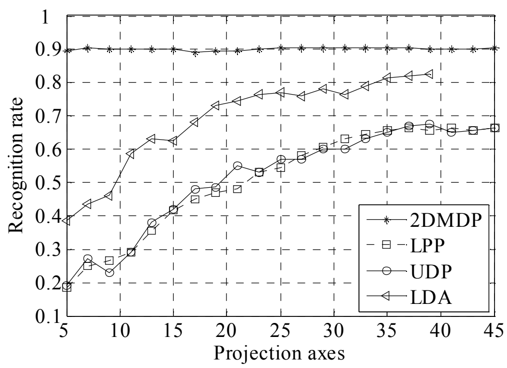
(a) Euclidean measure



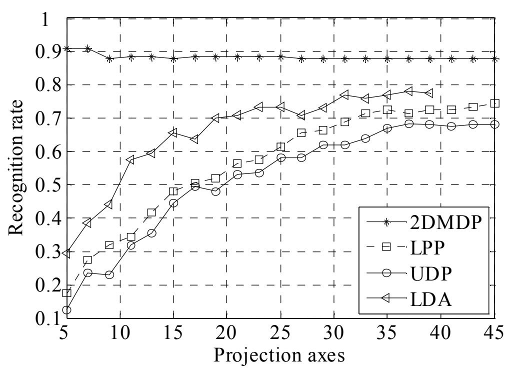
(b) Cosine measure

Figure 4. On the ORL database, recognition rate of LDA, LPP, UDP and 2DUDP versus the number of projection axis when the four samples per class are used

2DMDP versus the number of projection axes. Figure 4 chooses four training samples in each class; Figure5 chooses five training samples in each class. Euclidean measure is used on figure 4(a) and figure 5(a), and Cosine measure is used on figure 4(b) and figure 5(b). From figure 4 and figure 5, we can see: no matter we choose four and five training samples in each class, we can see that 2DMDP always achieves the highest recognition rate and more stable as the number of projection axes is varying from 5 to 45 (i.e. the number of projection axes is varying among 5, 7, 9, … 43, 45). And then recognition rate of LDA almost always is higher than recognition rate of LPP and UDP. From figure 4 and figure 5, it is demonstrated that 2DMDP is the most efficient among four discriminant methods.



(a) Euclidean measure



(b) Cosine measure

Figure 5. On the ORL database, recognition rate of LDA, LPP, UDP and 2DUDP versus the number of projection axis when the five samples per class are used

VI. CONCLUSIONS

Based on manifolds learning, this paper presents a new method of dimensionality reduction of high dimensional data. In this paper, we consider two quantities; local scatter and nonlocal scatter after the samples are projected in the modeling process for classification purposes. The method aim is to preserve the local structure of the image space. Its trait is to exploit image matrixes to directly construct local scatter matrix and nonlocal scatter matrix. Its discriminant criterion function is characterized by maximizing the difference between the nonlocal scatter and the local scatter after the samples are projected. The method is called the two-dimensional marginal discriminant projection (2DMDP). The main purpose of 2DMDP is to simultaneously maximize the nonlocal scatter of projected sample, and minimize the local scatter of projected sample. The experimental results on YALE face database and ORL face database show that the proposed method outperforms LPP and UDP in terms of recognition rate, and even outperforms LDA.

ACKNOWLEDGMENT

This work was supported by National Science Foundation of China under Grant (No. 60472060, No. 60632050). National Science Foundation of Jiangsu college and university (No. 06KJD520085) and Innovation Foundation of Nanjing Forestry University (No.2002-10)

REFERENCES

[1] M. Turk and A. Pentland, “Face Recognition Using Eigenfaces ”, Proc. IEEE Conf. on Computer Vision and Pattern Recognition, pp: 586-591, 1991.

[2] Peter N. Belhumeur, Joao P. Hespanha, and David J. Kriegman. “Eigenfaces vs. Fisherfaces: Recognition Using Class Specific Linear Projection”, IEEE Transactions on pattern analysis and machine intelligence, Vol.19(7), pp: 711—720. 1997

[3] L.F. Chen, H.Y. M. Liao, J.C. Lin, M.D. Kao, and G.J. Yu, “A New LDA- Based Face Recognition System which Can Solve the Small Sample Size Problem,” Pattern Recognition, vol. 33(10),pp. 1713-1726, 2000

[4] J.B. Tenenbaum, V. de Silva, and J.C. Langford, “A Global Geometric Framework for Nonlinear Dimensionality Reduction,” Science, vol. 290, pp: 2319-2322, 2000.

[5] S.T. Roweis and L.K. Saul, “Nonlinear Dimensionality Reduction by Locally Linear Embedding,” Science, vol. 290, pp:2323-2326, 2000.

[6] L.K. Saul and S.T. Roweis, “Think Globally, Fit Locally: Unsupervised Learning of Low Dimensional Manifolds,” J. Machine Learning Research, vol. 4, pp: 119-155, 2003.

[7] M. Belkin and P. Niyogi, “Laplacian Eigenmaps for Dimensionality Reduction and Data Representation,” Neural Computation, vol. 15, no. 6, pp. 1373-1396, 2003.

[8] X. He and P. Niyogi, “Locality Preserving Projections,” Proc. 16th Conf. Neural Information Processing Systems, 2003.

[9] X. He, S. Yan, Y. Hu, P. Niyogi, and H.-J. Zhang, “Face Recognition Using Laplacianfaces,” IEEE Trans. Pattern Analysis and Machine Intelligence, vol. 27(3), pp. 328-340, 2005.

[10] Jian Yang, David Zhang, Jing-yu Yang, and Ben Niu, “Globally Maximizing, Locally Minimizing: Unsupervised Discriminant Projection with Applications to Face and Palm Biometrics” IEEE Trans. Pattern Analysis and Machine Intelligence, vol. 29(4), pp. 650-664, 2007.

[11] Y. Z. Li, M. Feng, J.Y. Yang et al., “An Efficient Method of Nonlinear Feature Extraction Based on SVM.” IEEE 9th International Conference Control, Automation, Robotics and Vision(ICARCV2006)}. pp.216 - 221, 2006