Two-dimensional locality preserving projection based on Maximum Scatter Difference ⋆

Su-Jing Wang [a] Na Zhang [a] Xu-Jun Peng [b] Chun-Guang Zhou [a,*]

[a] *College of Computer Science and Technology, Jilin University, Changchun 130012, China*

[b] *CSE department, SUNY at Buffalo, Amherst, NY 14228, USA*

Abstract

In this paper, we propose a Two-dimensional locality preserving projection based on maximum scatter difference (2D-DLPP/MSD). 2D-LPP/MSD use additive principle to preserve the locality by maximizing the between-class scatter and within-class scatter instead of using multiplicative principle of 2D-DLPP. Theoretically, we also discuss the influence of balance factor α on performance and reveal the relations between 2D-LPP/MSD and 2D-DLPP. Experimental results on the ORL and Yale face databases show the effectiveness of the proposed 2D-DLPP/MSD.

Keywords: Face recognition; Locality preserving projection; Maximum scatter difference; Two dimensional

1 Introduction

As one of the most important biometric techniques, face recognition has gained lots of attentions in pattern recognition and machine learning areas. Usually, a 2D facial image is represented as a feature point in the high dimensional feature space. Its perceptually structure can be characterized by using a small set of meaningful parameters. Thus, dimensionality reduction techniques are commonly used before recognition.

PCA[1] is a widely used linear dimensionality reduction method by maximizing variance of projected feature in the projective subspace. Linear LDA[2] encodes discriminant information by maximizing the ratio between the between-class and within-class scatters. Seung[3] assumed that the high dimensional visual image information in real world lies on or is close to a smooth low dimensional manifold. Inspired by this idea, multiple manifold dimensionality reduction methods that preserve local structure of samples have been proposed. Locality Preserving Projections

⋆Project supported by (1)the National Natural Science Foundation of China under Grant No. 60873146, 60973092, 60903097, (2)project of science and technology innovation platform of computing and software science (985 engineering), (3)the Key Laboratory for Symbol Computation and Knowledge Engineering of the National Education Ministry of China.

*Corresponding author.

Email address: cgzhou@jlu.edu.cn (Chun-Guang Zhou).

(LPP)[4] aims to preserve the local structure of the original space in the projective subspace. Its performance is better than those of PCA and LDA for face recognition[5]. Discriminant Locality Preserving Projections (DLPP)[6] encodes discriminant information into LPP to further improve the discriminant performance of LPP for face recognition. Some other DLPP related works can be found in [7][8][9][10].

The above one dimensional methods applied on the 2D images have some potential problems: singularity of within-class scatter matrices, limited available projection directions, high computational cost and a loss of the underlying spatial structure information of the images.

To overcome the above problems, some researchers have attempted to treat the image as a matrix instead of a vector. Yang *et al.*[11] proposed a 2D-PCA algorithm to compute the image scatter matrix from the image matrix representations directly. Li and Yuan[12] presented a 2D-LDA which is an extension of the tranditional LDA by using the idea of the image matrix representations. Chen *et al.*[13] developed a 2D-LPP which directly extracts the proper features from image matrix representations by preserving the local structure of samples. Xu *et al.*[14] used discriminant information to construct the adjacency graph based on 2D-LPP. And Yu developed [15] a 2D-DLPP, a variation of 2D-LPP which uses DI. 2D-LPP and 2D-DLPP achieved better results in recognizing face, facial expression[16], gait[17], and palm[18] than the methods which preserve the global structure of samples such as 2D-PCA, 2D-LDA. These methods not only reduce the complexities of time and space, but also preserve spatial structure information of the 2D images.

In this paper, we propose a new two-dimensional discriminant locality preserving projection based on the maximum scatter difference (2D-LPP/MSD). Motivated by the idea of MSD[19], 2D-LPP/MSD seeks to maximize the difference, rather than the ratio, between the *locality preserving between-class scatter* and the *locality preserving within-class scatter*. The experimental results on the ORL and Yale face databases show the effectiveness of the proposed 2D-LPP/MSD method.

The rest of this paper is organized as follows: in Section 2, we briefly review the LDA, MSD and LPP; in Section 3, we introduce the two-dimensional discriminant locality preserving projection based on the maximum scatter difference; in Section 4, we analyze the influence of the parameter α on recognition performance and reveal the relations between 2D-LPP/MSD and 2D-DLPP, theoretically; in Section 5, the experimental results are reported and analyzed; finally in Section 6, conclusions are drawn and several issues for future works are discussed.

2 Review of the LDA, MSD and LPP algorithms

LDA is one of the linear discriminant dimensionality reduction algorithms. It tries to search for the directions which are most effective for discrimination by maximizing the ratio between the *between-class scatter matrix* $\mathbf{S}_b$ and the *within-class scatter matrix* $\mathbf{S}_w$. $\mathbf{S}_w$ and $\mathbf{S}_b$ are defined as

$$\mathbf{S}_w = \sum_{c=1}^{C} \sum_{\mathbf{x}_i \in \mathbf{X}_c} (\mathbf{x}_i - \overline{\mathbf{x}}_c)(\mathbf{x}_i - \overline{\mathbf{x}}_c)^T \tag{1}$$

$$\mathbf{S}_b = \sum_{c=1}^{C} N_c(\overline{\mathbf{x}}_c - \overline{\mathbf{x}})(\overline{\mathbf{x}}_c - \overline{\mathbf{x}})^T \tag{2}$$

where $\overline{\mathbf{x}}_c$ is the mean value of the cth class and N_c is the number of samples of cth class. The criterion function is defined as:

$$\max_{\mathbf{w}} \frac{\mathbf{w}^T \mathbf{S}_b \mathbf{w}}{\mathbf{w}^T \mathbf{S}_w \mathbf{w}} \tag{3}$$

MSD[19] is another linear discriminant dimensionality reduction algorithm. Its criterion function is obtained by using the additive principle[20] instead of the multiplicative principle to combine the two scatter matrices.

$$\max_{\mathbf{w}} \mathbf{w}^T (\mathbf{S}_b - \alpha \mathbf{S}_w) \mathbf{w} \tag{4}$$

where, the parameter α is a nonnegative constant which is a balance factor.

LPP[4] is a manifold dimensionality reduction which attempts to find a transformation matrix $\mathbf{A}$ that maps N samples $\mathbf{X} = [\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_N]$ to a set of the projected points $\mathbf{Y} = [\mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_N]$, where $\mathbf{X}$ is $D \times N$ matrix and $\mathbf{Y}$ is $d \times N$ matrix $(d \leq D)$. The objective function of LPP is as follows:

$$\min_{\mathbf{A}} \sum_{i,j} (\mathbf{y}_i - \mathbf{y}_j)^2 S_{ij} \tag{5}$$

where $y_i = \mathbf{A}^T x_i$ and the matrix $\mathbf{S}$ is defined as follows

$$S_{ij} = \begin{cases} exp(-\|x_i - x_j\|^2/t) & x_i(x_j) \text{ is among } k \text{ nearest neighbor of } x_j(x_i) \\ 0 & \text{otherwise.} \end{cases} \tag{6}$$

where t is a parameter that can be determined empirically. The Eq. (5) can be converted to the generalized eigenvalue problem as follows:

$$\mathbf{XLX}^T\mathbf{a} = \lambda \mathbf{XDX}^T\mathbf{a} \tag{7}$$

where $\mathbf{D}$ is a diagonal matrix; its entries are column (or row, since $\mathbf{S}$ is symmetric) sum of $\mathbf{S}$, $D_{ii} = \sum_j S_{ij}$. $\mathbf{L} = \mathbf{D} - \mathbf{S}$ is the Laplacian matrix.

Let the column vectors $\mathbf{a}_1, \mathbf{a}_2, \ldots, \mathbf{a}_d$ be the solutions of Eq. (7), ordered according to their eigenvalues, $\lambda_1, \lambda_2, \ldots, \lambda_d$. Thus, the transformation matrix $\mathbf{A}$ can be obtained as follows:

$$\mathbf{A} = [\mathbf{a}_1, \mathbf{a}_2, \ldots, \mathbf{a}_d] \tag{8}$$

3 2D-LPP based on Maximum Scatter Difference

We have a set $\mathcal{X}$ consisting of N samples coming from C classes:

$$\mathcal{X} = \{\mathbf{X}_1^1, \mathbf{X}_2^1, \ldots, \mathbf{X}_{N_1}^1, \mathbf{X}_1^2, \mathbf{X}_2^2, \ldots, \mathbf{X}_{N_2}^2, \ldots, \mathbf{X}_1^C, \mathbf{X}_2^C, \ldots, \mathbf{X}_{N_C}^C\} \tag{9}$$

where $\mathbf{X}_i^c \in \mathbb{R}^{m \times n}$ means the ith sample in the cth class. N_c is the number of samples in the cth class, and $N_1 + N_2 + \ldots + N_C = N$ is satisfied. The task is to obtain a projective matrix $\mathbf{V}$ which projects those N samples to a set of the projected points

$$\mathbf{Y}_i^c = \mathbf{X}_i^c \mathbf{V}, \quad i = 1, 2, \ldots, N_c, \quad c = 1, 2, \ldots, C. \tag{10}$$

where $\mathbf{Y}_i^c \in \mathbb{R}^{m \times d}$.

$\mathbf{W}^c$ is the *within-class similarity matrix* of cth class and its entry W^c_{ij} is the similarity between the samples $\mathbf{X}^c_i$ and $\mathbf{X}^c_j$, which is defined as: $W^c_{ij} = \exp(-\|\mathbf{X}^c_i - \mathbf{X}^c_j\|^2_F/t)$, where $\|\cdot\|$ is the Frobenius norm of matrix, *i.e.* $\|\mathbf{A}\|_F = \sqrt{\sum_i \sum_j A^2_{ij}}$.

We seek $\mathbf{V}$ such that the two samples of the same class $\mathbf{X}^c_i$ and $\mathbf{X}^c_j$ are close then feature points $\mathbf{Y}^c_i$ and $\mathbf{Y}^c_j$ are close as well. Thus we minimize the following objective:

$$
\begin{aligned}
&\frac{1}{2}\sum_{c=1}^{C}\sum_{i,j=1}^{N_c} \|\mathbf{X}^c_i\mathbf{V} - \mathbf{X}^c_j\mathbf{V}\|^2_F W^c_{ij} \\
&= \frac{1}{2}\sum_{c=1}^{C}\sum_{i,j=1}^{N_c} tr\left[(\mathbf{X}^c_i\mathbf{V} - \mathbf{X}^c_j\mathbf{V})^T(\mathbf{X}^c_i\mathbf{V} - \mathbf{X}^c_j\mathbf{V})\right] W^c_{ij} \\
&= \frac{1}{2}\sum_{c=1}^{C}\sum_{i,j=1}^{N_c} tr\left[\mathbf{V}^T(\mathbf{X}^c_i - \mathbf{X}^c_j)^T(\mathbf{X}^c_i - \mathbf{X}^c_j)\mathbf{V}\right] W^c_{ij} \\
&= tr\left\{\sum_{c=1}^{C}\sum_{i,j=1}^{N_c} \mathbf{V}^T\left[(\mathbf{X}^c_i)^T\mathbf{X}^c_i - (\mathbf{X}^c_i)^T\mathbf{X}^c_j\right]\mathbf{V}W^c_{ij}\right\} \\
&= tr\left\{\mathbf{V}^T\left\{\sum_{c=1}^{C}\left[\sum_{i=1}^{N_c}(\mathbf{X}^c_i)^T\mathbf{X}^c_i\sum_{j=1}^{N_c}W^c_{ij} - \sum_{i,j=1}^{N_c}(\mathbf{X}^c_i)^T\mathbf{X}^c_jW^c_{ij}\right]\right\}\mathbf{V}\right\} \\
&= tr\left\{\mathbf{V}^T\left\{\sum_{c=1}^{C}\left[(\mathbf{X}^c)^T(\mathbf{D}^c\otimes\mathbf{I}_m)\mathbf{X}^c - (\mathbf{X}^c)^T(\mathbf{W}^c\otimes\mathbf{I}_m)\mathbf{X}^c\right]\right\}\mathbf{V}\right\} \\
&= tr\left\{\mathbf{V}^T\mathbf{X}^T\left[(\mathbf{D}-\mathbf{W})\otimes\mathbf{I}_m\right]\mathbf{X}\mathbf{V}\right\} \\
&= tr\left[\mathbf{V}^T\mathbf{X}^T\left(\mathbf{L}\otimes\mathbf{I}_m\right)\mathbf{X}\mathbf{V}\right]
\end{aligned} \tag{11}
$$

where

$$
\mathbf{X}^c = \begin{bmatrix} \mathbf{X}^c_1 \\ \mathbf{X}^c_2 \\ \vdots \\ \mathbf{X}^c_{N_c} \end{bmatrix} \tag{12}
$$

and

$$
\mathbf{X} = \begin{bmatrix} \mathbf{X}^1 \\ \mathbf{X}^2 \\ \vdots \\ \mathbf{X}^C \end{bmatrix} \tag{13}
$$

$\mathbf{D}^c$ is a diagonal matrix, and its entries are column (or row, since $\mathbf{W}^c$ is symmetric) sum of $\mathbf{W}^c$, $D^c_{ii} = \sum_j W^c_{ij}$. $\mathbf{D}$ and $\mathbf{W}$ are separately composed of $\mathbf{D}^c$ and $\mathbf{W}^c$, that is,

$$
\mathbf{D} = \begin{bmatrix} \mathbf{D}^1 & & & & \\ & \ddots & & & \\ & & \mathbf{D}^c & & \\ & & & \ddots & \\ & & & & \mathbf{D}^C \end{bmatrix} \tag{14}
$$

and

$$\mathbf{W} = \begin{bmatrix} \mathbf{W}^1 & & & & \\ & \ddots & & & \\ & & \mathbf{W}^c & & \\ & & & \ddots & \\ & & & & \mathbf{W}^C \end{bmatrix} \tag{15}$$

$\mathbf{L} = \mathbf{D} - \mathbf{W}$. $\mathbf{I}_m$ is an identity matrix of order m and operator $\otimes$ is the Kronecher product of the matrices.

Meanwhile, we attempt to ensure that if the mean samples of two class $\overline{\mathbf{X}}_i$ and $\overline{\mathbf{X}}_j$ are close then feature matrix $\mathbf{Y}_i$ and $\mathbf{Y}_j$ are far. Thus we maximize the following objective:

$$\begin{aligned}
&\frac{1}{2}\sum_{i,j=1}^{C} \|\overline{\mathbf{X}}_i\mathbf{V} - \overline{\mathbf{X}}_j\mathbf{V}\|_F^2 B_{ij} \\
&= \frac{1}{2}\sum_{i,j=1}^{C} tr\left[\mathbf{V}^T(\overline{\mathbf{X}}_i - \overline{\mathbf{X}}_j)^T(\overline{\mathbf{X}}_i - \overline{\mathbf{X}}_j)\mathbf{V}\right] B_{ij} \\
&= \sum_{i,j=1}^{C} tr\left\{\mathbf{V}^T[(\overline{\mathbf{X}}_i)^T\overline{\mathbf{X}}_i - (\overline{\mathbf{X}}_i)^T\overline{\mathbf{X}}_j]\mathbf{V}\right\} B_{ij} \\
&= tr\left\{\sum_{i=1}^{C}\mathbf{V}^T\left[(\overline{\mathbf{X}}_i)^T\overline{\mathbf{X}}_i\sum_{j=1}^{C}B_{ij}\right]\mathbf{V} - \sum_{i,j=1}^{C}\mathbf{V}^T(\overline{\mathbf{X}}_i)^T B_{ij}\overline{\mathbf{X}}_j\mathbf{V}\right\} \\
&= tr\left\{\mathbf{V}^T\overline{\mathbf{X}}^T[(\mathbf{E} - \mathbf{B})\otimes\mathbf{I}_m]\overline{\mathbf{X}}\mathbf{V}\right\} \\
&= tr\left[\mathbf{V}^T\overline{\mathbf{X}}^T(\mathbf{H}\otimes\mathbf{I}_m)\overline{\mathbf{X}}\mathbf{V}\right]
\end{aligned} \tag{16}$$

where

$$\overline{\mathbf{X}} = \begin{bmatrix} \overline{\mathbf{X}}_1 \\ \overline{\mathbf{X}}_2 \\ \vdots \\ \overline{\mathbf{X}}_C \end{bmatrix} \tag{17}$$

$\mathbf{B}$ is the *between-class similarity matrix* and its entry B_{ij} is the similarity between the mean samples $\overline{\mathbf{X}}_i$ and $\overline{\mathbf{X}}_j$, and it is defined as: $B_{ij} = \exp(-\|\overline{\mathbf{X}}_i - \overline{\mathbf{X}}_j\|_F^2/t)$, where $\overline{\mathbf{X}}_i = \frac{1}{N_i}\sum_{k=1}^{N_i}\mathbf{X}_k^i$.

We suppose $\mathbf{S}_H = \overline{\mathbf{X}}^T(\mathbf{H}\otimes\mathbf{I}_m)\overline{\mathbf{X}}$ and $\mathbf{S}_L = \mathbf{X}^T(\mathbf{L}\otimes\mathbf{I}_m)\mathbf{X}$. In 2D-DLPP/MSD, we use the criterion function as follows:

$$\max_{\mathbf{V}} \mathbf{V}^T(\mathbf{S}_H - \alpha\mathbf{S}_L)\mathbf{V} \tag{18}$$

where the parameter α is a nonnegative constant. The maximin problem (18) can be thought of as the Rayleigh quotient[21] and obtain it's solution by computing the eigenvectors and eigenvalues:

$$(\mathbf{S}_H - \alpha\mathbf{S}_L)\mathbf{v} = \lambda(\mathbf{S}_H - \alpha\mathbf{S}_L)\mathbf{v} \tag{19}$$

Let the column vectors $v_1, v_2, \ldots, v_d$ be the solutions of Eq.(19), ordered according to their eigenvalues, $\lambda_1 > \lambda_2 > \ldots > \lambda_d$, Thus The optimal projection matrix $\mathbf{v} = [\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_d]$.

4 Relations between 2D-LPP/MSD and 2D-DLPP

It is easy to see that the difference between 2D-LPP/MSD and 2D-DLPP is the different principles [20]. When α in Eq.(18) is with a given value, 2D-LPP/MSD will degenerate into 2D-DLPP. In other words, 2D-LPP/MSD is the generalization of 2D-DLPP .

Without loss of generality, we assume $d = 1$, then the projective matrix $\mathbf{V}$ becomes a column N-dimensional vector $\mathbf{v}$. Let $\mathbf{v}(\alpha)$ be the optimal solution of Eq.18 when the parameter α has the value α^*

$$J(\alpha^*) = \mathbf{a}(\alpha^*)^T(\mathbf{S}_H - \alpha^*\mathbf{S}_L)\mathbf{a}(\alpha^*) \tag{20}$$

Theroem 1 *$J(\alpha^*)$ is a monotone decreasing function. When the locality preserving within-class scatter matrix $\mathbf{S}_L$ is nonsingular, $J(\alpha^*)$ is a strictly monotone decreasing function. And when α^* is approaching infinity, the limit of $J(\alpha^*)$ is negative infinity .*

Proof Let $\alpha_1^* < \alpha_2^*$, $\mathbf{v}_i$ is the unit eigenvector of the matrix $(\mathbf{S}_H - \alpha_i^*\mathbf{S}_L)$ corresponding to the largest eigenvalue, $i = 1, 2$. It is obvious that

$$\begin{aligned} & J(\alpha_1^*) = \mathbf{v}_1^{\mathrm{T}}(\mathbf{S}_H - \alpha_1^* \cdot \mathbf{S}_L)\mathbf{v}_1 \geq \mathbf{v}_2^{\mathrm{T}}(\mathbf{S}_H - \alpha_1^* \cdot \mathbf{S}_L)\mathbf{v}_2 \\ & = \mathbf{v}_2^{\mathrm{T}}(\mathbf{S}_H - \alpha_2^* \cdot \mathbf{S}_L)\mathbf{v}_2 + (\alpha_2^* - \alpha_1^*)\mathbf{v}_2^{\mathrm{T}}\mathbf{S}_L\mathbf{v}_2 \\ & = J(\alpha_2^*) + (\alpha_2^* - \alpha_1^*)\mathbf{v}_2^{\mathrm{T}}\mathbf{S}_L\mathbf{v}_2 \end{aligned} \tag{21}$$

$\mathbf{S}_L$ is positive definite, we have $\mathbf{v}_2^{\mathrm{T}}S_L\mathbf{v}_2 \geq 0$. Thus, $J(\alpha_1^*) \geq J(\alpha_2^*)$, i.e., $J(\alpha^*)$ is a monotone decreasing function. Especially, when $\mathbf{S}_L$ is non-singular, $\mathbf{S}_L$ is positive definite. Thus, for any unit vector $\mathbf{v}$, we always have

$$\mathbf{v}^{\mathrm{T}}S_L\mathbf{v} \geq \lambda_L > 0 \tag{22}$$

where λ_L is the smallest eigenvalue of $\mathbf{S}_L$. In this case, $J(\alpha_1^*) > J(\alpha_2^*))$, i.e., $J(\alpha^*)$ is a strictly monotone decreasing function.

Let λ_H denote the largest eigenvalue of the matrix $\mathbf{S}_H$. for any unit vector $\mathbf{v}$, the following inequality always holds:

$$\mathbf{v}^{\mathrm{T}}S_H\mathbf{v} \leq \lambda_H \tag{23}$$

From (22) and (23), we have

$$J(\alpha^*) = \mathbf{v}^{\mathrm{T}}(S_H - \alpha^* \cdot \mathbf{S}_L)\mathbf{v} = \mathbf{v}^{\mathrm{T}}\mathbf{S}_H\mathbf{v} - \alpha^*\mathbf{v}^{\mathrm{T}}\mathbf{S}_L\mathbf{v} \leq \lambda_H - \alpha^* \cdot \lambda_L \tag{24}$$

It is obvious that $\lim\limits_{\alpha^* \to \infty} J(\alpha^*) = -\infty$

From Theorem 1, we know that the parameter α is as small as possible in order to get better recognition rate.

Theroem 2 *If $\mathbf{S}_L$ is nonsingular, there exists a unique positive root α_0 of the equation $J(\alpha_0) = 0$. The unit eigenvector of the matrix $(\mathbf{S}_H - \alpha_0\mathbf{S}_L)$ corresponding to the largest eigenvalue is the solution of 2D-DLPP.*

Fig. 1: Sample images for one subject of the YALE database.

Proof Suppose $\mathbf{v}$ is the eigenvector of the largest eigenvalue λ_H. Then

$$J(0) = \mathbf{v}^{\mathrm{T}}\mathbf{S}_H\mathbf{v} = \lambda_H > 0 \tag{25}$$

From the proof of Theorem 1, we know that $J(\alpha^*) < 0$ when $\alpha^* > \frac{\lambda_H}{\lambda_L}$.

Since $J(\alpha)$ is a continuous function, there must exist a point α_0 in the interval $(0, \alpha^*)$ such that $J(\alpha_0) = 0$. Considering that $J(\alpha)$ is a strictly monotone function, we know that the point α_0 is unique.

From $J(\alpha_0) = 0$, i.e., $(\mathbf{S}_H - \alpha_0 \cdot \mathbf{S}_L)\mathbf{v}(\alpha_0) = 0$, one can obtain:

$$\mathbf{S}_H\mathbf{v}(\alpha_0) = \alpha_0 \cdot \mathbf{S}_L\mathbf{v}(\alpha_0) \tag{26}$$

$\mathbf{a}(\alpha_0)$ is the solution of 2D-DLPP.

From Theorem 1 and Theorem 2, when $0 < \alpha < \alpha_0$, the performance of 2D-LPP/MSD is superior to 2D-DLPP, theoretically.

5 Experiments and results

In this section, the experiments are conducted on the two well-known face databases, i.e., ORL and Yale[22], to evaluate the performance of 2D-LPP/MSD. 2D-LPP, 2D-DLPP and the proposed method are used for feature extraction. A nearest neighbor classifier with Euclidean distance is employed for classification in the projected space. For heat kernel $exp(-\frac{\|x-y\|^2}{t})$, parameter t is set as 1000. For k nearest neighbors in 2D-LPP, parameter k is set as 12.

5.1 Database

There are total of 165 gray scale images for 15 individuals where each individual has 11 images in Yale face database. The images demonstrate variations in lighting condition, facial expression (normal, happy, sad, sleepy, surprised, and wink). The sample images of one individual from the Yale database are showed in Fig. 1.

The ORL database collects images from 40 individuals, and 10 different images are captured for each individual. For each individual, the images with different facial expressions and details are obtained at different times. The face in the images may be rotated, scaled and be tilting in some degree. The sample images of one individual from the ORL database are shown in Fig. 2. For the purpose of computation efficiency, all images in the two face databases are resized to 32×32 pixels.

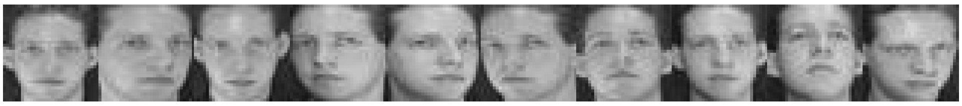

Fig. 2: Sample images for one subject of the ORL database.

5.2 Analysis on parameter α

The first 2, 5 and 8 images of each individual are selected for training, while the remaining images are used for testing on ORL and Yale face databases. The recognition accuracy of 2D-LPP/MSD algorithm over the variance of the dimensionality of subspaces and different values of the parameter α is demonstrated on ORL face database in Fig. 3. The recognition accuracy of 2D-LPP/MSD algorithm over the variance of the dimensionality of subspaces and different values of the parameter α is demonstrated on YALE face database in Fig. 4.

From the results, we can see that when the value of α is smaller, the performance of the proposed algorithm is better. The conclusion proves the theoretically analysis in Section 4.

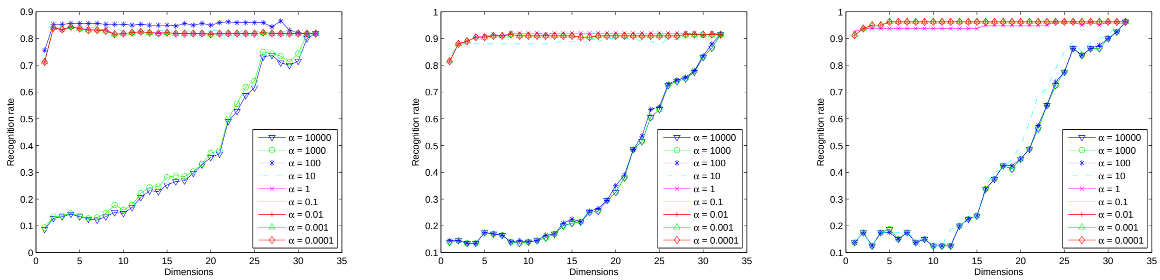

Fig. 3: Recognition accuracy V.S. α using the first 2, 5 and 8 sample images for one subject of the ORL database as training set and the remaining images as testing set.

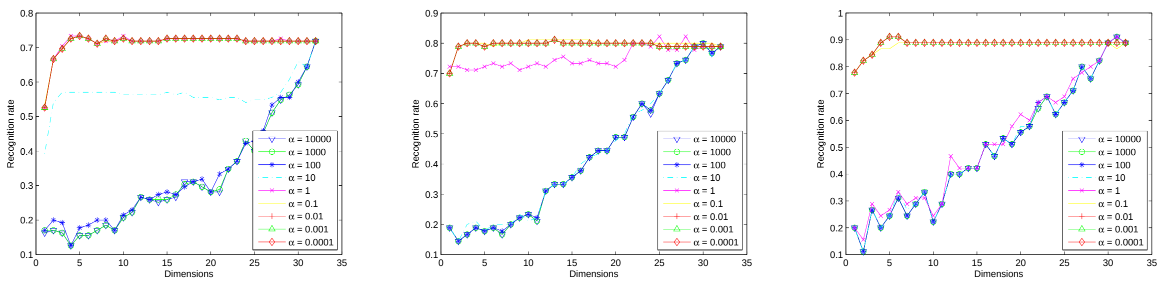

Fig. 4: Recognition accuracy V.S. α using the first 2, 5 and 8 sample images for one subject of the YALE database as training set and the remaining images as testing set.

5.3 Comparing 2D-LPP/MSD with 2D-LPP and 2D-DLPP

A number of experiments are implemented to compare the performances of 2D-LPP/MSD, 2D-LPP and 2D-DLPP with different number of training samples. The parameter α is set as 0.0001. Here, four tests are performed with different number of training samples on the ORL face database.

More specifically, in the kth test, we use the first k image samples per class for training and the remaining samples for testing on the ORL face database. Table 1 presents the top recognition accuracies of 2D-LPP/MSD, 2D-DLPP and 2D-LPP, which corresponds to different numbers of training samples. The value in parentheses denotes the dimension of feature vectors obtained by the three algorithms. Table 1 shows that the performance of 2D-LPP/MSD gets the best result using the first 7, 8 images per subject for training sets.

The experiment with the same setup is conducted on the Yale face database. Table 2 shows the proposed algorithm outperforms other methods for 5 tests out of 9 tests, and the mean accuracy of proposed method is also better than other methods.

Table 1: Comparison of the top recognition accuracy (%) of 2D-LPP/MSD versus 2D-DLPP and 2D-LPP on the ORL face database

Training samples/classs	5	6	7	8
2D-LPP/MSD	91.5(8)	96.88(5)	96.67(5)	96.25(5)
2D-DLPP	92.5(3)	96.88(4)	95.83(3)	95(3)
2D-LPP	85.5(3)	94.37(3)	95.83(3)	93.75(3)

Table 2: Comparison of the top recognition accuracy (%) of 2D-LPP/MSD versus 2D-DLPP and 2D-LPP on the Yale face database

Training samples/classs	2	3	4	5	6
2D-LPP/MSD	73.33(5)	80(5)	82.86(4)	81.11(13)	78.67(4)
2D-DLPP	54.07(3)	70.83(7)	73.33(6)	83.33(4)	78.67(3)
2D-LPP	68.89(9)	71.67(3)	77.14(3)	78.89(1)	76(2)
Training samples/classs	**7**	**8**	**9**	**10**	**Mean**
2D-LPP/MSD	93.33(5)	91.11(5)	90(4)	100(3)	85.6
2D-DLPP	91.67(6)	93.33(15)	90(9)	93.33(5)	80.95
2D-LPP	88.33(1)	88.89(7)	86.67(7)	93.33(1)	81.09

6 Conclusion

In this paper, we proposed a new two-dimensional locality preserving projection based on maximum scatter difference(2D-LPP/MSD). Compared with 2D-LPP, 2D-LPP/MSD can balance the relative merits of maximization the difference, rather than the ratio, between the *locality preserving between-class scatter* and the minimization of the *locality preserving within-class scatter* by introducing a parameter α. And we analyze the influence of the parameter α on recognition performance, theoretically. Meanwhile, the relations between 2D-LPP/MSD and 2D-DLPP are revealed. Experimental results on ORL and Yale face databases indicate that 2D-LPP/MSD performs significantly better than 2D-DLPP and 2D-LPP methods in terms of recognition accuracy.

References

[1] M. Turk, A. Pentland, Eigenfaces for recognition, Journal of Cognitive Neuroscience 3 (1) (1991) 71–86.

[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. doi:10.1109/34.598228.

[3] H. S. Seung, D. D. Lee, The manifold ways of perception, Science 290 (5500) (2000) 2268–2269.

[4] X. F. He, P. Niyogi, Locality preserving projections, in: Advances in neural information processing systems, Vol. 16, The MIT Press, 2004, pp. 153–160.

[5] X. F. He, S. C. Yan, Y. X. Hu, P. Niyogi, H. J. Zhang, Face recognition using laplacianfaces, IEEE Transactions on Pattern Analysis and Machine Intelligence 27 (3) (2005) 328–340.

[6] W. W. Yu, X. L. Teng, C. Q. Liu, Face recognition using discriminant locality preserving projections, Image and Vision Computing 24 (3) (2006) 239–248.

[7] L. Yang, W. Gong, X. Gu, W. Li, Y. Liang, Null space discriminant locality preserving projections for face recognition, Neurocomputing 71 (16-18) (2008) 3644–3649.

[8] L. Zhu, S. Zhu, Face recognition based on orthogonal discriminant locality preserving projections, Neurocomputing 70 (7-9) (2007) 1543–1546.

[9] J. Gui, C. Wang, L. Zhu, Locality preserving discriminant projections, in: ICIC'09: Proceedings of the Intelligent computing 5th international conference on Emerging intelligent computing technology and applications, Springer-Verlag, Berlin, Heidelberg, 2009, pp. 566–572.

[10] X. Zhao, X. Tian, Locality preserving fisher discriminant analysis for face recognition, in: ICIC'09: Proceedings of the 5th international conference on Emerging intelligent computing technology and applications, Springer-Verlag, Berlin, Heidelberg, 2009, pp. 261–269.

[11] J. Yang, D. Zhang, A. F. Frangi, J. Y. Yang, Two-dimensional PCA: a new approach to appearance-based face representation and recognition, IEEE Transactions on Pattern Analysis and Machine Intelligence 26 (1) (2004) 131–137.

[12] M. Li, B. Z. Yuan, 2D-LDA: A statistical linear discriminant analysis for image matrix, Pattern Recognition Letters 26 (5) (2005) 527–532.

[13] S. B. Chen, H. F. Zhao, M. Kong, B. Luo, 2D-LPP: A two-dimensional extension of locality preserving projections, Neurocomputing 70 (4-6) (2007) 912–921.

[14] Y. Xu, G. Feng, Y. N. Zhao, One improvement to two-dimensional locality preserving projection method for use with face recognition, Neurocomputing 73 (1-3) (2009) 245–249.

[15] W. Yu, Two-dimensional discriminant locality preserving projections for face recognition, Pattern Recognition Letters 30 (15) (2009) 1378–1383.

[16] R. C. Zhi, Q. Q. Ruan, Facial expression recognition based on two-dimensional discriminant locality preserving projections, Neurocomputing 71 (7-9) (2008) 1730–1734.

[17] E. Zhang, Y. W. Zhao, W. Xiong, Active energy image plus 2DLPP for gait recognition, Signal Processing 90 (7) (2010) 2295–2302.

[18] M. H. Wan, Z. H. Lai, J. Shao, Z. Jin, Two-dimensional local graph embedding discriminant analysis (2DLGEDA) with its application to face and palm biometrics, Neurocomputing 73 (1-3) (2009) 197–203.

[19] F. Song, D. Zhang, Q. Chen, J. Wang, Face recognition based on a novel linear discriminant criterion, Pattern Analysis & Applications 10 (3) (2007) 165–174.

[20] H. Eschenauer, J. Koski, A. Osyczka, Multicriteria design optimization, Springer-Verlag, 1990.

[21] R. A. Horn, C. A. Johnson, Matrix Analysis, Cambridge University Press, 1985.

[22] Yale face database `http://cvc.yale.edu/projects/yalefaces/yalefaces.html`.