Research Article

Face Recognition Using Double Sparse Local Fisher Discriminant Analysis

Zhan Wang,[1,2] Qiuqi Ruan,[1,2] and Gaoyun An[1,2]

[1]*Institute of Information Science, Beijing Jiaotong University, Beijing 100044, China*
[2]*Beijing Key Laboratory of Advanced Information Science and Network Technology, Beijing 100044, China*

Correspondence should be addressed to Zhan Wang; wz6308@hotmail.com

Local Fisher discriminant analysis (LFDA) was proposed for dealing with the multimodal problem. It not only combines the idea of locality preserving projections (LPP) for preserving the local structure of the high-dimensional data but also combines the idea of Fisher discriminant analysis (FDA) for obtaining the discriminant power. However, LFDA also suffers from the undersampled problem as well as many dimensionality reduction methods. Meanwhile, the projection matrix is not sparse. In this paper, we propose double sparse local Fisher discriminant analysis (DSLFDA) for face recognition. The proposed method firstly constructs a sparse and data-adaptive graph with nonnegative constraint. Then, DSLFDA reformulates the objective function as a regression-type optimization problem. The undersampled problem is avoided naturally and the sparse solution can be obtained by adding the regression-type problem to a ℓ^1 penalty. Experiments on Yale, ORL, and CMU PIE face databases are implemented to demonstrate the effectiveness of the proposed method.

1. Introduction

Dimensionality reduction tries to transform the high-dimensional data into lower-dimensional space in order to preserve the useful information as much as possible. It has a wide range of applications in pattern recognition, machine learning, and computer vision. A well-known approach for supervised dimensionality reduction is linear discriminant analysis (LDA) [1]. It tries to find a projection transformation by maximizing the between-class distance and minimizing the within-class distance simultaneously. In practical applications, LDA usually suffers from some limitations. First, LDA usually suffers from the undersampled problem [2]; that is, the dimension of data is larger than the number of training samples. Second, LDA can only uncover the global Euclidean structure. Third, the solution of LDA is not sparse, which cannot give the physical interpretation.

To deal with the first problem, many methods have been proposed. Belhumeur et al. [3] proposed a two-stage principal component analysis (PCA) [4] + LDA method, which utilizes PCA to reduce dimensionality so as to make the within-class

scatter matrix nonsingular, followed by LDA for recognition. However, some useful information may be compromised in the PCA stage. Chen et al. [5] extracted the most discriminant information from the null space of within-class scatter matrix. However, the discriminant information in the non-null space of within-class scatter matrix would be discarded. Huang et al. [6] proposed an efficient null-space approach, which first removes the null space of total scatter matrix. This method is based on the observation that the null space of total scatter matrix is the intersection of the null space of between-class scatter matrix and the null space of within-class scatter matrix. Qin et al. [7] proposed a generalized null space uncorrelated Fisher discriminant analysis technique that integrates the uncorrelated discriminant analysis and weighted pairwise Fisher criterion for solving the undersampled problem. Yu and Yang [8] proposed direct LDA (DLDA) to overcome the undersampled problem. It removes the null space of between-class scatter matrix and extracts the discriminant information that corresponds to the smallest eigenvalues of the within-class scatter matrix. Zhang et al. [9] proposed an exponential discriminant analysis (EDA) method to extract the most

discriminant information which is contained in the null space of the within-class scatter matrix.

To deal with the second problem, many methods have been developed for dimensionality reduction. These methods focus on finding the local structure of the original data space. Locality preserving projections (LPP) [10] was proposed to find an embedding subspace that preserves local information. One limitation of LPP is that it is an unsupervised method. Because the discriminant information is important to the classification tasks, some locality preserving discriminant methods have been proposed. Discriminant locality preserving projection (DLPP) [11] was proposed to improve the performance of LPP. Laplacian linear discriminant analysis (LapLDA) [12] tries to capture the global and local structure of the data simultaneously by integrating LDA with a locality preserving regularizer. Local Fisher discriminant analysis (LFDA) [13] was proposed to deal with the multimodal problem. It combines the ideas of Fisher discriminant analysis (FDA) [1] and LPP and maximizes between-class separability and preserves within-class local structure simultaneously. In LDA, the dimension of the embedding space should be less than the number of classes. This limitation can be solved by using the LFDA algorithm.

To deal with the third problem, many dimensionality reduction methods integrating the sparse representation theory have been proposed. These methods can be classified into two categories. The first category focuses on finding a subspace spanned by sparse vectors. The sparse projection vectors reveal which element or region of the patterns is important for recognition tasks. Sparse PCA (SPCA) [14] was proposed by using the least angle regression and elastic net to produce sparse principal components. Sparse discriminant analysis (SDA) [15] and sparse linear discriminant analysis (SLDA) [16] were proposed to learn a sparse discriminant subspace for feature extraction and classification in biological and medical data analysis. Both methods try to transform the original objective into a regression-type problem and add a lasso penalty to obtain the sparse projection axes. One disadvantage of these methods is that the number of sparse vectors is at most $C - 1$. C is the number of class. The second category focuses on the sparse reconstructive weight among the training samples. Graph embedding framework views many dimensionality reduction methods as the graph construction [17]. The k-nearest neighbor and the ε-ball based methods are two popular ways for graph construction. Instead of them, Cheng et al. built the ℓ^1-graph based on sparse representation [18]. The ℓ^1-graph has proved that it is efficient and robust to data noise. ℓ^1-graph based subspace learning methods include sparse preserving projections (SPP) [19] and discriminant sparse neighborhood preserving embedding (DSNPE) [20].

Motivated by ℓ^1-graph and sparse subspace learning, in this paper, we proposed double sparse local Fisher discriminant analysis (DSLFDA) for multimodal problem. It measures the similarity on the graph by integrating the sparse representation and nonnegative constraint. To obtain sparse projection vectors, the objective function can be transformed into a regression-type problem. Furthermore, the space

spanned by the solution of regression-type problem is identical to that spanned by the solution of original problem. The proposed DSLFDA has two advantages: (1) it remains the sparse characteristic of ℓ^1-graph; (2) to enhance the discriminant power of DSLFDA, the label information is used in the definition of local scatter matrices. Meanwhile, the projection vectors are sparse, which can make the physical meaning of the patterns clear. The proposed method is applied to face recognition and is examined using the Yale, ORL, and PIE face databases. Experimental results show that it can enhance the performance of LFDA effectively.

The rest of this paper is organized as follows. In Section 2, the LFDA algorithm is presented. The double sparse local Fisher discriminant analysis algorithm is proposed in Section 3. In Section 4, experiments are implemented to evaluate our proposed algorithm. The conclusions are given in Section 5.

2. Related Work

In this section, we give a brief of LDA and LFDA. Given a data set $X = [x_1, x_2, \ldots, x_n] \in \mathbb{R}^{m \times n}$ with each column corresponding to a data sample, $x_i \in \mathbb{R}^m$ ($1 \le i \le n$), the class label of x_i is set to $y_i \in \{1, 2, \ldots, c\}$, and c is the number of classes. We denote n_i as the number of samples in the ith class. Dimensionality reduction tries to map the point $x_i \in \mathbb{R}^m$ into $z_i \in \mathbb{R}^d$ ($d \ll m$) by the linear transformation:

$$z_i = V^T x_i \quad (i = 1, 2, \ldots, n), \text{ where } V \in \mathbb{R}^{m \times d}. \quad (1)$$

The above transformation can be written as matrix form:

$$Z = V^T X, \quad (2)$$

where $Z = [z_1, z_2, \ldots, z_n] \in \mathbb{R}^{d \times n}$.

2.1. Linear Discriminant Analysis. Linear discriminant analysis tries to find the discriminant vectors by the Fisher criterion, that is, the within-class distance is minimized and the between-class distance is maximized simultaneously. The within-class scatter matrix S_w and between-class scatter matrix S_b are, respectively, defined as follows:

$$S_w = \sum_{i=1}^{c} \sum_{x \in X_i} (x - m_i)(x - m_i)^T,$$

$$S_b = \sum_{i=1}^{c} n_i (m_i - m)(m_i - m)^T, \quad (3)$$

where X_i is the data set of class i. m_i is the mean of the samples in class i and m is the mean of the total data. LDA seeks the optimal projection matrix by maximizing the following Fisher criterion:

$$V_{\text{opt}} = \arg\max_{V} \frac{V^T S_b V}{V^T S_w V}. \quad (4)$$

The above optimization is equivalent to solving the following generalized eigenvalue problem:

$$S_b v_i = \lambda_i S_w v_i; \quad (5)$$

$V = [v_1, v_2, \ldots, v_d]$ consists of the eigenvectors of $S_w^{-1}S_b$ corresponding to the first d largest eigenvalues.

2.2. Local Fisher Discriminant Analysis.

2.2. Local Fisher Discriminant Analysis. Local Fisher discriminant analysis (LFDA) is also a discriminant analysis method. It aims to deal with the multimodal problem. The local within-class scatter matrix S_{lw} and the between-class scatter matrix S_{lb} are defined as

$$S_{\text{lw}} = \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n}W_{ij}^{\text{lw}}\left(x_i - x_j\right)\left(x_i - x_j\right)^T,$$

$$S_{\text{lb}} = \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n}W_{ij}^{\text{lb}}\left(x_i - x_j\right)\left(x_i - x_j\right)^T, \tag{6}$$

where

$$W_{ij}^{\text{lw}} = \begin{cases} \dfrac{A_{ij}}{n_k}, & x_i, x_j \text{ belong to } k\text{th class,} \\ 0, & \text{otherwise,} \end{cases}$$

$$W_{ij}^{\text{lb}} = \begin{cases} A_{ij}\left(\dfrac{1}{n} - \dfrac{1}{n_k}\right), & x_i, x_j \text{ belong to } k\text{th class,} \\ \dfrac{1}{n}, & \text{otherwise;} \end{cases} \tag{7}$$

A is the affinity matrix. $A_{ij} = \exp(-\|x_i - x_j\|^2/\eta_i\eta_j)$, and η_i is the local scaling of x_i defined by $\eta_i = \|x_i - x_i^{(k)}\|$ where $x_i^{(k)}$ is the kth nearest neighbor of x_i.

The objection function of LFDA is formulated as

$$V_{\text{opt}} = \arg\max_{V}\frac{\text{tr}\left(V^TS_{\text{lb}}V\right)}{\text{tr}\left(V^TS_{\text{lw}}V\right)}, \tag{8}$$

where $\text{tr}(\cdot)$ is the trace of a matrix. The projection matrix can be obtained by calculating the eigenvectors of the following generalized eigenvalue problem:

$$S_{\text{lb}}v_i = \lambda S_{\text{lw}}v_i, \quad i = 1, 2, \ldots, d; \tag{9}$$

$V = [v_1, v_2, \ldots, v_d]$. Because of the definition of matrix A, LFDA can effectively preserve the local structure of the data.

3. Double Sparse Local Fisher Discriminant Analysis

3.1. Graph Construction by Sparse Representation. In LFDA, the affinity matrices are defined by the local scaling method. This method can be regarded as an extension of k-nearest neighbor method. Recent research shows that the ℓ^1-graph is robust to data noise and efficient for finding the underlying manifold structure. Therefore, we defined the new affinity matrix $\overline{A}$ by sparse representation theory. Let

$\overline{A} = [a_1, a_2, \ldots, a_n]$; each a_i is a sparse vector and obtained by the following ℓ^1-minimization problem:

$$\begin{aligned} \min \quad & \|a_i\|_1 \\ \text{s.t} \quad & x_i = Xa_i \\ & 1 = \mathbf{1}^Ta_i \\ & a_{ij} \geq 0, \quad \forall j, \end{aligned} \tag{10}$$

where $a_i = [a_{i1}, \ldots, a_{i,i-1}, 0, a_{i,i+1}, \ldots, a_{i,n}]^T$ is a n-dimensional vector in which the ith element is equal to zero. $\mathbf{1} \in \mathbb{R}^n$ is a vector of all ones. The ℓ^1-minimization problem (10) can be solved by many efficient numerical algorithms. In this paper, the LARS algorithm [21] is used for solving problem (10). The matrix $\overline{A}$ can be seen as the similarity measurement by setting the matrix $\overline{A} = (\overline{A} + \overline{A}^T)/2$. Therefore, the new local scatter matrices can be defined as follows:

$$\overline{S}_{\text{lw}} = \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n}\overline{W}_{ij}^{\text{lw}}\left(x_i - x_j\right)\left(x_i - x_j\right)^T, \tag{11}$$

$$\overline{S}_{\text{lb}} = \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n}\overline{W}_{ij}^{\text{lb}}\left(x_i - x_j\right)\left(x_i - x_j\right)^T, \tag{12}$$

where $\overline{W}^{\text{lw}}$ and $\overline{W}^{\text{lb}}$ are the weight matrices and defined as

$$\overline{W}_{ij}^{\text{lw}} = \begin{cases} \dfrac{\overline{A}_{ij}}{\sqrt{n_k}}, & x_i, x_j \text{ belong to } k\text{th class,} \\ 0, & \text{otherwise,} \end{cases}$$

$$\overline{W}_{ij}^{\text{lb}} = \begin{cases} \overline{A}_{ij}\left(\dfrac{1}{\sqrt{n}} - \dfrac{1}{\sqrt{n_k}}\right), & x_i, x_j \text{ belong to } k\text{th class,} \\ \dfrac{1}{\sqrt{n}}, & \text{otherwise.} \end{cases} \tag{13}$$

The final objective function is described as follows:

$$V_{\text{opt}} = \arg\max_{V}\frac{V^T\overline{S}_{\text{lb}}V}{V^T\overline{S}_{\text{lw}}V}. \tag{14}$$

The optimal projection V can be obtained by solving the following generalized eigenvalue problem:

$$\overline{S}_{\text{lb}}v_i = \lambda\overline{S}_{\text{lw}}v_i, \quad i = 1, 2, \ldots, d. \tag{15}$$

When the matrix $\overline{S}_{\text{lw}}$ is nonsingular, the eigenvectors are obtained by the eigendecomposition of matrix $(\overline{S}_{\text{lw}})^{-1}\overline{S}_{\text{lb}}$. However, the projection matrix is not sparse.

3.2. Finding the Sparse Solution.

We first reformulate formulas (11) and (12) in matrix form. Consider

$$
\begin{aligned}
\overline{S}_{\text{lb}} &= \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n}\overline{W}_{ij}^{(\text{lb})}\left(x_i - x_j\right)\left(x_i - x_j\right)^T \\
&= \frac{1}{2}\sum_{i,j=1}^{n}\overline{W}_{ij}^{(\text{lb})}\left(x_i x_i^T + x_j x_j^T - x_i x_j^T - x_j x_i^T\right) \\
&= \sum_{i=1}^{n}\left(\sum_{j=1}^{n}\overline{W}_{ij}^{(\text{lb})}\right)x_i x_i^T - \sum_{i,j=1}^{n}\overline{W}_{ij}^{(\text{lb})}x_i x_j^T \\
&= X\left(\overline{D}^{(\text{lb})} - \overline{W}^{(\text{lb})}\right)X^T = X\overline{L}^{(\text{lb})}X^T,
\end{aligned}
\tag{16}
$$

where $\overline{D}^{(\text{lb})}$ is the diagonal matrix and the ith diagonal element is $\overline{D}_{i,i}^{(\text{lb})} = \sum_{j=1}^{n}\overline{W}_{ij}^{(\text{lb})}$, $\overline{L}^{(\text{lb})} = \overline{D}^{(\text{lb})} - \overline{W}^{(\text{lb})}$. Similarly, formula (12) can be expressed as

$$
\overline{S}_{\text{lw}} = X\overline{L}^{(\text{lw})}X^T,
\tag{17}
$$

where $\overline{L}^{(\text{lw})} = \overline{D}^{(\text{lw})} - \overline{W}^{(\text{lw})}$, $\overline{D}^{(w)}$ is the diagonal matrix, and the ith diagonal element is $\overline{D}_{i,i}^{(\text{lw})} = \sum_{j=1}^{n}\overline{W}_{ij}^{(\text{lw})}$.

Matrices $\overline{L}^{(\text{lb})}$ and $\overline{L}^{(\text{lw})}$ are always symmetric and positive semidefinite; therefore, the eigendecomposition of $\overline{L}^{(\text{lb})}$ and $\overline{L}^{(\text{lw})}$ can be expression as follows:

$$
\overline{L}^{(\text{lb})} = U_b \Sigma_b U_b^T, \qquad \overline{L}^{(\text{lw})} = U_w \Sigma_w U_w^T,
\tag{18}
$$

where Σ_b and Σ_w are the diagonal matrices. Their diagonal elements are the eigenvalues of matrices $\overline{L}^{(\text{lb})}$ and $\overline{L}^{(\text{lw})}$, respectively. So $\overline{S}_{\text{lb}}$ and $\overline{S}_{\text{lw}}$ can be rewritten as

$$
\begin{aligned}
\overline{S}_{\text{lb}} &= X\overline{L}^{(\text{lb})}X^T = XU_b\Sigma_b U_b^T X^T = XU_b\Sigma_b^{1/2}\left(XU_b\Sigma_b^{1/2}\right)^T \\
&= H_b H_b^T, \\
\overline{S}_{\text{lw}} &= X\overline{L}^{(\text{lw})}X^T = XU_w\Sigma_w U_w^T X^T = XU_w\Sigma_w^{1/2}\left(XU_w\Sigma_w^{1/2}\right)^T \\
&= H_w H_w^T,
\end{aligned}
\tag{19}
$$

where $H_b = XU_b\Sigma_b^{1/2}$ and $H_w = XU_w\Sigma_w^{1/2}$.

The following result which was inspired by [14, 16] gives the relationship between problem (10) and the regression-type problem.

Theorem 1. *Suppose that $\overline{S}_{lw}$ is positive definite; its Cholesky decomposition can be expressed as $\overline{S}_{lw} = F_w F_w^T$, where $F_w \in \mathbb{R}^{m \times m}$ is a lower triangular matrix. Let $V = [v_1, v_2, \ldots, v_d]$ be the eigenvector of problem (15) associated with the first d largest eigenvalues. Let $P = [p_1, p_2, \ldots, p_d]$ and $Q = [q_1, q_2, \ldots, q_d]$ be the optimal solution to the following problem:*

$$
\begin{aligned}
J(P,Q) &= \arg\min_{P,Q}\sum_{i=1}^{n}\left\|F_w^{-1}H_b(:,i) - PQ^T H_b(:,i)\right\|^2 \\
&\quad + \alpha\sum_{j=1}^{d}q_j^T S_{lw}q_j
\end{aligned}
\tag{20}
$$

$$
s.t. \quad P^T P = I,
$$

where $\alpha > 0$ and $H_b(:,i)$ is the ith column of H_b. Then the columns of Q span the same linear space as well as those of V.

To obtain sparse projection vectors, we add a ℓ^1 penalty to the objective function (20):

$$
\begin{aligned}
J_1(P,Q) &= \arg\min_{P,Q}\sum_{i=1}^{n}\left\|F_w^{-1}H_b(:,i) - PQ^T H_b(:,i)\right\| \\
&\quad + \alpha\sum_{j=1}^{d}q_j^T S_{\text{lw}}q_j + \zeta\sum_{j=1}^{d}\left\|q_j\right\|_1
\end{aligned}
\tag{21}
$$

$$
s.t. \quad P^T P = I.
$$

Generally speaking, it is difficult to compute the optimal P and Q simultaneously. An iterative algorithm was usually used for solving problem (21). For a fixed P, there exists an orthogonal matrix $\widehat{P}$ such that $[P, \widehat{P}]$ is $m \times m$ column orthogonal matrix. Then the first term of (21) can be rewritten as

$$
\begin{aligned}
&\sum_{i=1}^{n}\left\|F_w^{-1}H_b(:,i) - PQ^T H_b(:,i)\right\| \\
&= \left\|H_b^T F_w^{-T} - H_b^T QP^T\right\|^2 \\
&= \left\|H_b^T F_w^{-T}\left[P, \widehat{P}^T\right] - H_b^T QP^T\left[P, \widehat{P}^T\right]\right\|^2 \\
&= \left\|H_b^T F_w^{-T}P - H_b^T Q\right\|^2 + \left\|H_b^T F_w^{-T}\widehat{P}\right\|^2 \\
&= \sum_{j=1}^{d}\left\|H_b^T F_w^{-T}p_j - H_b^T q_j\right\|^2 + \left\|H_b^T F_w^{-T}\widehat{P}\right\|^2.
\end{aligned}
\tag{22}
$$

If P is fixed, then problem (21) is transformed into

$$
\begin{aligned}
\arg\min_{Q}\sum_{i=1}^{d}\left\|H_b^T F_w^{-T}p_j - H_b^T q_j\right\|^2 + \alpha\sum_{j=1}^{d}q_j^T S_{\text{lw}}q_j \\
+ \zeta\sum_{j=1}^{d}\left\|q_j\right\|_1
\end{aligned}
\tag{23}
$$

which is equivalent to d independent LASSO problem.

For a fixed Q, problem (21) is equivalent to minimizing the following problem with ignoring the constant terms:

$$
\sum_{i=1}^{n}\left\|F_w^{-1}H_b(:,i) - PQ^T H_b(:,i)\right\| = \left\|H_b^T G_w^{-T} - H_b^T QP^T\right\|^2,
\tag{24}
$$

FIGURE 1: Sample images of two individuals from Yale database.

which is subject to $P^T P = I$. The optimal solution can be obtained by computing the singular value decomposition

$$F_w^{-1} H_b H_b^T Q = U_P \Sigma_P V_P \tag{25}$$

and $P = U_p V_p^T$.

The algorithm procedure of DSLFDA is summarized as follows.

Input: the data matrix X.

Output: the sparse projection matrix Q.

(1) Calculate affinity matrix $\overline{A}$ by ℓ^1-minimization problem (10).

(2) Calculate matrix $H_b = X U_b \Sigma_b^{1/2}$ by (18) and matrix F_w by the Cholesky decomposition of $\overline{S}_{\text{lw}}$.

(3) Initialize matrix P as an arbitrary column orthogonal matrix.

(4) For given P, solve ℓ^1-minimization problem (23) which is equivalent to d independent LASSO problem.

(5) Calculate the SVD of $F_w^{-1} H_b H_b^T Q = U_P \Sigma_P V_P$ and update $P = U_p V_p^T$.

(6) Repeat steps 4 and 5 until converges.

4. Experimental Results

In this section, we use the proposed DSLFDA method for face recognition. Three face image databases, that is, Yale [22], ORL [23], and PIE [24], are used in the experiments. We compare our proposed algorithm with PCA, LDA, LPP, LFDA, SPCA, SPP, DSNPE, and SLDA. For simplicity, we use nearest neighbor classifier for classification task and the Euclidean metric is used as the distance measure.

4.1. Experiment on the Yale Face Database. The Yale face database contains 165 grayscale images of 15 individuals. Each individual has 11 images. These images were captured under lighting conditions (left-light, center-light, and right-light), with various facial expressions (normal, happy, sad, sleepy, surprised, and wink), and with facial details (with glasses or without). The original size of the images is 243×320 pixels. In
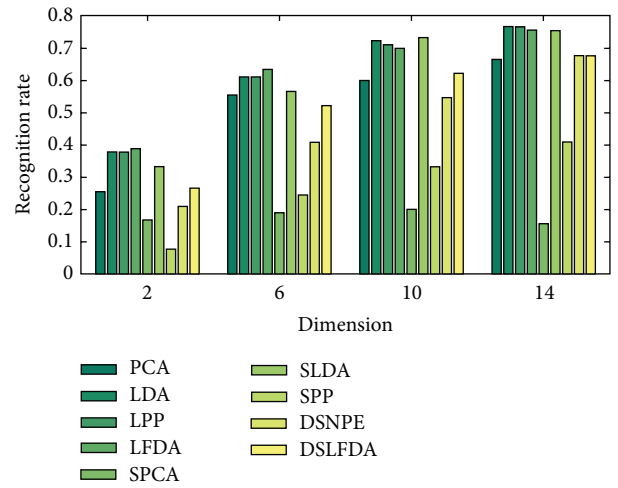

FIGURE 2: The recognition performance versus different dimensions on the Yale database.

our experiments, the face region of each original image was cropped based on the location of eyes. Each cropped image was resized to 32×32 pixels. Figure 1 shows the cropped sample images of two individuals from the Yale database.

In the first experiment, we randomly select l ($l = 2, 3, 4, 5, 6$) images per subject for training and the remaining images are for testing. 10 time runs were implemented for stable performance. The average rates are used as the final recognition accuracies. For LFDA, the parameter is set to $l - 1$ for simplicity. LPP is implemented in supervised model. For SPCA, we manually choose the sparse principal component in order to obtain the best performance. Table 1 shows the recognition accuracies of different methods with the corresponding dimension.

In the second experiment, we experiment with different dimensionalities of the projected space. Five images per individual were randomly selected for training, and the remaining images were used for testing. Figure 2 shows the performance of different methods.

4.2. Experiment on the ORL Face Database. The ORL database contains 400 images of 40 individuals. Each individual has 10 images. The images were captured at different times, under various light conditions, and with different facial

TABLE 1: The top recognition rates (%) and the corresponding dimensions on Yale database by different methods (mean ± std).

	2 trains	3 trains	4 trains	5 trains	6 trains
PCA	46.30 ± 3.26 (30)	51.58 ± 4.00 (44)	56.19 ± 4.16 (56)	57.33 ± 4.60 (74)	62.40 ± 4.06 (84)
LDA	45.11 ± 3.46 (10)	62.08 ± 4.31 (14)	70.86 ± 4.88 (14)	71.44 ± 5.19 (14)	77.22 ± 3.47 (14)
LPP	45.26 ± 3.52 (10)	62.83 ± 4.22 (14)	70.57 ± 4.59 (14)	72.22 ± 3.81 (16)	78.00 ± 3.46 (20)
LFDA	45.11 ± 3.46 (10)	62.50 ± 5.43 (14)	71.33 ± 5.07 (14)	72.33 ± 5.09 (18)	78.27 ± 3.72 (20)
SPCA	43.19 ± 3.22 (28)	49.83 ± 4.08 (44)	54.95 ± 3.70 (56)	56.78 ± 3.33 (44)	61.87 ± 5.11 (30)
SLDA	51.19 ± 5.78 (14)	63.85 ± 3.47 (14)	72.00 ± 4.76 (14)	72.56 ± 2.29 (14)	78.40 ± 2.42 (14)
SPP	46.59 ± 5.36 (16)	52.92 ± 3.63 (28)	57.67 ± 3.54 (48)	58.48 ± 3.76 (68)	64.53 ± 4.63 (72)
DSNPE	50.74 ± 5.26 (12)	63.58 ± 3.64 (30)	73.62 ± 5.04 (46)	75.89 ± 2.82 (54)	80.08 ± 2.61 (64)
DSLFDA	53.11 ± 5.36 (30)	65.17 ± 3.68 (36)	73.24 ± 5.03 (40)	74.44 ± 3.51 (42)	81.47 ± 2.77 (56)

FIGURE 3: Sample images of two individuals from ORL database.

expressions. The original size of the images is 243×320 pixels. The images were manually cropped and resized to 32 × 32 pixels. Figure 3 shows the cropped sample images of two individuals from the ORL database.

In the first experiment, we randomly select l ($l = 2, 3, 4, 5, 6$) images per subject for training and the remaining images are for testing. 10 time runs were implemented for stable performance. The average rates are used as the final recognition accuracies. The experimental parameters were set as in Section 4.1.

In the second experiment, we experiment with different dimensionalities of the projected space. Five images per individual were randomly selected for training, and the remaining images were randomly selected for testing. Figure 4 shows the performance of different methods (Table 2).

4.3. Experiment on the PIE Face Database. The CMU PIE face database contains 41368 images of 68 individuals. The images were captured under 13 different poses, under 43 different illumination conditions, and with 4 different expressions. In our experiments, we choose a subset (C29) that contains 1632 images of 68 individuals. These were manually cropped and resized to 32 × 32 pixels. Figure 5 shows the cropped sample images of two individuals from CMU PIE database.
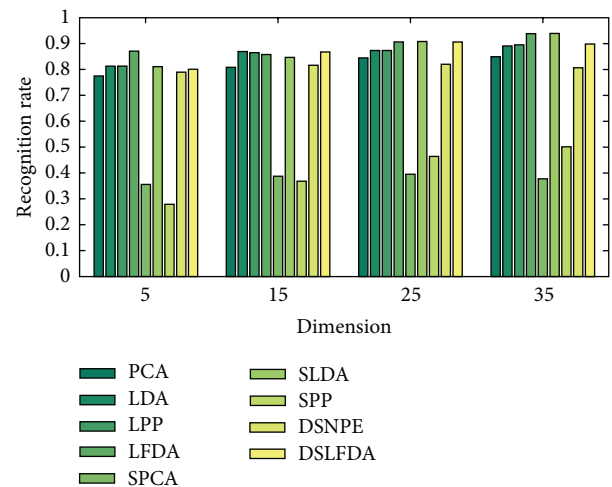

FIGURE 4: The recognition performance versus different dimensions on the ORL database.

In the first experiment, we randomly select l ($l = 3, 6, 9, 12, 15$) images per subject for training and the remaining images are for testing. 10 time runs were implemented for stable performance. The average rates are used as the final

TABLE 2: The top recognition rates (%) and the corresponding dimensions on ORL database by different methods (mean ± std).

	2 trains	3 trains	4 trains	5 trains	6 trains
PCA	71.63 ± 3.06 (79)	79.14 ± 1.91 (113)	83.87 ± 1.77 (159)	86.85 ± 2.01 (189)	89.94 ± 2.72 (203)
LDA	75.87 ± 3.42 (27)	85.96 ± 2.61 (39)	91.54 ± 1.51 (39)	93.85 ± 1.63 (39)	95.63 ± 1.89 (39)
LPP	78.78 ± 2.80 (39)	86.25 ± 2.93 (39)	90.46 ± 1.83 (39)	94.40 ± 1.71 (39)	95.94 ± 1.96 (41)
LFDA	82.13 ± 3.19 (39)	87.00 ± 2.97 (37)	90.83 ± 1.73 (37)	94.50 ± 1.90 (39)	96.06 ± 2.74 (39)
SPCA	71.16 ± 2.80 (79)	78.50 ± 2.42 (117)	84.04 ± 1.82 (131)	86.80 ± 2.28 (183)	90.56 ± 2.03 (73)
SLDA	82.03 ± 2.82 (39)	90.21 ± 1.85 (39)	93.79 ± 1.26 (39)	96.05 ± 1.83 (39)	97.56 ± 1.12 (39)
SPP	79.53 ± 2.22 (57)	83.50 ± 2.08 (63)	86.58 ± 2.10 (99)	87.40 ± 2.63 (95)	88.19 ± 3.19 (93)
DSNPE	78.97 ± 3.47 (39)	86.43 ± 1.92 (55)	92.92 ± 1.59 (73)	94.60 ± 1.96 (71)	95.94 ± 2.29 (89)
DSLFDA	81.16 ± 2.88 (69)	90.25 ± 2.51 (67)	94.38 ± 1.54 (81)	96.65 ± 1.20 (87)	97.88 ± 1.03 (95)

FIGURE 5: Sample images of two individuals from CMU PIE database.

recognition accuracies. The experimental parameters were set as in Section 4.1. Table 3 shows the recognition accuracies of different methods with the corresponding dimension.

In the second experiment, we experiment with different dimensionalities of the projected space. Fifteen images per individual were randomly selected for training, and the remaining images were randomly selected for testing. Figure 6 shows the performance of different methods.

5. Discussion and Conclusion

5.1. Discussion. Based on the above experimental results, we can conclude the following observations.

(1) For each method, the recognition rate increases with the increase of training sample sizes. The supervised extension of LPP can effectively improve the performance. PCA and SPCA achieve the worst results in all experiments; meanwhile, the performance of SPCA is inferior to that of PCA on all face databases. The reason may be that the number of nonzero variables for each component is selected equally.

(2) For LDA and SLDA, the dimensionalities of projected subspace are $C - 1$ at most; C is the number of classes. LFDA and DSLFDA can overcome this limitation; hence, we may project the original high-dimensional
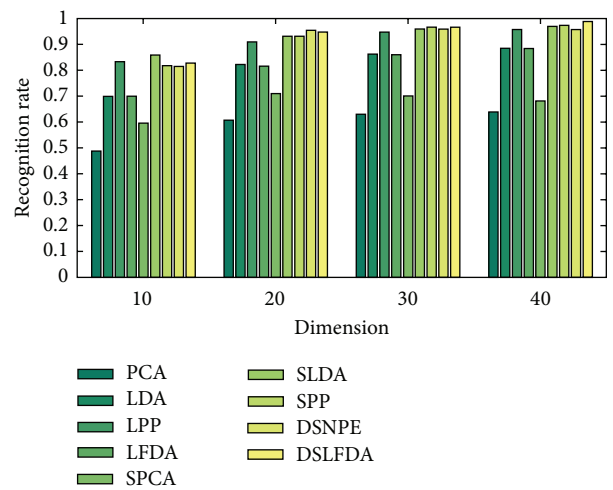

FIGURE 6: The recognition performance versus different dimensions on the CMU PIE database.

data into a low-dimensional subspace whose dimensionality is larger than the number of classes.

(3) From Table 3, LPP and SLDA outperform LFDA on the CMU PIE database. However, DSLFDA can

TABLE 3: The top recognition rates (%) and the corresponding dimensions on CMU PIE database by different methods (mean ± std).

	3 trains	6 trains	9 trains	12 trains	15 trains
PCA	36.11 ± 0.85 (155)	55.05 ± 0.94 (179)	66.81 ± 2.08 (387)	76.80 ± 1.66 (415)	83.45 ± 1.77 (431)
LDA	78.40 ± 1.43 (67)	87.70 ± 1.30 (67)	90.17 ± 1.04 (67)	91.67 ± 0.46 (67)	92.37 ± 0.56 (67)
LPP	78.72 ± 1.19 (79)	89.40 ± 0.96 (187)	91.08 ± 0.86 (163)	92.22 ± 0.52 (179)	92.79 ± 0.65 (179)
LFDA	78.68 ± 1.49 (99)	88.88 ± 1.09 (139)	90.70 ± 0.68 (143)	91.83 ± 0.51 (83)	92.43 ± 0.58 (127)
SPCA	33.25 ± 0.60 (63)	52.51 ± 1.05 (347)	65.03 ± 2.14 (499)	75.40 ± 1.97 (695)	83.32 ± 2.02 (751)
SLDA	78.17 ± 1.38 (67)	88.46 ± 1.23 (67)	92.15 ± 1.09 (67)	94.68 ± 0.36 (67)	96.36 ± 0.73 (67)
SPP	75.11 ± 0.97 (67)	87.26 ± 1.04 (87)	90.59 ± 1.35 (87)	93.42 ± 0.56 (99)	95.36 ± 0.80 (99)
DSNPE	78.93 ± 1.11 (119)	88.35 ± 1.22 (79)	91.77 ± 1.28 (83)	94.09 ± 0.60 (67)	96.05 ± 0.49 (51)
DSLFDA	78.35 ± 1.24 (147)	88.99 ± 1.20 (263)	92.31 ± 1.10 (211)	95.10 ± 0.47 (191)	96.63 ± 0.62 (251)

achieve better performance than other methods. This point shows that DSLFDA improves not only the performance of LFDA but also the performance of sparse-based method, such as SLDA. The proposed DSLFDA algorithm constructs the graph on the original data and obtains the nonnegative similarity measurement. This is different from SPP and DSNPE.

(4) From the experimental results, we obtain that SPP can get competitive performance on CMU PIE database, rather than ORL and Yale databases. The reason may be that the sparse representation needs abundant training samples. Conversely, the nonnegative similarity measurement in DSLFDA is adaptive and can overcome the drawback of sparse representation.

(5) DSNPE can be regarded as an extension of SPP. It can extract the discriminant information and perform better than SPP. On the Yale database, DSNPE can achieve the best recognition performance when the training samples per individual are four and five.

5.2. Conclusion. In this paper, we proposed a sparse projection method, called DSLFDA, for face recognition. It defines a novel affinity matrix that describes the relationships of points on the original high-dimensional data. The sparse projection vectors are obtained by solving the ℓ^1-optimization problem. Experiments on Yale, ORL, and CMU PIE face databases indicate that DSLFDA can get competitive performance compared to other dimensionality reduction methods.

We only focus on supervised learning in this paper. Because a large amount of unlabeled data is available in practical applications, semisupervised learning has attracted much attention in recent years [25–27]. One of our future works is to extend our approach under the semisupervised learning framework. On the other hand, DSLFDA needs the local within-class scatter matrix positive definite. We add an identity matrix to the local within-class scatter matrix for regularization. This may motivate us to find the regularization method to approximate the local within-class scatter matrix well.

Conflict of Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

Acknowledgments

References

[1] K. Fukunaga, *Introduction to Statistical Pattern Recognition*, Computer Science and Scientific Computing, Academic Press, Boston, Mass, USA, 2nd edition, 1990.

[2] D. Tao, X. Tang, X. Li, and X. Wu, "Asymmetric bagging and random subspace for support vector machines-based relevance feedback in image retrieval," *IEEE transactions on pattern analysis and machine intelligence*, vol. 28, no. 7, pp. 1088–1099, 2006.

[3] P. N. Belhumeur, J. P. Hespanha, and D. J. Kriegman, "Eigenfaces vs. fisherfaces: recognition using class specific linear projection," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 19, no. 7, pp. 711–720, 1997.

[4] M. Turk and A. Pentland, "Eigenfaces for recognition," *Journal of Cognitive Neuroscience*, vol. 3, no. 1, pp. 71–86, 1991.

[5] L. F. Chen, H. Y. M. Liao, M. T. Ko, J. C. Lin, and G. J. Yu, "New LDA-based face recognition system which can solve the small sample size problem," *Pattern Recognition*, vol. 33, no. 10, pp. 1713–1726, 2000.

[6] R. Huang, Q. Liu, H. Lu, and S. Ma, "Solving the small sample size problem of LDA," in *Proceedings of the 16th International Conference on Pattern Recognition*, vol. 3, pp. 29–32, IEEE, Quebec City, Canada, September 2002.

[7] A. K. Qin, P. N. Suganthan, and M. Loog, "Generalized null space uncorrelated Fisher discriminant analysis for linear dimensionality reduction," *Pattern Recognition*, vol. 39, no. 9, pp. 1805–1808, 2006.

[8] H. Yu and J. Yang, "A direct LDA algorithm for high-dimensional data: with application to face recognition," *Pattern Recognition*, vol. 34, no. 10, pp. 2067–2070, 2001.

[9] T. Zhang, B. Fang, Y. Y. Tang, Z. Shang, and B. Xu, "Generalized discriminant analysis: a matrix exponential approach," *IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics*, vol. 40, no. 1, pp. 186–197, 2010.

[10] X. He, S. Yan, Y. Hu, P. Niyogi, and H.-J. Zhang, "Face recognition using Laplacianfaces," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 27, no. 3, pp. 328–340, 2005.

[11] W. Yu, X. Teng, and C. Liu, "Face recognition using discriminant locality preserving projections," *Image and Vision Computing*, vol. 24, no. 3, pp. 239–248, 2006.

[12] J. Chen, J. Ye, and Q. Li, "Integrating global and local structures: a least squares framework for dimensionality reduction," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR '07)*, pp. 1–8, Minneapolis, Minn, USA, June 2007.

[13] M. Sugiyama, "Dimensionality reduction of multimodal labeled data by local fisher discriminant analysis," *Journal of Machine Learning Research*, vol. 8, pp. 1027–1061, 2007.

[14] H. Zou, T. Hastie, and R. Tibshirani, "Sparse principal component analysis," *Journal of Computational and Graphical Statistics*, vol. 15, no. 2, pp. 265–286, 2006.

[15] L. Clemmensen, T. Hastie, D. Witten, and B. Ersboll, "Sparse discriminant analysis," *Technometrics*, vol. 53, no. 4, pp. 406–413, 2011.

[16] Z. Qiao, L. Zhou, and J. Z. Huang, "Sparse linear discriminant analysis with applications to high dimensional low sample size data," *IAENG International Journal of Applied Mathematics*, vol. 9, no. 1, pp. 48–60, 2009.

[17] S. Yan, D. Xu, B. Zhang, H.-J. Zhang, Q. Yang, and S. Lin, "Graph embedding and extensions: a general framework for dimensionality reduction," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 29, no. 1, pp. 40–50, 2007.

[18] B. Cheng, J. Yang, S. Yan, Y. Fu, and T. S. Huang, "Learning with l1-graph for image analysis," *IEEE Transactions on Image Processing*, vol. 19, no. 4, pp. 858–866, 2010.

[19] L. Qiao, S. Chen, and X. Tan, "Sparsity preserving projections with applications to face recognition," *Pattern Recognition*, vol. 43, no. 1, pp. 331–341, 2010.

[20] J. Gui, Z. Sun, W. Jia, R. Hu, Y. Lei, and S. Ji, "Discriminant sparse neighborhood preserving embedding for face recognition," *Pattern Recognition*, vol. 45, no. 8, pp. 2884–2893, 2012.

[21] B. Efron, T. Hastie, I. Johnstone, and R. Tibshirani, "Least angle regression," *The Annals of Statistics*, vol. 32, no. 2, pp. 407–499, 2004.

[22] F. S. Samaria and A. C. Harter, "Parameterisation of a stochastic model for human face identification," in *Proceedings of the 2nd IEEE Workshop on Applications of Computer Vision*, pp. 138–142, December 1994.

[23] P. N. Bellhumer, J. Hespanha, and D. Kriegman, "Eigenfaces vs. fisherfaces: recognition using class specific linear projection," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 19, no. 7, pp. 711–720, 1997.

[24] T. Sim, S. Baker, and M. Bsat, "The CMU pose, illumination, and expression database," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 25, no. 12, pp. 1615–1618, 2003.

[25] O. Chapelle, B. Schölkopf, and A. Zien, *Semi-Supervised Learning*, The MIT Press, Cambridge, Mass, USA, 2006.

[26] F. Wang and C. Zhang, "Label propagation through linear neighborhoods," *IEEE Transactions on Knowledge and Data Engineering*, vol. 20, no. 1, pp. 55–67, 2008.

[27] F. Shang, L. C. Jiao, and F. Wang, "Semi-supervised learning with mixed knowledge information," in *Proceedings of the 18th ACM International Conference on Knowledge Discovery and Data Mining (KDD '12)*, pp. 732–740, ACM, Beijing, China, August 2012.

Advances in
Operations Research

Advances in
Decision Sciences

Journal of
Applied Mathematics

Algebra

Journal of
Probability and Statistics

The Scientific
World Journal

International Journal of
Differential Equations

International Journal of
Combinatorics

Advances in
Mathematical Physics

Hindawi

Submit your manuscripts at
http://www.hindawi.com

Journal of
Complex Analysis

Journal of
Mathematics

Mathematical Problems
in Engineering

Abstract and
Applied Analysis

Discrete Dynamics in
Nature and Society

International
Journal of
Mathematics and
Mathematical
Sciences

Journal of
Discrete Mathematics

Journal of
Function Spaces

International Journal of
Stochastic Analysis

Journal of
Optimization