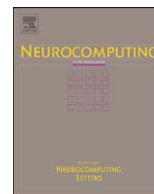




ELSEVIER

Contents lists available at ScienceDirect

Neurocomputing

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

One improvement to two-dimensional locality preserving projection method for use with face recognition

Yong Xu^{a,*}, Ge Feng^a, Yingnan Zhao^b

^a Bio-Computing Research Center, Shenzhen Graduate School, Harbin Institute of Technology, China

^b School of Computer and Software, Nanjing University of Information Science & Technology, China

ARTICLE INFO

Article history:

Received 12 April 2009

Received in revised form

31 August 2009

Accepted 5 September 2009

Communicated by J. Yang

Available online 26 September 2009

Keywords:

Locality preserving projection

Two-dimensional locality preserving projection

Feature extraction

Dimension reduction

Face recognition

ABSTRACT

While locality preserving projection (LPP) is directly applicable to only vector data, two-dimensional locality preserving projection (2DLPP) is directly applicable to two-dimensional data. As a result, 2DLPP is computationally more efficient than LPP. On the other hand, when determining the transform axes, both conventional 2DLPP and LPP do not exploit the class label information of training samples, the use of which is usually advantageous for producing good classification result. In order to exploit the class label information, we proposed one novel LPP method, i.e. two-dimensional discriminant supervised LPP (2DDSLPP). We also analyzed the characteristics and advantages of 2DDSLPP and presented the difference and relationship between 2DDSLPP and other methods. Compared with two-dimensional discriminant LPP (2DDLPP), 2DDSLPP has a stronger capability to preserve the distance relation of samples from different classes. We used two face databases to test 2DDSLPP and several other two-dimensional dimensionality reduction methods. Experimental results show that 2DDSLPP can obtain a higher classification right rate.

© 2009 Elsevier B.V. All rights reserved.

1. Introduction

In the field of face recognition, many dimensionality reduction methods, such as principal component analysis (PCA) [1], linear discriminant analysis (LDA) [2,3], and locality preserving projection (LPP) [4–10], have been developed over the past few decades. All these methods reduce the dimension of original data by transforming the data into a lower-dimensional space. PCA can be considered as the best data representation method. This is because if the transform results obtained using different methods are, respectively, exploited to reconstruct the original data, the mean squared error between the original data and the data reconstructed using the PCA transform result is the minimum. LDA is a supervised dimensionality reduction method that seeks the following transform axes: the transform results of the data points of different classes will be far as much as possible from each other and the transform results of the data points of the same class will be close as much as possible to each other. We note that when transforming samples into the new space, both PCA and LDA do not take into account whether the local structure of samples is preserved. On the other hand, in many classification applications

such as in the application where the nearest neighbor classifier is used, to preserve the local structure information is also important. LPP is a locality structure preserving method that tries to preserve the intrinsic geometry of the samples. LPP also appears not to be sensitive to noise and outliers [4].

Though LPP has been applied in many domains, it still has limitations. For example, when applied to recognition problems, conventional LPP has the following shortcoming: since conventional LPP is an unsupervised method, when exploiting the training samples to solve the transform axes it cannot exploit the class label of the training samples, the use of which is usually advantageous for producing good classification result. Under this circumstance, the supervised locality preserving projection (SLPP) was proposed. Since SLPP takes advantage of the class label information, it usually obtains a higher classification accuracy than conventional LPP [5]. Based on the idea of the linear discriminant analysis, researchers also proposed the discriminant locality preserving projection (DLPP) [6] and the orthogonal discriminant locality preserving projection method (ODLPP) [11].

When we apply conventional LPP, SLPP, DLPP and ODLPP to two-dimensional data such as images, we must transform these data into one-dimensional vectors in advance. The resulting vectors usually lead to a high-dimensional vector space and covariance matrix, which brings a large computational burden. This also usually causes the small sample size (SSS) problem, which means that the eigenvalue equation corresponding to LPP cannot be directly solved because of the matrix singularity [12]. In

* Corresponding author at: Department of Computer Science & Technology, Shenzhen Graduate School, Harbin Institute of Technology, Shenzhen, Guangdong 518055, China. Tel./fax: +86 25 84315510.

E-mail address: laterfall286@yahoo.com (Y. Xu).

order to overcome this problem, PCA has ever been used as a preprocessing step of conventional LPP [7]. However, we note that the goal and essence of LPP is indeed quite different from PCA. This implies that the use of PCA actually will offend the locality preserving projection goal of LPP. In this paper we refer to all vector-based LPP methods including LPP, SLPP, DLPP and ODLPP as one-dimensional LPP.

In order to make the locality preserving projection methodology directly applicable to matrix data, researchers also proposed two-dimensional LPP (2DLPP) [8,9]. Compared with one-dimensional LPP, 2DLPP has several advantages such as that it can avoid the SSS problem and reduce the computational cost. On the other hand, 2DLPP is still an unsupervised method.

In this paper, we propose one improvement to 2DLPP. We refer to it as two-dimensional discriminant supervised LPP (2DDSLPP). 2DDSLPP tries to obtain the projection space in which data points of the same class will preserve their original neighbor relationship. In the projection space obtained using 2DDSLPP, while the data points from different classes are far from each other, two data points from two different classes are also partially subject to the following constraint: the larger the distance between the original two samples is, the larger the distance between the two data points should be. Comparative experiments on 2DPCA [13,14], 2DLPP, 2DDLPP [10], 2DLDA and 2DDSLPP show that 2DDSLPP can obtain a high classification accuracy.

This paper is organized as follows: Section 2 provides a brief review of 2DLPP. Section 3 introduces the proposed improvement to 2DLPP. Section 4 presents the comparison analysis on 2DDSLPP and 2DDLPP as well as 2DLDA. Section 5 describes the experimental results. Finally Section 6 offers the conclusion.

2. Two-dimensional locality preserving projection (2DLPP)

Suppose the image sample set is $\{X_1, X_2, \dots, X_M\}$, where X_i is the matrix representation of the i th image sample. The size of each sample is $m \times n$. The transform axis is denoted as \mathbf{a} and the transform result is $Y_i = \mathbf{a}^T X_i$. It should be pointed out that the transform result Y_i obtained using one transform axis of 2DLPP is a vector, whereas the transform result obtained using one transform axis of one-dimensional LPP is a scalar. The set $\{Y_1, Y_2, \dots, Y_M\}$ forms the sample projection space. The objective function of 2DLPP is

$$\min \sum_{ij} \|Y_i - Y_j\|^2 W_{ij}, \quad (1)$$

where $\|\cdot\|$ means the l_2 norm, W is the similarity matrix. Entries of W are defined as follows: if X_i is among the k -nearest-neighbors of X_j or X_j is among the k nearest-neighbors of X_i , then $W_{ij} = \exp(-\|X_i - X_j\|^2/t)$; otherwise, $W_{ij} = 0$. W is indeed a symmetric matrix. Parameter t is set to a positive real number. According to [4], we can obtain

$$\frac{1}{2} \sum_{ij} \|Y_i - Y_j\|^2 W_{ij} = \mathbf{a}^T X(L \otimes I_n) X^T \mathbf{a}, \quad (2)$$

where $X = [X_1 \ X_2 \ \dots \ X_M]$ is the sample matrix that consists of the matrices of all the samples, and D is a diagonal matrix defined as $D_{ii} = \sum_j W_{ij}$, $L = D - W$, \otimes represents the Kronecker product. The Kronecker product can be briefly illustrated as follows. If A is an m by n matrix and B is a p by q matrix, then the result of $A \otimes B$ is the mp by nq matrix

$$E = A \otimes B = \begin{Bmatrix} a_{11}B & \dots & a_{1n}B \\ \vdots & \ddots & \vdots \\ a_{m1}B & \dots & a_{mn}B \end{Bmatrix}.$$

We can use the constraint $\mathbf{a}^T X(D \otimes I_n) X^T \mathbf{a} = 1$ (I_n is an n order identity matrix) to remove an arbitrary scaling factor of the

transform axis [7]. Under this constraint, we know that the optimal transform axis of 2DLPP will be the eigenvector corresponding to the smallest eigenvalue of the following eigenvalue equation:

$$X(L \otimes I_n) X^T \mathbf{a} = \lambda X(D \otimes I_n) X^T \mathbf{a}. \quad (3)$$

3. Improvement to 2DLPP

For image applications, since the matrices in Eq. (3) has a much lower dimensionality than those in the eigenvalue equation of the one-dimensional LPP, 2DLPP usually does not encounter the small sample size (SSS) problem. On the other hand, 2DLPP is still an unsupervised method. This means that 2DLPP neglects the class label information of training samples when exploiting them to solve its transform axes. Suppose that there are two samples, respectively, from two classes and one sample is among the k neighbors of the other sample (this might occur in the border of two classes), then 2DLPP will transform them into close points, which is not helpful for classification. In order to overcome this shortcoming of 2DLPP, we propose the two-dimensional discriminant supervised LPP, i.e. 2DDSLPP.

3.1. Description of 2DDSLPP

The objective function of 2DDSLPP is defined as

$$\min \frac{\sum_{ij} \|Y_i - Y_j\|^2 S_{ij}}{\sum_{ij} \|Y_i - Y_j\|^2 S_{Dij}}, \quad (4)$$

where S_s is the similarity matrix of samples from the same class. S_D is the similarity matrix of the samples from different classes. Their definitions are as follows:

$$S_{Sij} = \begin{cases} \exp(-\|X_i - X_j\|^2/t) & \text{if } x_i, x_j \text{ are from the same class} \\ 0 & \text{otherwise} \end{cases}$$

and

$$S_{Dij} = \begin{cases} \exp(-\|X_i - X_j\|^2/t) & \text{if } x_i, x_j \text{ are from different classes} \\ 0 & \text{otherwise} \end{cases}.$$

The objective function (4) consists of two parts: i.e. the numerator part and the denominator part. The numerator part constrains the transform results of samples from the same class, and the denominator part constrains the transform results of samples from different classes. The goal of the objective function is to minimize $\sum_{ij} \|Y_i - Y_j\|^2 S_{ij}$ and to maximize $\sum_{ij} \|Y_i - Y_j\|^2 S_{Dij}$. Clearly the objective function requires that neighbor samples from the same class should be transformed into data points close to each other and samples from different classes should be transformed into data points far from each other.

Adopting a similar analysis method on the objective function of 2DLPP, the objective function (4) can be transformed into as follows:

$$\min \frac{\sum_{ij} \|Y_i - Y_j\|^2 S_{ij}}{\sum_{ij} \|Y_i - Y_j\|^2 S_{Dij}} = \min \frac{\mathbf{a}^T X(L_1 \otimes I_n) X^T \mathbf{a}}{\mathbf{a}^T X(L_2 \otimes I_n) X^T \mathbf{a}}, \quad (5)$$

where $L_1 = D_1 - S_s$, $D_{1ii} = \sum_j S_{Sij}$, $L_2 = D_2 - S_D$, $D_{2ii} = \sum_j S_{Dij}$. Let $\mathbf{a}^T X(L_2 \otimes I_n) X^T \mathbf{a} = c (c \neq 0)$, where c is a constant, then we can define the following Lagrange function:

$$L(\mathbf{a}, \lambda) = \mathbf{a}^T X(L_1 \otimes I_n) X^T \mathbf{a} + \lambda(c - \mathbf{a}^T X(L_2 \otimes I_n) X^T \mathbf{a}). \quad (6)$$

In other words, objective function (5) is equivalent to the objective function $\min \sum_{ij} \|Y_i - Y_j\|^2 S_{ij}$ under the condition that $\sum_{ij} \|Y_i - Y_j\|^2 S_{Dij}$ is fixed. Clearly the solution to minimize (6) will occur under the condition that the partial derivative of $L(\mathbf{a}, \lambda)$

respect to \mathbf{a} equals zero, i.e. $\partial L(\mathbf{a}, \lambda) / \partial \mathbf{a} = 0$. As a result, we know that the optimal solution of (6) should be the eigenvector corresponding to the minimum eigenvalue of the following generalized eigenvalue problem:

$$X(L_1 \otimes I_n)X^T \mathbf{a} = \lambda X(L_2 \otimes I_n)X^T \mathbf{a}. \tag{7}$$

The rationales of 2DSLPP can be described as follows: First, to minimize $\sum_{i,j} \|Y_i - Y_j\|^2 S_{D_{ij}}$ requires that samples from the same class preserve their distance relationship well. This is similar to 2DLPP except that 2DSLPP takes the samples from the same class as the neighbors of one sample while 2DLPP determines the neighbors in terms of only the distance metric and does not take the class relationship into account. Second, as we know, if two original samples, respectively, from two classes have a large distance, then $S_{D_{ij}}$ has a relatively small value. As a result, to make $\sum_{i,j} \|Y_i - Y_j\|^2 S_{D_{ij}}$ have a fixed value indeed implies that if two original samples from two different classes have a large distance, their transform results should also have a relatively large distance. On the other hand, if two original samples respectively from two classes have a small distance, then the transform results of the two samples should also have a relatively small distance. This means that besides that 2DSLPP can preserve the neighbor relationship for samples from the same class, 2DSLPP is also helpful to preserve the distance relation for samples from different classes.

3.2. 2DDSLPP-based feature extraction procedure

The 2DDSLPP-based feature extraction procedure consists of the following three steps: The first step solves the eigenvectors of the corresponding eigenvalue equation. The second step exploits the eigenvalues to determine transform axes. We formally describe the three steps as follows:

Step 1: Solve Eq. (7) to obtain the eigenvectors and eigenvalues.

Step 2: Determine transform axes. Suppose that the eigenvectors corresponding to the d eigenvalues $\lambda_1 < \lambda_2 < \dots < \lambda_d$ are $\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_d$. If r transform axes are needed, then we should select the first r eigenvectors as the transform axes.

Step 3: Let $A = [\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_d]$. Transform X_i into Y_i using $y_i = A^T X_i$.

4. Comparison between 2DDLPP, 2DLDA and 2DDSLPP

Superficially, 2DDSLPP seems to be formally similar to 2DDLPP and 2DLDA. However, 2DDSLPP is also obviously different from them. In order to investigate the similarity and the difference, we compare the proposed 2DDSLPP with 2DDLPP [10] and 2DLDA [15,16] in this section.

4.1. Comparison between 2DDSLPP and 2DLDA

Let \mathbf{x} be the transform axis of 2DLDA, then the transformation from X_i to Y_i will be performed by $Y_i = X_i \mathbf{x}$. Suppose there are C classes and each class has n_c samples. The objective function of 2DLDA is

$$J_{2DLDA}(\mathbf{x}) = \max \frac{\mathbf{x}^T S_B \mathbf{x}}{\mathbf{x}^T S_W \mathbf{x}} = \max \frac{\mathbf{x}^T \sum_{i=1}^C (\bar{X}_i - \bar{X})^T (\bar{X}_i - \bar{X}) \mathbf{x}}{\mathbf{x}^T \sum_{c=1}^C \sum_{i=1}^{n_c} (X_i^c - \bar{X}_i)^T (X_i^c - \bar{X}_i) \mathbf{x}}, \tag{8}$$

where \bar{X} is the mean matrix of all the samples, \bar{X}_i is the mean matrix of the i th class, and X_i^c denotes the i -th sample from the c th

class. Exploiting the following equation:

$$\begin{aligned} & \frac{\mathbf{x}^T \sum_{i=1}^C (\bar{X}_i - \bar{X})^T (\bar{X}_i - \bar{X}) \mathbf{x}}{\mathbf{x}^T \sum_{c=1}^C \sum_{i=1}^{n_c} (X_i^c - \bar{X}_i)^T (X_i^c - \bar{X}_i) \mathbf{x}} \\ &= \frac{\sum_{i=1}^C \|\bar{X}_i \mathbf{x} - \bar{X} \mathbf{x}\|^2}{\sum_{c=1}^C \sum_{i=1}^{n_c} \|X_i^c \mathbf{x} - \bar{X}_i \mathbf{x}\|^2} \\ &= \frac{\sum_{i=1}^C \|\bar{Y}_i - \bar{Y}\|^2}{\sum_{c=1}^C \sum_{i=1}^{n_c} \|Y_i^c - \bar{Y}_i\|^2}, \end{aligned}$$

we can transform the objective function of 2DLDA into

$$J_{2DLDA}(\mathbf{x}) = \max \frac{\sum_{i=1}^C \|\bar{Y}_i - \bar{Y}\|^2}{\sum_{c=1}^C \sum_{i=1}^{n_c} \|Y_i^c - \bar{Y}_i\|^2}. \tag{9}$$

We note that the objective function as shown in (4) of 2DDSLPP is identical to the following function:

$$J = \max \frac{\sum_{i,j} \|Y_i - Y_j\|^2 S_{D_{ij}}}{\sum_{i,j} \|Y_i - Y_j\|^2 S_{S_{ij}}}. \tag{10}$$

From (9), we know that the goal of 2DLDA is to make each class mean far as much as possible from the mean of all the samples, while making each sample close as much as possible to its class mean. From (10), we know that 2DDSLPP attempts to separate the samples from different classes while requiring the samples of the same classes to be close to each other. 2DDSLPP is different from 2DLDA as follows: first, 2DLDA evaluates the separability in terms of the distance between class means or the distance between the sample and its class mean, whereas 2DDSLPP evaluate the separability in terms of the distance between samples. Second, in 2DLDA, samples from the same class are dealt with in the same way and no similarity matrix is used to express the similarity between different samples from the same class. In other words, we can also say that 2DLDA uses the same similarity parameter “1” to represent the similarity between all the samples from the same class. However, 2DDSLPP uses a variable parameter, i.e. similarity coefficient, to represent the similarity between two samples from the same class. As a result, while 2DDSLPP requires that samples from the same class be close to each other, it also tries to preserve the neighbor relationship (locality structure) of samples from the same class. Based on the analysis shown in Section 3 and the objective function of 2DDSLPP, we can conclude as follows: While 2DDSLPP requires that samples from different classes to be far from each other, it also partially requires that the larger the distance between two original samples from different classes is, the larger the distance between their transform results should be. In contrast, while implementing the transform, 2DLDA does not take into account the locality structure of samples. Indeed, 2DDLPP can also be viewed as a combination of the ideas of both 2DLDA and 2DLPP.

4.2. Comparison between 2DDSLPP and 2DDLPP

Suppose there are C classes and each class has n_c samples, then the description of 2DDLPP is as follows: First, the objective function of 2DDLPP is

$$\min \frac{\sum_{c=1}^C \sum_{i,j=1}^{n_c} \|Y_i^c - Y_j^c\|^2 S_{ij}^c}{\sum_{i,j=1}^C \|M_i - M_j\|^2 W_{ij}}, \tag{11}$$

where both S^c and W are similarity matrices, and M_i, M_j , respectively, represent the transform results of the sample means of the i th and j th class, i.e. $M_i = 1/n_c \sum_{k=1}^{n_c} Y_k^i, M_j = 1/n_c \sum_{k=1}^{n_c} Y_k^j$. Y_i^c denotes the transform result of the i -th sample from the c th class. The entries of S^c represent the similarity of two samples from the same class. S^c are defined as follows: if X_i and X_j belong to the same class, then $S_{ij}^c = \exp(-\|X_i - X_j\|^2/t)$; otherwise, $S_{ij}^c = 0$.



Fig. 1. Some face images from the ORL database.

The entries of W represent the similarity of the means of two classes and are determined as follows: $W_{ij} = \exp(-\|F_i - F_j\|^2/t)$, where F_i is the sample mean of the i th class, i.e. $F_i = 1/n_c \sum_{k=1}^{n_c} X_k^i$.

We can compare 2DDSLPP and 2DDLPP by analyzing their objective functions as shown in (4) and (11). We note that the numerators of (4) and (11) are the same, which is actually the weighted sum of the distance between the transform results of two samples from the same class and the weight is the similarity of the two samples. The main difference between the two objective functions is from the denominator part. In (4), $\sum_{i,j} \|Y_i - Y_j\|^2 SD_{ij}$ represents the weighted sum of the distances between two samples from different classes. However, in (11), $\sum_{i,j=1}^c \|M_i - M_j\|^2 W_{ij}$ stands for the weighted sum of the distances between means of the samples of different classes (also referred to as class means). Consequently, while 2DDLPP requires that the means of different classes be far from each other, it also tries to partially preserve the distance relation of the class means. That is, it requires that the farther from each other the class means are, the farther from each other their transform results should be. In this sense, while 2DDLPP preserves the distance relation of the class means, it is possible that 2DDSLPP does not perform well in preserving the distance relation between some samples from different classes, especially in the case where the class has a complex distribution. We can conclude that 2DDSLPP can perform well in preserving the locality structure of within-class samples, whereas it might fail in preserving the distance relation between some samples from different classes.

2DDSLPP also requires that different classes to be far from each other. It achieves this by constraining the transform results of samples rather than the class means. In other words, while 2DDSLPP requires that samples from different classes be far from each other, it also tries to preserve the distance relation of these samples. That is, it requires that the farther from each other the original samples from different classes are, the farther from each other their transform results should be. As 2DDSLPP directly imposes the distance constraint on the samples from different classes rather than on the class means, it has a stronger capability to preserve the distance relation of samples from different classes than 2DDLPP.

5. Experimental results

In this section we test 2DDSLPP, 2DLPP, 2DPCA, 2DLDA and 2DDLPP using two face databases. We use the nearest neighbor classifier to classify transform results of samples obtained using different methods.

5.1. Experiment on the ORL database of faces

The ORL face database includes 400 face images from 40 subjects [17]. The images include variations in facial expression (smiling/not smiling, open/closed eyes) and facial detail. The subjects are in an upright, frontal position with tolerance for some tilting and rotation of up to 20°. Each of the face images contains

Table 1

The mean and the standard deviation of the 10 top recognition rates (%) on the ORL database.

Method	G2/8	G3/7	G4/6	G5/5
2DPCA	75.38 ± 3.1	84.54 ± 3.1	88.87 ± 3.2	90.05 ± 1.5
2DLDA	79.46 ± 4.8	88.67 ± 3.2	90.46 ± 2.5	93.17 ± 1.6
2DLPP	72.16 ± 4.2	83.13 ± 3.9	87.79 ± 3.4	91.69 ± 2.5
2DDLPP	78.56 ± 3.8	87.34 ± 3.7	90.54 ± 3.4	93.52 ± 2.4
2DDSLPP	79.56 ± 4.0	89.33 ± 4.7	91.17 ± 3.4	94.18 ± 2.2

112 × 92 pixels. Several face images of the ORL database are shown in Fig. 1.

We segmented the ORL database into a number of training sets and testing sets and tested the methods using all of them. We used Gx/y to denote the groups of the training sets consisting of x samples per subject and the corresponding testing sets consisting of the remaining y samples per subject. There are the following four groups: G2/8, G3/7, G4/6 and G5/5. When we performed experiments on “G2/8”, we arbitrarily selected two samples per subject as training samples and took the remaining eight samples of each subject as testing samples. In this subsection, for each “ Gx/y ”, we arbitrarily used 10 training sets and the corresponding testing sets to test the methods. We showed the mean and the standard deviation of the top accuracies of the experiments on the 10 training sets and the corresponding testing sets of each “ Gx/y ” in Table 1. Table 1 shows that 2DDSLPP obtains a higher recognition rate than 2DLPP. In addition, Table 1 also shows that 2DDSLPP performs better than both 2DDLPP and 2DLDA.

5.2. Experiments on the AR face database

The AR face database contains 4000 images of frontal view faces with different facial expressions, illumination conditions, and occlusions (sun glasses and scarf) [18].

We first resized the images to 40 by 50 pixels using the down-sampling method in [19]. Several AR samples after down-sampling are shown in Fig. 2. Because the training and testing using the whole database is very time consuming, we exploited only the samples of the first 40 subjects to test all the methods. We run the experiments for the following four groups of the training and testing sets: G6/20, G8/18, G10/16, G12/24, where Gx/y is also defined as in Section 5.1. For each “ Gx/y ”, we arbitrarily used 10 training sets and the corresponding testing sets to test the methods. We showed the mean and the standard deviation of the top accuracies of the experiments on the 10 training sets and the corresponding testing sets of each “ Gx/y ” in Table 2.

6. Conclusion

In this paper, we proposed and tested one improvement to 2DLPP, i.e. 2DDSLPP. While 2DDSLPP tries to preserve the local structure of the samples from the same class, it also integrates the



Fig. 2. Some face images generated from the AR database by using the down-sampling method in [19].

Table 2

The mean and the standard deviation of the 10 top recognition rates (%) on the AR database.

Method	G6/20	G8/18	G10/16	G12/24
2DPCA	45.23 ± 4.3	47.92 ± 6.3	55.80 ± 4.0	57.77 ± 4.9
2DLDA	83.48 ± 5.6	85.72 ± 8.8	92.97 ± 4.4	94.45 ± 4.5
2DLPP	46.66 ± 16.3	47.67 ± 6.0	53.70 ± 4.5	55.18 ± 4.5
2DDLPP	86.41 ± 4.0	88.01 ± 7.5	93.36 ± 3.4	94.95 ± 3.1
2DDSLPP	86.44 ± 4.6	88.67 ± 7.5	93.70 ± 3.8	95.68 ± 3.7

advantages of LPP and LDA in a good way. 2DDSLPP can obtain the subspace which well discriminates different classes and is partially subject to the essence of LPP. Our analysis shows that while 2DDLPP preserves the distance relation of the class means, it is possible that 2DDLPP does not perform well in preserving the distance relation between some samples from different classes, especially in the case where the class has a complex distribution. However, 2DDSLPP has a stronger capability to preserve the distance relation of samples from different classes than 2DDLPP. The experimental results show that 2DDSLPP can obtain a higher classification accuracy than 2DDLPP, 2DPCA and 2DLDA.

Acknowledgments

This work was partially supported by the Program for New Century Excellent Talents in University (NCET-08-0156, NCET-08-0155), the NSFC under Grant nos. 60602038, 60803090 and 60702076, the National High-Tech Research and Development Plan of China (863) under Contract no. 2007AA01Z195.

References

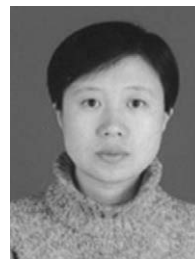
- [1] M. Turk, A.P. Pentland, Face recognition using eigenfaces, IEEE Conference on Computer Vision and Pattern Recognition (1991) 586–591.
- [2] P.N. Belhumeur, J.P. Hespanha, D.J. Kriegman, Eigenfaces vs. Fisherfaces: recognition using class specific linear projection., IEEE Transactions on Pattern Analysis and Machine Intelligence 19 (7) (1997) 711–720.
- [3] Y. Xu, J.-Y. Yang, Z. Jin, A novel method for Fisher discriminant analysis, Pattern Recognition 37 (2) (2004) 381–384.
- [4] X. He, P. Niyogi, Locality preserving projections, in: Proceedings of the Conference on Advances in Neural Information Processing Systems, 2003.
- [5] Z. Zheng, F. Yang, W. Tan, Gabor feature-based face recognition using supervised locality preserving projections, Signal Processing 87 (2007) 2473–2483.
- [6] W. Yu, X. Teng, C. Liu, Face recognition using discriminant locality preserving projections, Image and Vision Computing 24 (2006) 239–248.
- [7] X.F. He, S. Yan, Y. Hu, P. Niyogi, H.J. Zhang, Face recognition using laplacianfaces, IEEE Transactions on Pattern Analysis and Machine Intelligence 27 (3) (2005) 328–340.
- [8] S. Chen, H. Zhao, 2D-LPP: a two-dimensional extension of locality preserving projections, Neurocomputing 70 (2007) 912–921.
- [9] D. Hu, G. Feng, Z. Zhou, Two-dimensional locality preserving projections (2DLPP) with its application to palmprint recognition, Pattern Recognition 40 (2007) 339–342.
- [10] R. Zhi, Q. Ruan, Facial expression recognition base on two-dimensional discriminant locality preserving projections, Neurocomputing 70 (2007) 1543–1546.
- [11] L. Zhu, S. Zhu, Face recognition based on orthogonal discriminant locality preserving projections, Neurocomputing 70 (2007) 1543–1546.
- [12] Y. Xu, D. Zhang, A new solution scheme of unsupervised locality preserving projection method for the SSS problem, in: N. da Vitoria Lobo, et al. (Eds.), Lecture Notes in Computer Science, Vol. 5342, SSPR&SPR 2008, 2008, pp. 785–791.
- [13] J. Yang, D. Zhang, A.F. Frangi, J.-Y. Yang, Two-dimensional PCA: a new approach to appearance-based face representation and recognition, IEEE Transactions on Pattern Analysis and Machine Intelligence (2004) 131–137.
- [14] Y. Xu, D. Zhang, J. Yang, J.-Y. Yang, An approach for directly extracting features from matrix data and its application in face recognition, Neurocomputing 71 (10–12) (2008) 1857–1865.
- [15] M. Li, B. Yuan, 2D-LDA: a statistical linear discriminant analysis for image matrix, Pattern Recognition Letter 26 (2005) 527–532.
- [16] J. Yang, D. Zhang, Y. Xu, J.-Y. Yang, Two-dimensional discriminant transform for face recognition, Pattern Recognition 38 (7) (2005) 1125–1129.
- [17] <<http://www.cl.cam.ac.uk/research/dtg/attarchive/facedatabase.html>>.
- [18] <http://cobweb.ecn.purdue.edu/~aleix/aleix_face_DB.html>.
- [19] Y. Xu, Z. Jin, Down-sampling face images and low-resolution face recognition, in: Innovative Computing Information and Control on Third International Conference (ICIC '08), 2008, pp. 392.



Yong Xu received his B.S. and M.S. degrees at Air Force Institute of Meteorology (China) in 1994 and 1997, respectively. He then received his Ph.D. degree in pattern recognition and intelligence system at the Nanjing University of Science and Technology (NUST) in 2005. From May 2005 to April 2007, he worked at Shenzhen graduate school, Harbin Institute of Technology (HIT) as a postdoctoral research fellow. Now he is an associate professor at Shenzhen graduate school, HIT. He also acts as a research assistant researcher at the Hong Kong Polytechnic University from August 2007 to June 2008. His current interests include pattern recognition, biometrics, and machine learning. He has published more than 40 scientific papers.



Ge Feng obtained her B.S. degree at Zhengzhou University in 2006. She worked for her M.S. degree in Computer Science & Technology at Shenzhen graduate school, Harbin Institute of Technology during 2006. 9–2008.12.



Yingnan Zhao graduated in computer science from Shaanxi Normal University in 1996 and received her M.S. degree in computer science from Shenyang Ligong University in 2002 and Ph.D. degree in pattern recognition from Nanjing University of Science and Technology in 2005, respectively. Currently, she is a vice professor in Nanjing University of Information Science & Technology. Her main research interests are feature extraction, face recognition and image processing.