Neurocomputing 73 (2010) 2696–2707

Contents lists available at ScienceDirect

Neurocomputing

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

Locality preserving discriminant projections for face and palmprint recognition

Jie Gui [a,b], Wei Jia [a,*], Ling Zhu [a,b], Shu-Ling Wang [a,c], De-Shuang Huang [a]

[a] Hefei Institute of Intelligent Machines, Chinese Academy of Sciences, PO Box 1130, Hefei, Anhui 230031, China
[b] Department of Automation, University of Science and Technology of China, Hefei, Anhui 230027, China
[c] School of Computer and Communication, Hunan University, Chuangsha, Hunan 410082, China

ARTICLE INFO

Article history:
Received 31 January 2010
Received in revised form
27 April 2010
Accepted 28 April 2010
Communicated by J. Yang
Available online 11 May 2010

Keywords:
Subspace learning
Locality preserving discriminant projections
Locality preserving projections
Biometrics
Face recognition
Palmprint recognition

ABSTRACT

A new subspace learning algorithm called locality preserving discriminant projections (LPDP) is proposed by adding the criterion of maximum margin criterion (MMC) into the objective function of locality preserving projections (LPP). LPDP retains the locality preserving characteristic of LPP and utilizes the global discriminative structures obtained from MMC, which can maximize the between-class distance and minimize the within-class distance. Thus, our proposed LPDP combining manifold criterion and Fisher criterion has more discriminanting power, and is more suitable for recognition tasks than LPP, which considers only the local information for classification tasks. Moreover, two kinds of tensorized (multilinear) forms of LPDP are also derived in this paper. One is iterative while the other is non-iterative. The proposed LPDP method is applied to face and palmprint biometrics and is examined using the Yale and ORL face image databases, as well as the PolyU palmprint database. Experimental results demonstrate the effectiveness of the proposed LPDP method.

1. Introduction

In computer vision and pattern recognition fields, the dimension of signals is usually very high. For example, a 64×64 pixel face image can be viewed as a point in a 4096-dimensional vector space. In fact, many data in high-dimensional space are perhaps superfluous and not necessary for classification. A typical way to solve this problem is to adopt dimensionality reduction techniques to find intrinsic structures or features of the data.

Up to now, an enormous volume of literature has been devoted to develop dimensionality reduction methods. Based on how to utilize the label information from patterns, these algorithms can be broadly divided into three classes: unsupervised, supervised, and semi-supervised. Supervised methods include linear discriminant analysis (LDA) [1], maximum margin criterion (MMC) [2], marginal Fisher analysis (MFA) [3], locality sensitive discriminant analysis (LSDA) [4], etc. Unsupervised methods are those such as principal component analysis (PCA) [1] and also include some manifold learning algorithms, such as locally linear embedding (LLE) [5], ISOMAP [6], and Laplacian eigenmaps (LE) [7,8] as well as its linear extension locality preserving projections (LPP) [9,10]. Semi-supervised methods [11–13] can use unlabeled data for promoting supervised methods. In this paper, we focus on linear dimensionality reduction technique, namely subspace learning, mainly due to its simplicity, solid mathematical background, efficiency, and effectiveness in many areas such as face and palmprint recognition. In recent years, some important progress has been made in the research of dimensional reduction techniques. Among them, three strategies should be highlighted. The first one is the kernel method, which uses a linear classifier algorithm to solve non-linear problems by mapping the original non-linear observations into a higher-dimensional space [14,15]. The second one is the manifold learning method, which is based on the idea that the data points are actually samples from a low-dimensional manifold that is embedded in a high-dimensional space. Manifold learning methods aim to uncover the proper parameters in order to find a low-dimensional representation of the data. The last one is matrix and tensor embedding methods. Matrix embedding methods can extract a feature matrix using a straightforward image projection technique [16,17]. In addition, tensor embedding methods can represent the image ensembles by a higher-order tensor and extract low-dimensional features using multilinear algebra methods [18]. Yan et al. [3] proposed a general framework for dimensionality reduction named as Graph Embedding, in which several representative methods such as PCA, LDA and LPP are included as special cases. To further boost the discriminating power of the manifold learning techniques, some recent discriminant learning methods combining the Fisher criterion with manifold criterion have been proposed: MFA, ubiquitously supervised subspace

* Corresponding author.
E-mail address: icg.jiawei@gmail.com (W. Jia).

0925-2312/$ - see front matter
doi:10.1016/j.neucom.2010.04.017

learning [19], etc. Recently, some new methods integrating the theory of sparse representation, compressed sensing and subspace learning have been proposed, and have been successfully applied to face recognition [20–22].

Among all the subspace learning methods, PCA and LDA [1] are two most famous ones and have become the most popular techniques for face recognition, palmprint recognition, and other biometrics [1,23–30].

Generally, PCA projects the original data into a low-dimensional space, which is spanned by the eigenvectors associated with the largest eigenvalues of the covariance matrix of all the data points. PCA finds the optimal representation of the input data by means of minimizing reconstruction error. For linearly embedded manifolds, PCA is guaranteed to discover the intrinsic dimensionality of the manifold and produce a compact representation. However, PCA does not take into consideration the label information of the input data at all, which will probably lose much useful information, and weaken the recognition accuracy, especially when encountering a large number of sample points [24].

Unlike PCA, LDA is a supervised method which takes full consideration of the class labels for patterns. It is generally believed that the class information can make the recognition algorithm more discriminative. Thus, LDA can enhance data separability more than PCA. LDA projects the original data into an optimal subspace by a linear transformation. The transformation matrix consists of the eigenvectors, which can maximize the ratio of the trace of the between-class scatter matrix to the trace of the within-class scatter matrix.

One limitation of PCA and LDA is that they only exploit the linear global Euclidean structure but fail to use the locality information. LPP is a new linear dimensionality reduction algorithm. It can preserve the locality property and achieve higher recognition rates than PCA and LDA for biometrics applications such as face and palmprint recognition [10,30]. LPP is derived from a manifold learning algorithm, i.e., LE. Compared to most manifold learning algorithms, LPP possesses a remarkable advantage that it can generate an explicit map. This map is linear and can be easily computed, like PCA and LDA.

However, LPP suffers from a limitation that it does not encode discriminant information, which is very important for recognition tasks. Recently, several modified LPP algorithms have been proposed to make use of the label information. Yu et al. [31] presented a discriminant locality preserving projections (DLPP) algorithm to improve the classification performance of LPP and applied it to face recognition. Null space discriminant locality preserving projections (NDLPP) [32] was proposed to avoid the small sample size problem of DLPP by solving an eigenvalue problem in null space. Zhu and Zhu [33] introduced an orthogonal discriminant locality preserving projections (ODLPP) method based on OLPP [34]. Cai et al. [4] proposed a locality sensitive discriminant analysis (LSDA) method by discovering the local manifold structure. Specifically, in LSDA, the data points are mapped into a subspace in which the nearby points with the same label are close to each other while the nearby points with different labels are far apart.

In this paper, we propose a locality preserving discriminant projections (LPDP) algorithm by combining LPP and MMC methods, which can be viewed as a new algorithm integrating Fisher criterion and manifold criterion. It is well known that MMC is a method proposed to maximize the trace of the difference of the between-class scatter matrix and within-class scatter matrix from which LDA can be derived by incorporating some constraints. However, MMC has the same weakness as LDA that both of them neglect the locality information. LPP preserves the local structures of samples, while MMC preserves the global discriminative structures. Obviously, the two algorithms can complement each other. Thus, LPDP is proposed by introducing MMC into the objective function of LPP, which has two advantages: (1) It retains the locality preserving characteristic. If two points are close in the original space, they are close as well in the low-dimensional space. (2) It emphasizes the discriminative information by incorporating MMC, which can make the class mean vectors have a wide spread and make every class scatter in a small space. Experimental results show that it is more suitable for recognition tasks than LPP.

The remainder of this paper is organized as follows: In Section 2 we will introduce our LPDP method in detail, and its tensorized forms will be derived in Section 3. The experimental results for applying the LPDP method to face and palmprint recognition will be presented in Section 4, followed by the discussions in Section 5. Finally, the conclusions are given in Section 6.

2. Locality preserving discriminant projections (LPDP)

Generally speaking, if there are N samples, $X=[x_1,x_2,\ldots,x_N]$, in a high-dimensional space, and the dimensionality of x_i is D, the aim of linear dimensionality reduction algorithms is to find a transformation matrix V that can map these N samples to low-dimensional vectors $Y=[y_1,y_2,\cdots y_N]$, where Y is $d\times N$ matrix ($d \ll D$). Of course, different algorithms try to minimize different objective functions for 'good mapping'.

2.1. Locality preserving projections (LPP)

In [9,10], the objective function of LPP is defined as

$$\min \sum_{ij} \|y_i - y_j\|^2 W_{ij} \tag{1}$$

where $y_i = V^T x_i$ and the matrix $W=(W_{ij})$ is a similarity matrix. The weight W_{ij} incurs a heavy penalty when neighboring points x_i and x_j are mapped far apart. Therefore, minimizing the objective function is an attempt to ensure that if x_i and x_j are "close" then y_i and y_j are close as well. A possible way of defining W is introduced as follows:

$W_{ij}=1$, if x_i is among k nearest neighbors of x_j or x_j is among k nearest neighbors of x_i; otherwise, $W_{ij}=0$. Here we assume that k is the number of neighbors. The justification for this choice of weights can be traced back to [9,10].

Suppose V is a transformation matrix, that is, $Y=V^T X$. By some simple algebra formulations, the objective function can be reduced to

$$\begin{aligned}\frac{1}{2}\sum_{ij}\|y_i-y_j\|^2 W_{ij} &= \frac{1}{2} tr\left(\sum_{ij}(V^T x_i - V^T x_j)(V^T x_i - V^T x_j)^T W_{ij}\right)\\ &= tr(\sum_i V^T x_i D_{ii} x_i{}^T V - \sum_{ij} V^T x_i W_{ij} x_j{}^T V)\\ &= tr(V^T X (D-W) X^T V)\end{aligned} \tag{2}$$

where $X=[x_1,x_2,\ldots,x_N]$, D is a diagonal matrix and D_{ii} is column (or row) sum of W, $D_{ii}=\sum_j W_{ij}$, $L=D-W$ is the Laplacian matrix.

In addition, a constraint is proposed: $V^T X D X^T V = I$.

Finally, the minimization problem reduces to finding

$$\begin{aligned}&\arg\min\, tr(V^T X L X^T V)\\ &\quad \text{s.t.}\ \ V^T X D X^T V = I\end{aligned} \tag{3}$$

Thus, the transformation matrix that minimizes the objective functions is given by the minimum eigenvalues solution to the generalized eigenvalues problem.

$$XLX^T v = \lambda X D X^T v \tag{4}$$

It is easy to show that the matrices XLX^T and XDX^T are symmetric and positive semidefinite. The vectors υ_i that minimize the objective function are given by minimum eigenvalues solutions to the generalized eigenvalues problem. Let the column vectors υ_0, $\upsilon_1,\ldots$, υ_{d-1} be the solutions of Eq. (4), ordered according to their eigenvalues, $\lambda_0,\lambda_1,\ldots,\lambda_{d-1}$. Thus, the embedding is written as follows:

$$x_i \to y_i = V^T x_i, V = [\upsilon_0,\upsilon_1,\ldots,\upsilon_{d-1}] \tag{5}$$

where y_i is a d-dimensional vector, and V is a $D \times d$ matrix.

2.2. Modified maximizing margin criterion (MMMC)

In [2,35], the objective function of MMC is written as

$$J_1 = \max\left\{\sum_{ij} p_i p_j (d(m_i,m_j) - s(m_i) - s(m_j))\right\} \tag{6}$$

where p_i and p_j are the prior probability of class i and j, respectively; m_i and m_j are the mean vectors of class i and j, respectively. Here $d(m_i, m_j)$, $s(m_i)$ and $s(m_j)$ are defined as

$$d(m_i,m_j) = \|m_i - m_j\| \tag{7}$$

$$s(m_i) = tr(S_i) \tag{8}$$

$$s(m_j) = tr(S_j) \tag{9}$$

where S_j is the covariance matrix of class j.

Thus, the optimized function can be derived as follows:

$$J_2 = max\ tr(S_b - S_w) \tag{10}$$

The matrix S_b is called the between-class scatter matrix and S_w is called the within-class scatter matrix. The dimension of $tr(S_b)$ and $tr(S_w)$ may be different in practical applications, so it is more reasonable to add a rescaling coefficient to S_w in order to balance the difference in dimension. This method is called modified MMC (MMMC) [36,37]. Here, we first translate the data to suitable places, and then rescale the data with the same label to their centroids and all the centroids are kept unchanged. So S_b is still preserved and S_w is rescaled. Consequently, the derivations are stated below

$$S'_b = \sum_{i=1}^{c} N_i (m'_i - m')(m'_i - m')^T = \sum_{i=1}^{c} N_i (m_i - m)(m_i - m)^T = S_b$$

$$S'_w = \sum_{i=1}^{c}\sum_{k=1}^{N_i} (X'_{ik} - m'_i)(X'_{ik} - m'_i)^T = \sum_{i=1}^{c}\sum_{k=1}^{N_i} \alpha (X_{ik} - m_i)(X_{ik} - m_i)^T = \alpha S_w$$

where α is the rescaling coefficient.

As a result, the MMMC can be obtained and rewritten in the following form:

$$J_3 = \max\ tr(S_b - \alpha S_w)$$

It is clear that MMC becomes a special case of MMMC when α equals to 1, so the best result of MMMC will be better or at least not worse than MMC.

2.3. Locality preserving discriminant projections (LPDP)

In this section, we will discuss the solution of LPDP. If a linear transformation $Y = V^T X$ can maximize J_3, an optimal subspace for classification will be explored. This is because the linear transformation aims to project a pattern closer to patterns in the same class but farther from those in different classes, which is exactly the goal for classification. That is to say, to find an optimal linear subspace for classification means to maximize the following optimized function:

$$J_4 = \max\ tr(V^T (S_b - \alpha S_w) V)$$

If the linear transformation obtained by LPP can satisfy J_4 simultaneously, the discriminability of the data will be improved greatly. Thus, the solution for LPDP can be represented as the following multi-object optimized problem:

$$\begin{cases} \min\ \mathrm{tr}\left(V^T XLX^T V\right) \\ \max\ \mathrm{tr}\left(V^T (S_b - \alpha S_w) V\right) \\ s.t. V^T XDX^T V = I \end{cases} \tag{11}$$

The solution to the constrained multi-object optimized problem is to find a subspace preserving the locality property and maximize the margin between different classes simultaneously, so it can be changed into the following constrained problem:

$$\begin{aligned} &\min\ tr\left(V^T \left(XLX^T - (S_b - \alpha S_w)\right) V\right) \\ &\qquad s.t.\ \ V^T XDX^T V = I \end{aligned} \tag{12}$$

Eq. (12) can be solved by Lagrangian multiplier method

$$\frac{\partial}{\partial V} tr\left(V^T \left(XLX^T - (S_b - \alpha S_w)\right) V - \lambda_i \left(V^T XDX^T V - I\right)\right) = 0$$

where λ_i is the Lagrangian multiplier. Thus we get

$$\left(XLX^T - (S_b - \alpha S_w)\right) V_i = \lambda_i XDX^T V_i \tag{13}$$

where V_i is the generalized eigenvector of $(XLX^T - (S_b - \alpha S_w))$ and XDX^T; λ_i is the corresponding eigenvalue.

Let the column vectors $V_0, V_1, \ldots, V_{d-1}$ be the solutions of Eq. (13), ordered according to their first d smallest eigenvalues $\lambda_0, \lambda_1, \ldots, \lambda_{d-1}$. Thus, the embedding is written as follows:

$$x_i \to y_i = V^T x_i, \quad V = [V_0, V_1, \ldots, V_{d-1}]$$

where y_i is a d-dimensional vector and V is a $D \times d$ matrix.

The main procedure for the locality preserving discriminant projections algorithm is summarized in Table 1.

2.4. Connection with graph construction

Generally, a graph is used to characterize data geometry (e.g., manifold) and thus plays an important role in data analysis including machine learning, such as dimensionality reduction [3], semi-supervised learning [38] and spectral clustering [7]. However, how to establish high-quality graphs is still an open problem, as stated in [39]. Recently, the graph construction problem has attracted increasing attention [40,41]. We will show in the following that LPDP provides a new graph construction and edge weight assignment method.

Eq. (13) can be rewritten as

$$\left(X(D-W)X^T - (S_t - (\alpha+1)S_w)\right) V_i = \lambda_i XDX^T V_i$$

$$(XWX^T + S_t - (\alpha+1)S_w) V_i = \hat{\lambda}_i XDX^T V_i \tag{14}$$

where $\hat{\lambda}_i = 1 - \lambda_i$.

Table 1
Locality preserving discriminant projections.

Input: Training set $X = \{(x_i, y_i)\}_{i=1}^{N}$, the number of nearest neighbors
Output: $D \times d$ feature matrix V extracted from X

(1) Project the image set $\{x_i\}$ into the PCA subspace by throwing away the smallest principal components;
(2) Construct the nearest neighbor graph and put an edge between nodes i and j if x_i and x_j are "close";
(3) Construct the weights for each edge;
(4) Do eigenvalue decomposition using Eq. (13), construct $D \times d$ feature matrix V whose columns consist of the eigenvectors corresponding to its d smallest eigenvalues

From [3] we know that the total scatter matrix S_t and within-class scatter matrix S_w can be written as

$$S_t = \frac{1}{N} X \left(I - \frac{1}{N} ee^T \right) X^T \tag{15}$$

$$S_w = X \left(I - \sum_{k=1}^{c} \frac{1}{N_k} e^k (e^k)^T \right) X^T \tag{16}$$

where e is an N-dimensional vector of all ones and e^k is an N-dimensional vector with $e^k(i) = 1$ if the label of x_i is k; 0 otherwise.

So Eq. (14) is equivalent to

$$X\left(W + \frac{1}{N}\left(I - \frac{1}{N}ee^T\right) - (\alpha+1)\left(I - \sum_{k=1}^{c}\frac{1}{N_k}e^k(e^k)^T\right)\right)X^T V_i = \hat{\lambda}_i XDX^T V_i$$

$$X\left(W + \left(\frac{1}{N}I - (\alpha+1)I - \frac{1}{N^2}ee^T\right) + (\alpha+1)\left(\sum_{k=1}^{c}\frac{1}{N_k}e^k(e^k)^T\right)\right)X^T V_i = \hat{\lambda}_i XDX^T V_i \tag{17}$$

Thus we have

$$X\hat{W}X^T V_i = \hat{\lambda}_i XDX^T V_i \tag{18}$$

where

$$\hat{W} = W + \tilde{W}, \tilde{W} = \left(\frac{1}{N}I - (\alpha+1)I - \frac{1}{N^2}ee^T\right) + (\alpha+1)\left(\sum_{k=1}^{c}\frac{1}{N_k}e^k(e^k)^T\right)$$

As we know, the graph construction of LPP fails to use the global discriminative information while that of MMC fails to utilize the locality information. However, we can see from $\widehat{W}$ first that LPDP preserves the locality characteristic since W still exists and second that it adds the global discriminative information through $\tilde{W}$. From what has been discussed above, it can be concluded that LPDP can build a new graph construction and edge weight assignment method, integrating both local information and global discriminative information. Thus, it will be more robust than LPP. An example is given in Fig. 1, in which there exist two classes of two-dimensional samples. It can be seen that the projection direction of LPDP is better than that of LPP according to the classification in the one-dimensional space. The failure of LPP may be due to its unsupervised characteristic.

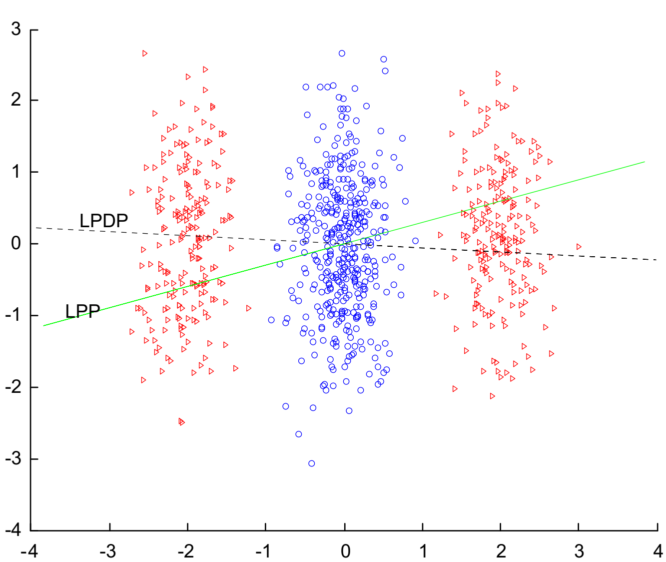


Fig. 1. Illustration of two classes of samples in two-dimensional space and the projection directions.

3. Multilinearization via tensorization

Till now, the LPDP method we have introduced is based on the basic assumption that the data are in vectorized representations. Here, we will derive its corresponding tensor version. The tensor-based criterion for LPDP can be described as: Given N data points $X_1, X_2 \ldots X_N$ in $R^{D_1} \otimes R^{D_2}$, we want to find two transformation matrices P of size $D_1 \times d_1$ and Q of size $D_2 \times d_2$ that map these N points to a set of points $Y_1, Y_2 \ldots Y_N \in R^{d_1} \otimes R^{d_2} (d_1 < D_1, d_2 < D_2)$, such that Y_i "represents" X_i, where $Y_i = P^T X_i Q$. Our method is of particular applicability in the special case where $X_1, X_2, \ldots, X_N \in \mathrm{M}$ compose a non-linear submanifold embedded in $R^{D_1} \otimes R^{D_2}$. In the following, we develop two kinds of tensorized LPDP. One is iterative, while the other is non-iterative.

3.1. Tensor locality preserving discriminant projections (TLPDP)

He et al. [42] proposed tensor subspace analysis (TSA), which is the tensor version of LPP. TSA incurs a heavy penalty if neighboring points X_i and X_j are mapped far apart. Therefore, minimizing it is an attempt to ensure that if X_i and X_j are "close" then $P^T X_i Q$ and $P^T X_j Q$ are "close" as well. We propose TLPDP by combining TSA and the tensor version of improved MMC, which can map the data into an optimal subspace for classification. Thus the discriminability of the data will be improved.

The objective function of TLPDP is given as follows:

$$\min_{P,Q} \sum_{ij} \frac{1}{2} \| P^T X_i Q - P^T X_j Q \|^2 W_{ij} - tr\left(P^T \left(S_b^Q - \alpha S_w^Q\right) P\right)$$

where

$$S_b^Q = \sum_{i=1}^{c} N_i \left(\overline{\phi_i} - \overline{\phi}\right)\left(\overline{\phi_i} - \overline{\phi}\right)^T \quad \overline{\phi_i} = \frac{1}{N_i}\sum_{k=1}^{N_i} X_k Q \quad \overline{\phi} = \frac{1}{N}\sum_{k=1}^{N} X_k Q$$

$$S_w^Q = \sum_{i=1}^{c}\sum_{k=1}^{N_i}\left(X_k Q - \overline{\phi_i}\right)\left(X_k Q - \overline{\phi_i}\right)^T \tag{19}$$

The definition of W_{ij} is the same as the one in Section 2.1.

By simple linear algebra, the objective function can be transformed to

$$tr\left(P^T\left((D_Q - W_Q) - \left(S_b^Q - \alpha S_w^Q\right)\right)P\right)$$

where

$$D_Q = \sum_i D_{ii} X_i Q Q^T X_i^T \quad \text{and} \quad W_Q = \sum_{ij} W_{ij} X_i Q Q^T X_j^T \tag{20}$$

Similarly, the objective function can also be transformed to

$$tr(Q^T((D_P - W_P) - (S_b^P - \alpha S_w^P))Q)$$

where

$$D_P = \sum_i D_{ii} X_i^T P P^T X_i, \quad \text{and} \quad W_P = \sum_{ij} W_{ij} X_i^T P P^T X_j$$

$$S_b^P = \sum_{i=1}^{c} N_i \left(\overline{\phi_i} - \overline{\phi}\right)\left(\overline{\phi_i} - \overline{\phi}\right)^T \quad \overline{\phi_i} = \frac{1}{N_i}\sum_{k=1}^{N_i} X_k^T P \quad \overline{\phi} = \frac{1}{N}\sum_{k=1}^{N} X_k^T P$$

$$S_w^P = \sum_{i=1}^{c}\sum_{k=1}^{N_i}\left(X_k^T P - \overline{\phi_i}\right)\left(X_k^T P - \overline{\phi_i}\right)^T \tag{21}$$

Therefore, we should minimize Eq. (20) and Eq. (21) at the same time. The weighted variance in the tensor subspace is defined as

$$\mathrm{var}(Y) = \sum_i \|Y_i\|^2 D_{ii} = \sum_i tr\left(D_{ii} Y_i{}^T Y_i\right) = \sum_i tr(D_{ii} Q^T X_i^T P P^T X_i Q)$$

$$= tr\left(Q^T\left(\sum_i D_{ii} X_i^T P P^T X_i\right)Q\right) = tr(Q^T D_P Q)$$

Similarly

$$\|Y_i\|^2 = tr\left(Y_i Y_i^T\right), \text{ so we also have}$$

$$\mathrm{var}(Y) = \sum_i tr\left(D_{ii} Y_i Y_i^T\right) = tr(P^T(\sum_i D_{ii} X_i Q Q^T X_i^T)P) = tr(P^T D_Q P)$$

Finally, we get the following optimization problems:

$$\min_{P,Q} \frac{tr\left(P^T\left((D_Q - W_Q) - \left(S_b^Q - \alpha S_w^Q\right)\right)P\right)}{tr(P^T D_Q P)} \tag{22}$$

$$\min_{P,Q} \frac{tr(Q^T((D_P - W_P) - (S_b^P - \alpha S_w^P))Q)}{tr(Q^T D_P Q)} \tag{23}$$

The main procedure for the tensor locality preserving discriminant projections (TLPDP) is summarized in Table 2.

3.2. Non-iterative tensor locality preserving discriminant projections (NTLPDP)

The aim of TLPDP is to solve the left projection matrix P and the right projection matrix Q simultaneously. Usually, there is no closed-form solution to P and Q for TLPDP, so an iterative optimization method should be employed in a way similar to GLRAM [43] and MDA [18]. Obviously, the iterative method is time-consuming. In fact, there is no need to solve the left projection matrix and right projection matrix at the same time. A solution can be obtained if we solve one first and then solve the other. Our method of non-iterative tensor locality preserving discriminant projections (NTLPDP) is developed based on this idea.

First, we solve the left projection matrix P, i.e., compressing 2D-data in vertical direction, making the discriminant information pack into a small number of rows. The corresponding objective function is defined as follows:

$$\min_P \sum_{ij} \frac{1}{2} \|P^T X_i - P^T X_j\|^2 W_{ij} - tr(P^T(S_b^N - \alpha S_w^N)P)$$

$$\text{where } S_b^N = \sum_{i=1}^{c} N_i\left(\overline{\phi_i} - \overline{\phi}\right)\left(\overline{\phi_i} - \overline{\phi}\right)^T \quad \overline{\phi_i} = \frac{1}{N_i}\sum_{k=1}^{N_i} X_k \quad \overline{\phi} = \frac{1}{N}\sum_{k=1}^{N} X_k$$

$$S_w^N = \sum_{i=1}^{c}\sum_{k=1}^{N_i}\left(X_k - \overline{\phi_i}\right)\left(X_k - \overline{\phi_i}\right)^T \tag{24}$$

Since $\|A\|^2 = tr(AA^T)$, we have

$$\begin{aligned}
&\sum_{ij} \frac{1}{2} \|P^T X_i - P^T X_j\|^2 W_{ij} - tr(P^T(S_b^N - \alpha S_w^N)P) \\
&= \sum_{ij} \frac{1}{2} tr\left(W_{ij} P^T (X_i - X_j)(P^T(X_i - X_j))^T\right) - tr(P^T(S_b^N - \alpha S_w^N)P) \\
&= \sum_{ij} \frac{1}{2} tr\left(W_{ij} P^T (X_i - X_j)(X_i - X_j)^T P\right) - tr(P^T(S_b^N - \alpha S_w^N)P) \\
&= tr\left(\sum_i D_{ii} P^T X_i X_i^T P - \sum_{ij} W_{ij} P^T X_i X_j^T P - P^T(S_b^N - \alpha S_w^N)P\right) \\
&= tr(P^T((D_N - W_N) - (S_b^N - \alpha S_w^N))P)
\end{aligned}$$

$$\text{where } D_N = \sum_i D_{ii} X_i X_i^T \quad W_N = \sum_{ij} W_{ij} X_i X_j^T \tag{25}$$

The corresponding constraint, which aims at maximizing the global variance, is defined as

$$\mathrm{var}(Y) = \sum_i tr\left(D_{ii} Y_i Y_i^T\right) = tr\left(P^T\left(\sum_i D_{ii} X_i X_i^T\right)P\right) = tr(P^T D_N P)$$

Finally, we get the following optimization functions:

$$\min_P \frac{tr(P^T((D_N - W_N) - (S_b^N - \alpha S_w^N))P)}{tr(P^T D_N P)} \tag{26}$$

In fact, (26) corresponds to the trace ratio problem [44–47], whose solution can be solved in an approximate way as follows:

$$((D_N - W_N) - (S_b^N - \alpha S_w^N))p = \lambda_1 D_N p \tag{27}$$

Eq. (27) is equivalent to

$$(W_N + (S_b^N - \alpha S_w^N))p = (1 - \lambda_1) D_N p = \lambda D_N p \tag{28}$$

After getting the left transform P, we can solve the right projection matrix. The second transform $Z = YQ$ performs the compression of 2D-data in horizontal direction, eliminating the correlations between rows of image Y and making its discriminant information further compact into a small number of columns.

The objective function is given as follows:

$$\begin{aligned}
&\min_Q \sum_{ij} \frac{1}{2} \|Y_i Q - Y_j Q\|^2 W_{ij} - tr(Q^T(S_b^P - \alpha S_w^P)Q) \\
&= tr(Q^T((D_P - W_P) - (S_b^P - \alpha S_w^P))Q)
\end{aligned} \tag{29}$$

Eq. (29) and Eq. (21) look like the same to each other and have the same constraint, so they have the same solution. Consequently, the main procedure for the non-iterative tensor locality preserving discriminant projections (NTLPDP) is summarized in Table 3.

Table 2
Tensor locality preserving discriminant projections.

Input: Training set $X = \{(X_i, y_i)\}_{i=1}^N$, the number of nearest neighbors, the maximum iteration steps $T_{\max}$
Output: the left projection matrix P and the right projection matrix Q
(1) Initialize $P = I_{D_1}$, $Q = I_{D_2}$, where I_{D_i} denotes the $D_i * D_i$ identity matrix;
(2) For $i = 1, 2, \ldots, T_{\max}$
  For $j = 1, 2, \ldots, n$
    (a) $Y_i = P^T X_i Q$
    (b) Compute D_Q, W_Q, S_b^Q, S_w^Q using Eqs. (19) and (20)
    (c) Perform eigenvalue decomposition using Eq. (22) to obtain P
    (d) Compute D_P, W_P, S_b^P, S_w^P using Eq. (21)
    (e) Perform eigenvalue decomposition using Eq. (23) to obtain Q
  end
end
(3) Return (P, Q)

4. Experimental results

In this section, we conduct a set of experiments to verify the effectiveness of the proposed LPDP method. Three image databases were used, including two benchmark face databases ORL and Yale, and the PolyU palmprint image database.

To facilitate choosing the reduced dimensions of tensor methods, we cropped and resized all images to be square, so only one parameter was needed in the experiments. The reduced

Table 3
Non-iterative tensor locality preserving discriminant projections (NTLPDP).

Input: Training set $X = \{(X_i, y_i)\}_{i=1}^N$, the number of nearest neighbors
Output: the left projection matrix P and the right projection matrix Q
(1) Compute D_N, W_N, S_b^N, S_w^N using Eqs. (24) and (25);
(2) Perform eigenvalue decomposition using Eq. (28) to obtain P;
(3) Perform eigenvalue decomposition using Eq. (29) to obtain Q;
(4) Return (P, Q)

dimensions of the images were then in the form of $m \times m$, where m was the height or width of the reduced image.

In each experiment, the image set was partitioned into a training set and test set with different numbers. For ease of representation, the experiments were named as p-train, which means that p images per individual were selected for training and the remaining images for test. To robustly evaluate the performance of different algorithms in different training and testing conditions, we selected images randomly and repeated the experiment 20 times in each condition. We exhibited the results in the form of mean recognition rate with standard deviation. The experiments and the corresponding results on the palmprint image database are in different forms, since palmprint images were captured in two sessions.

We compare LPDP, TLPDP and NTLPDP with several representative dimensional reduction methods such as PCA [1], LDA [1], MMC [2], LPP [9,10], NDLPP [32], locality sensitive discriminant analysis (LSDA) [4], marginal Fisher analysis (MFA) [3], multilinear discriminant analysis (MDA) [18] and concurrent subspaces analysis (CSA) [48]. The nearest neighbor classifier is employed for classification. For LPP, LPDP and NDLPP, the number of nearest neighbors k is taken to be $p-1$ as done in [30] where p is the number of images per individual selected for training. The empirically determined parameter α is taken to be 1. How to set parameters of MFA [3] is still an open problem. We therefore empirically set the parameters k_1 and k_2 of MFA in all experiments. Specifically, we set k_1 to be 2 for 2-train. Otherwise we choose the best k_1 between 2 and $p-1$ at sampled intervals of 1 and choose the value with the best MFA

Fig. 2. Eleven cropped and resized samples of one person in Yale face database.

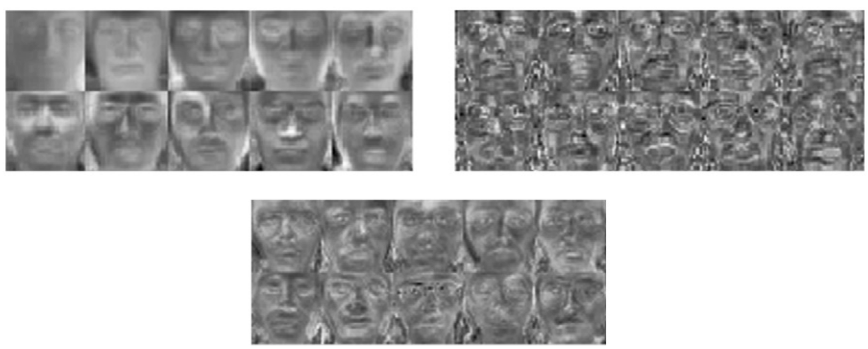


Fig. 3. The top 10 Eigenfaces, Fisherfaces and LPDPfaces of Yale dataset.

Table 4
The maximal average recognition rates (percent) across 20 runs on the Yale database and the corresponding standard deviations (std) and dimensions (shown in parentheses).

Methods	2 Train	3 Train	4 Train	5 Train
Baseline	42.63 ± 3.79 (1024)	48.08 ± 4.28 (1024)	52.86 ± 4.19 (1024)	55.44 ± 3.86 (1024)
PCA	42.63 ± 3.79 (29)	48.08 ± 4.28 (44)	52.86 ± 4.19 (59)	55.44 ± 3.86 (74)
LDA	45.19 ± 5.10 (10)	59.42 ± 4.62 (13)	68.95 ± 5.87 (13)	74.89 ± 3.52 (14)
MMC	50.93 ± 4.60 (14)	61.08 ± 4.84 (14)	67.43 ± 5.20 (13)	72.11 ± 3.42 (14)
LPP	57.19 ± 5.51 (14)	67.92 ± 4.25 (14)	75.14 ± 5.46 (16)	77.22 ± 3.5 (14)
NDLPP	56.11 ± 5.28 (14)	69.70 ± 3.66 (14)	77.47 ± 4.60 (14)	81.77 ± 3.71 (14)
MFA	54.30 ± 5.92 (13)	67.13 ± 4.95 (15)	75.52 ± 5.07 (15)	80.44 ± 4.61 (26)
LSDA	56.51 ± 5.01 (14)	68.52 ± 4.21 (14)	74.45 ± 3.54 (14)	79.80 ± 4.26 (14)
LPDP	**56.74 ± 5.90 (14)**	**71.75 ± 4.50 (14)**	**78.90 ± 3.86 (16)**	**81.78 ± 3.75 (13)**
MDA	33.70 ± 5.53 (15^2)	53.13 ± 4.97 (15^2)	63.19 ± 4.81 (15^2)	68.06 ± 4.78 (14^2)
CSA	45.07 ± 3.94 (7^2)	50.42 ± 3.85 (6^2)	55.14 ± 4.02 (7^2)	57.67 ± 3.67 (11^2)
TLPDP	50.52 ± 3.77 (8^2)	61.75 ± 4.40 (9^2)	68.28 ± 4.74 (7^2)	71.83 ± 3.24 (7^2)
NTLPDP	47.03 ± 5.93 (7^2)	60.45 ± 4.63 (9^2)	68.52 ± 4.69 (7^2)	71.0 ± 3.32 (9^2)

Methods	6 Train	7 Train	8 Train
Baseline	58.80 ± 5.28 (1024)	59.67 ± 5.29 (1024)	63.44 ± 5.47 (1024)
PCA	59.13 ± 5.29 (30)	59.83 ± 6.14 (33)	64.33 ± 5.70 (50)
LDA	79.27 ± 4.69 (14)	79.83 ± 6.73 (13)	83.22 ± 5.51 (14)
MMC	75.40 ± 4.29 (14)	77.83 ± 5.02 (14)	80.89 ± 5.12 (15)
LPP	81.6 ± 4.94 (14)	82.25 ± 4.69 (14)	84.11 ± 5.21 (15)
NDLPP	84.60 ± 3.83 (14)	87.41 ± 3.91 (14)	89.88 ± 3.70 (14)
MFA	84.13 ± 4.08 (23)	86.92 ± 3.80 (22)	89.89 ± 3.01 (23)
LSDA	83.01 ± 4.23 (14)	84.41 ± 5.34 (14)	86.56 ± 2.15 (14)
LPDP	**86.73 ± 3.95 (14)**	**88.17 ± 3.24 (14)**	**90.67 ± 2.35 (14)**
MDA	72.27 ± 4.36 (14^2)	74.58 ± 3.74 (10^2)	78.56 ± 5.06 (15^2)
CSA	61.87 ± 3.68 (5^2)	62.75 ± 5.57 (12^2)	66.44 ± 5.29 (11^2)
TLPDP	73.6 ± 4.0 (7^2)	75.0 ± 3.76 (9^2)	78.55 ± 4.63 (10^2)
NTLPDP	75.4 ± 4.68 (7^2)	76.50 ± 4.07 (5^2)	81.5 ± 5.81 (8^2)

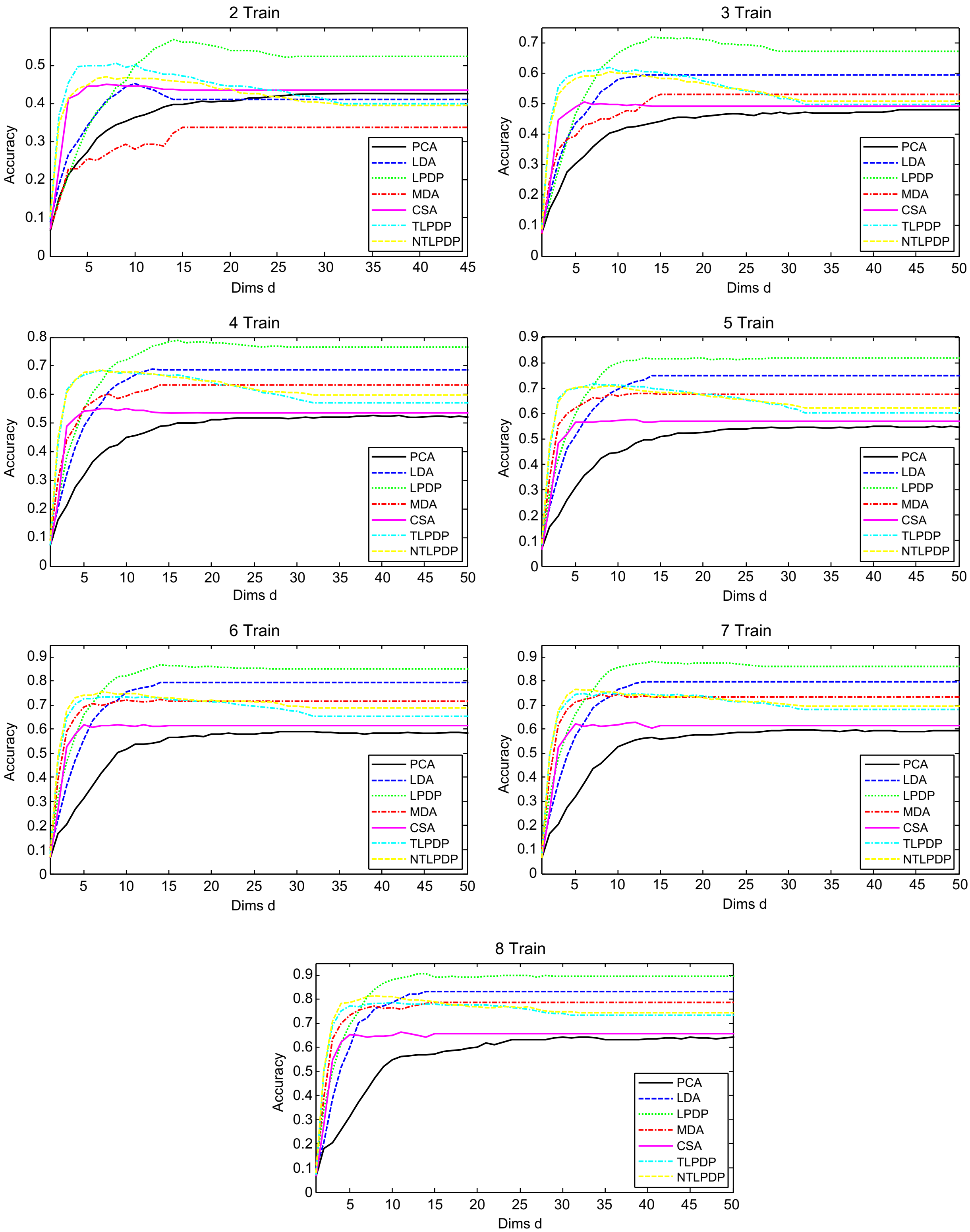


Fig. 4. Recognition accuracy vs. dimensionality on Yale database with 2–8 images for each individual randomly selected for training.

Fig. 5. Ten cropped and resized samples of one person in ORL face database.

performance. We similarly choose the best k_2 between 20 and $8c$ at sampled intervals of 10 where c is the number of classes.

4.1. Experiment using the Yale database

The Yale face database was constructed at the Yale Center for Computation Vision and Control. There are 165 images of 15 individuals (each person providing 11 different images). The images demonstrate variations in lighting condition (left-light, center-light and right-light), facial expression (normal, happy, sad, sleepy, surprised and wink), and with or without glasses. All images were also in grayscale and cropped and resized to the resolution of 32×32 pixels. We pre-processed the data by normalizing each face vector to the unit. Fig. 2 shows one object from Yale database.

The top ten Eigenfaces, Fisherfaces and LPDPfaces of Yale images are shown in Fig. 3. From this figure we can see that LPDPfaces may contain more discriminant information than Eigenfaces and Fisherfaces.

For each person, p images (p varying from 2 to 8) were randomly selected for training, and the rest were used as test samples. The training set was used to learn a face subspace. Recognition was then performed in the subspaces. In general, the recognition rates vary with the dimension of the face subspace. Table 4 shows the maximal average recognition rates across 20 runs of each method under nearest neighbor classifier and their corresponding standard deviations (std) and dimensions, where the best results are highlighted in bold face font. Due to the space limitation of Fig. 4, we only draw the recognition rate curves of PCA, LDA, LPDP, MDA, CSA, TLPDP and NDLPDP. For the baseline method, we simply performed face recognition in the original 1024-dimensional image space. Note that the upper bound of the dimensionality of LDA is $c-1$ where c is the number of individuals [1].

As can be seen, our algorithm outperformed all other methods while the PCA method performed the worst in all cases. It is very interesting that the PCA method and the baseline method have the same performance when p varies from 2 to 5. It can also be seen that for the tensor-based methods, NTLPDP outperforms MDA all the time and TLPDP outperforms MDA in most cases. Apparently, only MDA suffers from the convergence problems when training set is very small. It is clear that NTLPDP, TLPDP and MDA aim at classification while CSA aims at reconstruction, so NTLPDP, TLPDP and MDA outperform CSA in most cases.

4.2. Experiment using the ORL database

The ORL (Olivetti Research Laboratory) database contains 400 face images of 40 persons (10 samples per individual). These images were captured at different times and have different variations including expression (open or closed eyes, smiling or non-smiling) and facial details (glasses or no glasses). The images were taken with a tolerance for some tilting and rotation of the face up to 20°. All images were in grayscale and cropped and resized to the resolution of 32×32 pixels. We pre-processed the data by normalizing each face vector to the unit. Ten cropped sample images of one person in the ORL database after the scale normalization are displayed in Fig. 5.

For each individual, p images (p varying from 2 to 8) were randomly selected for training and the rest were used for testing. The experimental design is the same as in Section 4.1. The maximal average recognition rate, the corresponding dimensionality and the standard deviations across 20 runs of tests of each method are shown in Table 5. The best results are highlighted in bold face font. Due to space limitation, we only draw the recognition rate curves of PCA, LDA, LPDP, MDA, CSA, TLPDP and NDLPDP in Fig. 6.

As can be seen, our LPDP algorithm performed the best in all cases. Moreover, the optimal dimensionality obtained by LPDP and LDA is much lower than that obtained by PCA. It would be interesting to note that when there are only two training samples for each individual, the best performance of LDA is no longer obtained in a $c-1(=39)$ dimensional subspace, but a 22-dimensional subspace. It can be seen that for the tensor-based methods when the size of training set is

Table 5
The maximal average recognition rates (percent) across 20 runs on the ORL database and the corresponding standard deviations (std) and dimensions (shown in parentheses).

Methods	2 Train	3 Train	4 Train	5 Train
Baseline	66.81 ± 3.41 (1024)	77.02 ± 2.55 (1024)	81.73 ± 2.27 (1024)	86.65 ± 2.38 (1024)
PCA	66.81 ± 3.41 (79)	77.02 ± 2.55 (119)	81.73 ± 2.27 (159)	86.65 ± 2.38 (198)
LDA	70.16 ± 4.54 (22)	82.05 ± 2.25 (39)	88.73 ± 2.13 (39)	92.48 ± 1.76 (39)
MMC	75.45 ± 3.11 (39)	83.45 ± 1.90 (39)	88.83 ± 2.15 (38)	92.47 ± 2.13 (37)
LPP	78.09 ± 3.2 (39)	85.84 ± 1.62 (39)	90.25 ± 1.63 (39)	93.28 ± 1.92 (39)
NDLPP	78.92 ± 3.79 (39)	88.57 ± 1.68 (39)	93.56 ± 1.76 (39)	95.97 ± 1.56 (39)
MFA	72.27 ± 3.50 (30)	83.23 ± 2.25 (30)	88.73 ± 2.40 (30)	92.47 ± 2.58 (30)
LSDA	76.71 ± 2.14 (39)	85.07 ± 2.58 (39)	90.56 ± 3.25 (39)	93.62 ± 5.24 (39)
LPDP	**81.11 ± 3.10 (37)**	**91.14 ± 1.69 (39)**	**95.15 ± 1.38 (39)**	**97.70 ± 1.06 (40)**
MDA	71.44 ± 0.41 (15^2)	87.11 ± 2.82 (13^2)	92.67 ± 1.77 (9^2)	95.42 ± 1.58 (12^2)
CSA	69.03 ± 2.90 (9^2)	79.11 ± 2.34 (11^2)	83.81 ± 2.07 (11^2)	87.50 ± 2.31 (11^2)
TLPDP	78.07 ± 3.39 (11^2)	87.89 ± 1.99 (9^2)	92.18 ± 1.55 (9^2)	94.50 ± 1.72 (9^2)
NTLPDP	79.37 ± 3.96 (11^2)	88.23 ± 2.08 (9^2)	92.33 ± 2.25 (13^2)	94.52 ± 2.13 (9^2)

Methods	6 Train	7 Train	8 Train
Baseline	88.94 ± 2.38 (1024)	91.42 ± 2.26 (1024)	91.75 ± 2.04 (1024)
PCA	88.78 ± 2.34 (199)	90.96 ± 2.08 (104)	91.13 ± 1.94 (76)
LDA	93.63 ± 2.55 (39)	94.79 ± 2.76 (39)	96.06 ± 2.00 (39)
MMC	93.88 ± 2.34 (34)	94.92 ± 1.77 (36)	95.69 ± 1.96 (38)
LPP	94.66 ± 1.91 (39)	95.79 ± 1.9 (39)	96.81 ± 1.43 (38)
NDLPP	96.62 ± 1.72 (39)	97.04 ± 1.95 (39)	98.18 ± 1.37 (37)
MFA	94.53 ± 1.90 (30)	95.71 ± 2.12 (30)	97.06 ± 1.83 (30)
LSDA	95.31 ± 2.45 (39)	95.64 ± 1.98 (39)	96.50 ± 1.43 (39)
LPDP	**98.25 ± 1.16 (40)**	**98.63 ± 1.39 (42)**	**99.06 ± 0.90 (40)**
MDA	96.53 ± 1.51 (10^2)	96.88 ± 1.17 (8^2)	97.81 ± 1.71 (10^2)
CSA	90.59 ± 2.17 (11^2)	92.17 ± 2.3 (12^2)	93.06 ± 1.7 (12^2)
TLPDP	95.63 ± 1.60 (9^2)	96.62 ± 2.09 (11^2)	97.69 ± 1.48 (10^2)
NTLPDP	95.96 ± 1.81 (10^2)	96.54 ± 1.58 (9^2)	97.31 ± 1.63 (10^2)

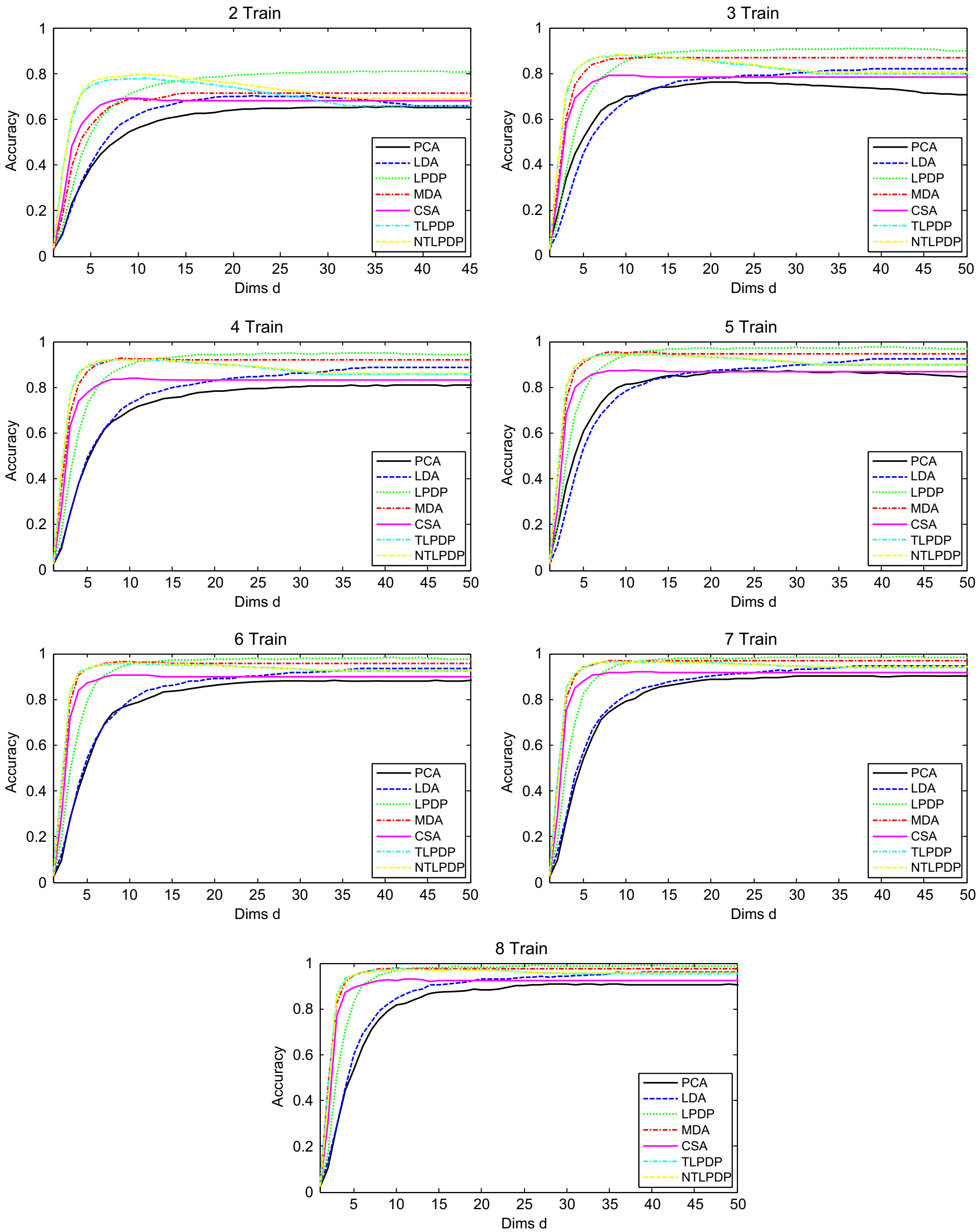


Fig. 6. Recognition accuracy vs. dimensionality on ORL database with 2–8 images for each individual randomly selected for training.

smaller, NTLPDP and TLPDP outperform MDA, and when the size of training set is larger, MDA outperforms NTLPDP and TLPDP.

4.3. Experiment using the PolyU palmprint database

Our last experiment was tested using the Hong Kong Polytechnic University (PolyU) palmprint database which contains 600 gray-scale palmprint images from 100 palms (http://www4.comp.polyu.edu.hk/~biometrics/). The database was collected in two sessions, and the average interval between the first and the second collection was two months. After preprocessing, the palmprint image was normalized to 128×128 ROI image. That is, the size of all the original palmprint images is 128×128 pixels. Some samples from the PolyU palmprint database are shown in Fig. 7.

According to the protocol of this database, the images captured in the first session were used for training and the images captured in the second session for testing. Thus, for each palm class, there are three training samples and three test samples. As the training set and test set are fixed, we only give the recognition rates of different algorithms. The maximal recognition rate of each method and the corresponding reduced dimension are listed in Table 6. The recognition rate curves of PCA, LDA, LPDP, MDA, CSA, TLPDP and NDLPDP vs. the variation of dimensions are shown in Fig. 8.

From Table 6, we can see that the performances of MFA, LSDA were poor. Maybe since the number of the training samples for each class is too small compared to the number of classes of palmprint data, the limited number of the training samples for each class cannot well represent the distribution of the palmprint data, which makes MFA and LSDA suffer from overfitting. Among all methods, LPDP achieves the highest recognition rate, i.e., 99.7 percent. Fig. 8 shows that LPDP consistently performs better than PCA and LDA, irrespective of the dimensional variation. These results demonstrate that LPDP is a good tool for palmprint recognition.

4.4. *The relationship between α and the performances*

The parameter α is defined in Eq. (13). We empirically set it to be one in the previous experiments. In this subsection, we try to examine the impact of parameter α on the performance of LPDP. The relationship between α and the performance is shown in Table 7.

From this table we can see that the value of α has no impact on the performance for the ORL face database. For the Yale face database and the PolyU palmprint database, the maximal average recognition rate and the reduced dimensions are the same when α is equal to 0.01, 0.1 and 1. When α is equal to 10 and 100, the performances become a little worse. Thus setting α to be 1 is a rather good choice.

5. Discussion

It can be seen from the results for Yale and ORL face databases that the vector-based LPDP outperforms those matrix-based methods. On the ORL database the latter is just slightly worse than the former, while on the Yale database the performance of the latter is far behind that of the former. The difference between these two databases is that variations in lighting conditions in Yale database are much more severe, which cannot be represented effectively by the rows and columns of the samples, and these variations hinder the recognition rates of the matrix-based methods.

In addition, from the experimental results, it can also be seen that MDA outperformed the LDA on the ORL face database, which may indicate that the matrix-based methods can outperform the vector-based methods in face recognition applications. However, LDA outperformed MDA on the Yale face database.

From all the experimental results we can reach the following conclusions:

(1) The discriminative analysis based methods clearly perform better than unsupervised methods as the former incorporate label information.
(2) Although NTLPDP gave the final solution in only one step, its performance is better than TLPDP in most cases on the ORL and Yale face databases. The advantage of NTLPDP over TLPDP is attributed to its closed-form solution. When the size of the training set is large, NTLPDP gave better results than TLPDP on the Yale face database. However, NTLPDP did better than TLPDP on the ORL face database when the training set size is small. They have the same performance on the PolyU palmprint database.

Besides the recognition rates, the computational time of training is also important for real applications. We list the training time of all methods obtained from Yale database in Table 8, and draw the training time curves of some representative methods in Fig. 9 due to space limitation. All of our experiments were performed on a

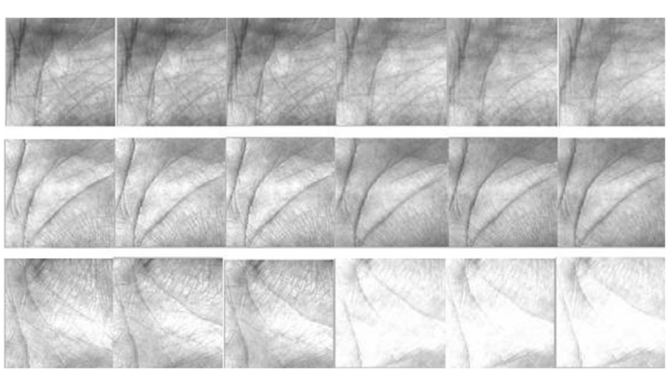


Fig. 7. Some typical samples of the cropped images found in the PolyU palmprint database.

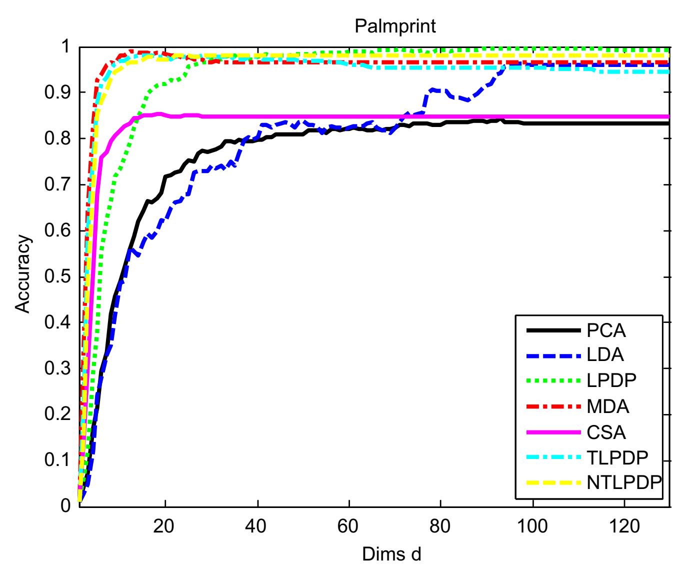


Fig. 8. Recognition accuracy vs. dimensionality on the PolyU palmprint database.

Table 6
The maximal recognition rates (percent) on the PolyU palmprint database and the corresponding dimensions.

Methods	Baseline	PCA	LDA	MMC	LPP	NDLPP	MFA	LSDA
Recognition rate	84.67	84.33	96.33	91.72	98.3	93	80.33	82.33
Dimension	16384	93	96	74	140	99	30	29

Methods	LPDP	MDA	TLPDP	NTLPDP	CSA
Recognition rate	**99.7**	99	98	98	85.33
Dimension	71	12^2	14^2	22^2	18^2

Table 7
The maximal average recognition rate and the corresponding standard deviations (percent) with the reduced dimensions for varying α on PolyU palmprint database, Yale and ORL face databases.

Dataset	$\alpha=0.01$	$\alpha=0.1$	$\alpha=1$	$\alpha=10$	$\alpha=100$
Yale	90.67 ± 2.7 (14)	90.67 ± 2.7 (14)	90.67 ± 2.4 (14)	90.22 ± 3.3 (14)	88.67 ± 4.4 (14)
ORL	99.06 ± 0.9 (40)	99.06 ± 0.9 (40)	99.06 ± 0.9 (40)	99.06 ± 0.9 (40)	99.06 ± 0.9 (40)
Palm	99.7 (71)	99.7 (71)	99.7 (71)	99.3 (89)	99 (98)

Table 8
Computational time (20 random splits) on Yale (seconds).

Methods	2 Train	3 Train	4 Train	5 Train	6 Train	7 Train	8 Train
PCA	0.207891	0.228146	0.273381	0.323161	0.373804	0.432467	0.497120
LDA	0.266633	0.286927	0.340625	0.379094	0.438158	0.499526	0.559267
MMC	0.639684	0.798576	0.902642	1.010750	1.101070	1.233276	1.370592
LPP	0.409055	0.453418	0.544829	0.629830	0.725992	0.752047	0.893495
NDLPP	0.424097	0.447398	0.551118	0.587129	0.640912	0.721610	0.772408
MFA	8.138273	8.237768	8.250829	8.331281	8.386955	8.479645	8.683751
LSDA	0.397071	0.440674	0.555785	0.672741	0.787031	0.888814	1.094846
LPDP	0.428333	0.453638	0.555530	0.649204	0.788537	0.897398	1.118984
MDA	5.851131	5.980118	6.166407	6.874212	7.664904	8.437457	8.545175
CSA	1.290729	1.709189	2.167877	2.662728	3.188897	3.548019	4.672861
TLPDP	6.819219	10.157505	15.217071	20.662618	25.143033	33.410781	43.073959
NTLPDP	0.963369	1.284985	1.765389	2.261859	2.850467	3.497947	4.132550

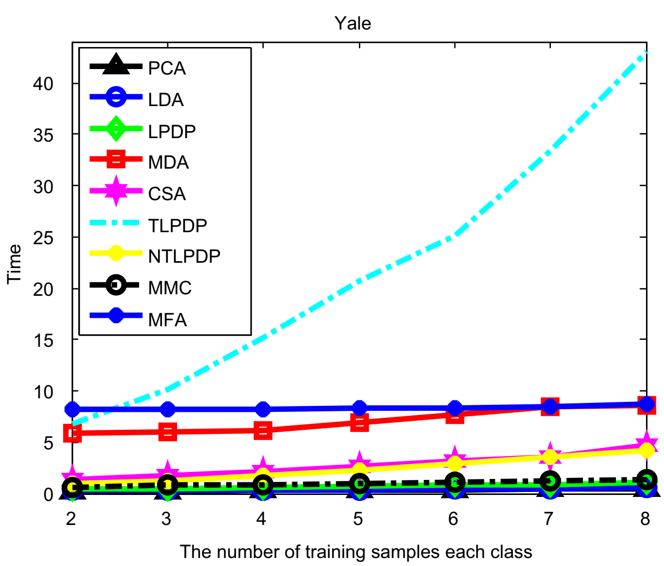


Fig. 9. Computational time (20 random splits) on Yale (seconds).

computer with a Dual-Core Intel Xeon 2.13 GHz, 12 GB RAM, and configured with Microsoft Windows Vista and Matlab 7.4. From the experimental results, we can see that although LPDP consumes a little more time than PCA, LDA, LSDA and LPP, it is more efficient than most of the other methods. Meanwhile, the training time of NTLPDP is much less than that of TLPDP.

6. Conclusions

In this paper, we presented an LPDP algorithm for dimensionality reduction. Two contributions were made in this paper. First, we presented LPDP by introducing MMC into the objective function of LPP. MMC can only preserve the global discriminant structure, while the local geometrical structure of the data is ignored. As for LPP, it can only preserve the local structure of the samples, while the global discriminant information is ignored. The most prominent property of LPDP is the complete preservation of both global discriminant and local structure of the data. Second, we developed two kinds of tensorized LPDP and found that the performance of the non-iterative version is better than that of the iterative version in most cases. Our algorithm was applied to the Yale and ORL face image databases as well as the PolyU palmprint database. Experimental results indicated the promising performance of the proposed method. Our future work is to extend LPDP to non-linear form by kernel trick.

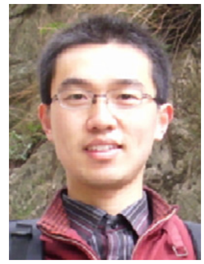
Acknowledgments

This work was supported by the grants of the National Science Foundation of China, nos. 60705007, 60905023, 60975005, 30900321, 60973153, 60873012 and 60805021, the grant of the Guide Project of Innovative Base of Chinese Academy of Sciences (CAS), no. KSCX1-YW-R-30, and the Knowledge Innovation Program of the Chinese Academy of Sciences.

References

[1] 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.
[2] H.F. Li, T. Jiang, K.S. Zhang, Efficient and robust feature extraction by maximum margin criterion, IEEE Transactions on Neural Networks 17 (1) (2006) 157–165.
[3] S.C. Yan, D. Xu, B.Y. Zhang, et al., Graph embedding and extensions: a general framework for dimensionality reduction, IEEE Transactions on Pattern Analysis and Machine Intelligence 29 (1) (2007) 40–51.
[4] D. Cai, X. He, K. Zhou et al., Locality sensitive discriminant analysis, in: Proceedings of 2007 International Joint Conference on Artificial Intelligence (IJCAI'07), Hyderabad, India, 2007.
[5] S.T. Roweis, L.K. Saul, Nonlinear dimensionality reduction by locally linear embedding, Science 290 (5500) (2000) 2323–2326.
[6] J.B. Tenenbaum, V. de Silva, J.C. Langford, A global geometric framework for nonlinear dimensionality reduction, Science 290 (5500) (2000) 2319–2324.
[7] M. Belkin, P. Niyogi, Laplacian eigenmaps and spectral techniques for embedding and clustering, Advances in Neural Information Processing Systems, 14, 2002, pp. 585–591.
[8] M. Belkin, P. Niyogi, Laplacian eigenmaps for dimensionality reduction and data representation, Neural Computation 15 (6) (2003) 1373–1396.
[9] X.F. He, P. Niyogi, Locality preserving projections, Advances in Neural Information Processing Systems, 16, , 2004 153–160.
[10] X.F. He, S.C. Yan, Y.X. Hu, et al., Face recognition using Laplacianfaces, IEEE Transactions on Pattern Analysis and Machine Intelligence 27 (3) (2005) 328–340.

[11] M. Belkin, P. Niyogi, Semi-supervised learning on Riemannian manifolds, Machine Learning 56 (1–3) (2004) 209–239.
[12] D. Zhang, Z.-H. Zhou, S. Chen, Semi-supervised dimensionality reduction, in: Proceedings of the seventh SIAM International Conference on Data Mining (SDM'07), Minneapolis, MN, 2007, pp. 629–634.
[13] D. Cai, X. He, J. Han, Semi-supervised discriminant analysis, in: IEEE International Conference on Computer Vision (ICCV' 07), Rio de Janeiro, Brazil, 2007.
[14] B. Scholkopf, A. Smola, K.R. Muller, Nonlinear component analysis as a kernel eigenvalue problem, Neural Computation 10 (5) (1998) 1299–1319.
[15] G. Baudat, F.E. Anouar, Generalized discriminant analysis using a kernel approach, Neural Computation 12 (10) (2000) 2385–2404.
[16] J. Yang, D. Zhang, A.F. Frangi, et al., 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.
[17] J. Yang, D. Zhang, X. Yong, et al., Two-dimensional discriminant transform for face recognition, Pattern Recognition 38 (7) (2005) 1125–1129.
[18] S.C. Yan, D. Xu, Q. Yang, et al., Multilinear discriminant analysis for face recognition, IEEE Transactions on Image Processing 16 (1) (2007) 212–220.
[19] J.C. Yang, S.C. Yan, T.S. Huang, Ubiquitously supervised subspace learning, IEEE Transactions on Image Processing 18 (2) (2009) 241–249.
[20] J. Wright, A.Y. Yang, A. Ganesh, et al., Robust face recognition via sparse representation, IEEE Transactions on Pattern Analysis and Machine Intelligence 31 (2) (2009) 210–227.
[21] P.O. Hoyer, Non-negative matrix factorization with sparseness constraints, Journal of Machine Learning Research 5 (2004) 1457–1469.
[22] X.R. Pu, Z. Yi, Z.M. Zheng et al., Face recognition using Fisher non-negative matrix factorization with sparseness constraints, Advances in Neural Networks—Isnn 2005, Part 2, Proceedings, 3497, 112–117, 2005.
[23] J.H. Friedman, Regularized discriminant-analysis, Journal of the American Statistical Association 84 (405) (1989) 165–175.
[24] A.M. Martinez, A.C. Kak, PCA versus LDA, IEEE Transactions on Pattern Analysis and Machine Intelligence 23 (2) (2001) 228–233.
[25] J.P. Ye, T. Xiong, Computational and theoretical analysis of null space and orthogonal linear discriminant analysis, Journal of Machine Learning Research 7 (2006) 1183–1204.
[26] J. Ye, Least squares linear discriminant analysis, in: The Twenty-Fourth International Conference on Machine Learning (ICML 2007), 2007, pp. 1087–1093.
[27] D. Cai, X.F. He, J.W. Han, SRDA: an efficient algorithm for large-scale discriminant analysis, IEEE Transactions on Knowledge and Data Engineering 20 (1) (2008) 1–12.
[28] S.W. Ji, J.P. Ye, Generalized linear discriminant analysis: a unified framework and efficient model selection, IEEE Transactions on Neural Networks 19 (10) (2008) 1768–1782.
[29] W.S. Zheng, J.H. Lai, P.C. Yuen, et al., Perturbation LDA: learning the difference between the class empirical mean and its expectation, Pattern Recognition 42 (5) (2009) 764–779.
[30] J. Yang, D. Zhang, J.Y. Yang, et al., Globally maximizing, locally minimizing: unsupervised discriminant projection with applications to face and palm biometrics, IEEE Transactions on Pattern Analysis and Machine Intelligence 29 (4) (2007) 650–664.
[31] 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.
[32] L.P. Yang, W.G. Gong, X.H. Gu, et al., Null space discriminant locality preserving projections for face recognition, Neurocomputing 71 (16–18) (2008) 3644–3649.
[33] L. Zhu, S.N. Zhu, Face recognition based on orthogonal discriminant locality preserving projections, Neurocomputing 70 (7–9) (2007) 1543–1546.
[34] D. Cai, X.F. He, J.W. Han, et al., Orthogonal laplacianfaces for face recognition, IEEE Transactions on Image Processing 15 (11) (2006) 3608–3614.
[35] J. Liu, S.C. Chen, X.Y. Tan, et al., Comments on ''Efficient and robust feature extraction by maximum margin criterion, IEEE Transactions on Neural Networks 18 (6) (2007) 1862–1864.
[36] B. Li, C.H. Zheng, D.S. Huang, Locally linear discriminant embedding: an efficient method for face recognition, Pattern Recognition 41 (12) (2008) 3813–3821.
[37] K. Fukunaga, Introduction to Statistical Pattern Recognition, second ed., Academic Press Professional, 1990.
[38] X. Zhu, Semi-supervised learning literature survey, 〈http://www.cs.wisc.edu/~jerryzhu/pub/ssl_survey.pdf〉.
[39] W. Liu, S.-F. Chang, Robust multi-class transductive learning with graphs, in: Proceedings of 2009 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'09), 2009.
[40] S.I. Daitch, J.A. Kelner, D.A. Spielman, Fitting a graph to vector data, in: International Conference on Machine Learning (ICML'2009), 2009.
[41] L.S. Qiao, S.C. Chen, X.Y. Tan, Sparsity preserving projections with applications to face recognition, Pattern Recognition 43 (1) (2010) 331–341.
[42] X. He, D. Cai, P. Niyogi, Tensor subspace analysis, Advances in Neural Information Processing Systems, , 2006 499.
[43] J.P. Ye, Generalized low rank approximations of matrices, Machine Learning 61 (1–3) (2005) 167–191.
[44] Y.Q. Jia, F.P. Nie, C.S. Zhang, Trace ratio problem revisited, IEEE Transactions on Neural Networks 20 (4) (2009) 729–735.
[45] S.C. Yan, X.O. Tang, Trace quotient problems revisited, in: Computer Vision—Eccv 2006, Part 2, Proceedings, 2006, pp. 232–244.
[46] Y.F. Guo, S.J. Li, J.Y. Yang, et al., A generalized Foley–Sammon transform based on generalized fisher discriminant criterion and its application to face recognition, Pattern Recognition Letters 24 (1–3) (2003) 147–158.
[47] H. Wang, S. Yan, D. Xu et al., Trace ratio vs. ratio trace for dimensionality reduction, in: IEEE Conference on Computer Vision and Pattern Recognition (CVPR07), 2007.
[48] D. Xu, S.C. Yan, L. Zhang, et al., Reconstruction and recognition of tensor-based objects with concurrent subspaces analysis, IEEE Transactions on Circuits and Systems for Video Technology 18 (1) (2008) 36–47.



Jie Gui received the B.E. degree in Computer Science in Hohai University in 2004. He is currently working toward the Ph.D. degree in the Department of Automation, The University of Science and Technology of China. His research interests are machine learning, pattern recognition and image processing.

Wei Jia received the B.Sc. degree in Informatics from Center of China Normal University, Wuhan, China, in 1998, the M.Sc. degree in Computer Science from Hefei University of technology, Hefei, China, in 2004 and the Ph.D. degree in Pattern Recognition and Intelligence System from University of Science and Technology of China, Hefei, China, in 2008. From June 2007 to February 2008, he was a research assistant in Biometrics Research Centre, Department of Computing, Hong Kong Polytechnic University. He is currently an assistant research fellow in Hefei Institute of Intelligent Machines, Chinese Academy of Science. His research interests include palmprint recognition, pattern recognition and image processing.

Lin Zhu received B.Sc. degree in Mechanical Engineering and Automation from Nanjing University of Science and Technology in 2006. He is a Master-Doctoral Program student in department of automation, University of Science and Technology of China, Hefei, China. His research interests are artificial neural networks, neuroinformatics and data mining.

Shu-Lin Wang was born in Sichuan, China. Currently, he is working as a Postdoctoral Fellow in Intelligent Computing Laboratory, Hefei Institute of Intelligent Machines, Chinese Academy of Sciences, China. He received his Ph.D. degree in Computer Science and Technology from the National University of Defense Technology, China, in 2008 (Advisor: Prof. Huowang Chen, Academician and Prof. Ji Wang). He also received his M.Sc. degree in Computer Application from the National University of Defense Technology, China, in 1997 (Advisor: Prof. Deqing Zhang), and obtained his B.Sc. degree in Computer Application from China University of Geosciences in 1989. He also worked in Hunan University as an associate professor. His current research interest includes Bioinformatics, Software Engineering, and Complex System.

De-Shuang Huang received the B.Sc. degree in Electronic Engineering from the Institute of Electronic Engineering, Hefei, China, in 1986, the M.Sc. degree in Electronic Engineering from the National Defense University of Science and Technology, Changsha, China, in 1989 and the Ph.D. degree in Electronic Engineering from Xidian University, Xian, China, in 1993. From 1993 to 1997, he was a Postdoctoral Student at the Beijing Institute of Technology, Beijing, China, and the National Key Laboratory of Pattern Recognition, Chinese Academy of Sciences (CAS), Beijing. In 2000, he was a professor, and joined the Institute of Intelligent Machines, CAS, as a member of the Hundred Talents Program of CAS. He has published over 190 papers and, in 1996, published a book entitled *Systematic Theory of Neural Networks for Pattern Recognition*. His research interests include pattern recognition, machine leaning, bioinformatics and image processing.