Journal of Electronic Imaging

JElectronicImaging.org

Locality sensitive discriminant projection for feature extraction and face recognition

Yi-Kang Wei
Cong Jin

SPIE. IS&T imaging.org

Yi-Kang Wei, Cong Jin, "Locality sensitive discriminant projection for feature extraction and face recognition," *J. Electron. Imaging* **28**(4), 043028 (2019), doi: 10.1117/1.JEI.28.4.043028.

Locality sensitive discriminant projection for feature extraction and face recognition

Yi-Kang Wei and Cong Jin*
Central China Normal University, School of Computers, Wuhan, China

Abstract. As an effective feature extraction method, locality sensitive discriminant analysis (LSDA) utilizes the neighbor relationship of data to characterize the manifold structure of data and uses label information of data to adapt to classification tasks. However, the performance of LSDA is affected by outliers and the destruction of local structure. Aiming at solving the limitations of LSDA, a locality sensitive discriminant projection (LSDP) algorithm is proposed. LSDP minimizes the distance of intraclass neighbor samples to maintain local structure and minimizes the intraclass non-neighbor samples to increase the compactness of intraclass samples after projection. The problem of outliers is alleviated by increasing the compactness of intraclass samples in subspace. At the same time, we redefine the weights of interclass neighbor samples to maintain the neighbor relationship of different labels samples. Holding the local structure of interclass samples maintains the manifold structure of data. Experiments on face datasets demonstrate the effectiveness of the LSDP algorithm. [DOI: 10.1117/1.JEI.28.4.043028]

Keywords: feature extraction; locality sensitive discriminant projection; outliers; manifold learning.

Paper 190196 received Feb. 27, 2019; accepted for publication Aug. 7, 2019; published online Aug. 24, 2019.

1 Introduction

Dimensionality reduction is a common data analysis and processing method in computer vision, pattern recognition, machine learning, and other fields.[1–3] In face recognition,[4,5] data visualization, and other fields,[6–9] it is usually necessary to extract effective low-dimensional features from high-dimensional data to facilitate data analysis and processing. Feature extraction algorithms for dimensionality reduction can not only reduce computer's storage space and computing cost but also remove redundant information of data and extract essential features of data.[10,11] Using the features extracted by dimensionality reduction algorithm achieves better results in classification tasks.

In practical application, the target of dimensionality reduction algorithm is different in different fields. The goal of the dimensionality reduction algorithm in data visualization is to reduce the dimensions of data in the original space to two- (2-D) to three- (3-D) dimensions to show the distribution of data. The goal of the dimensionality reduction algorithm used for the classification task is to project the data in the original space into the low-dimensional subspace for the classification task.[12–15] There are many feature extraction algorithms proposed in recent years.[16] The matrix exponential-based discriminant locality preserving projections (LPP) algorithm was proposed in Ref. 17, to solve the small sample size (SSS) problem by an efficient method. A nonlinear supervised dimensionality reduction was proposed in Ref. 18, which has achieved generalization by constructing a smooth and regular interpolation function. A new linear discriminant analysis (LDA) algorithm was proposed in Ref. 19, which was based on $L1$-norm maximization and LPP. An unsupervised robust discriminative manifold embedding algorithm was proposed in Ref. 20, which aims to solve the problem caused by noise data. Although these algorithms have been well utilized and have achieved good results, they do not address the issues of concern in this paper. We focus on the dimensionality reduction algorithm in the field of face recognition as follows.

Principle component analysis (PCA)[21,22] and LDA[23] are widely used for dimensionality reduction. PCA is an unsupervised algorithm, which projects the data along the direction of maximal variance. PCA can retain the diversity of data and has been successfully applied to dimensionality reduction. In face recognition field, PCA is also called eigenface. However, PCA does not use the label information of data, which leads to the failure of PCA to achieve better results in classification tasks. In contrast to the unsupervised algorithm PCA, LDA uses the label information of data to construct intraclass and interclass scatter matrixes to make the intraclass samples closer after projecting and the interclass samples further apart after projected. LDA makes the data in subspace after projected more suitable for classification tasks. Recently, a large number of algorithms[24–26] have been proposed on the basis of LDA. When the sample's dimension is higher than the number of samples, the intraclass scatter matrix of LDA is singular, which leads to the SSS problem.[23] In order to solve the SSS problem of LDA, maximum margin criterion[25] has been put forward. A general approach to solve the SSS problem is using PCA to preprocess the data to ensure the nonsingularity of the matrix before using LDA, which is called fisherface in face recognition field. Both PCA and LDA use the global geometric structure of data to obtain the projection direction.

Research shows that face data have a low-dimensional manifold structure. The manifold learning algorithm that uses the local geometric structure of the data to maintain manifold structure has achieved remarkable results in face recognition. The manifold learning algorithm can be divided into two types of linear algorithm and nonlinear algorithm. Classical nonlinear manifold learning algorithms include

*Address all correspondence to Cong Jin, E-mail: jincong@mail.ccnu.edu.cn

1017-9909/2019/$28.00 © 2019 SPIE and IS&T

isomap,[27] locally linear embedding (LLE),[28] and Laplacian eigenmap (LE).[29] These nonlinear algorithms use the local relationship to keep the manifold structure of data. However, nonlinear manifold learning algorithms are affected by out-of-sample problem.[30] These nonlinear algorithms get the low-dimensional features from all the samples, so the mapping between high-dimensional data and low-dimensional data is latent, which leads to the new sample's low-dimensional representation cannot be directly obtained.

Some works have focused on the linear dimension reduction algorithms to solve the out-of-sample problem of nonlinear algorithms. Neighborhood preserving embedding[31] was put forward to improve LLE. LPP[32] was put forward for avoiding the limitations of LE. LPP is an unsupervised algorithm to maintain the local structure of data. A large number of algorithms[33–36] have been proposed based on LPP. In order to enhance the discriminant ability of LPP, some algorithms, such as locality sensitive discriminant analysis (LSDA)[36] and locality discriminant projection,[37] were proposed by using label information of samples. LSDA algorithm has achieved good results in face recognition task, and a large number of algorithms have been proposed to improve the LSDA algorithm. Among them, stable locality sensitive discriminant analysis[38] may improve LSDA from the perspective of maintaining similarity and diversity of samples. There are other related studies.[39,40] For example, improved LSDA (ILSDA)[41] combines the local and global geometric structure of data to solve the outlier problem of LSDA. The ILSDA algorithm integrates the intraclass scatter to LSDA to make the intraclass samples closer to the intraclass center after projected, which impairs the local structure of intraclass samples. Discriminant maximum margin projection (DMMP) aims at preserving the local neighborhood information as well as maximizing the margin of interclass neighbor samples.[42] Yan et al. proposed marginal Fisher analysis (MFA),[43] which used the local structure of data to independently construct the intrinsic graph and penalty graph, the neighbor samples of the same class are closer to each other after projected, and the neighbor samples of different class are far away from each other after projected. MFA provides a general framework for manifold learning dimension reduction algorithm based on graph embedding method. In contrast to MFA, LSDA only constructs the neighbor graph of samples and then divides the neighbor graph of samples into intraclass neighbor graph and interclass neighbor graph according to the label information of samples. It is very easy to meet a situation where the class of a sample is different from the classes of its k-nearest neighbor samples for LSDA, which leads the intraclass outliers problem.[41]

LSDA is sensitive to intraclass outliers because the algorithm only uses the local structure of data. The intraclass outlier problem of LSDA can be solved by expanding the values of k-nearest neighbor samples. If the value of k-nearest neighbor samples of a sample is set as the number of all samples, then each sample is connected by intraclass samples, and the intraclass outlier problem will not exist. The influence of outliers can also be relieved by combining the global structure of data.[41] At the same time, while maximizing the distance of interclass neighbor samples enhances the discriminant ability of algorithm, it destroys the local structure of samples. The interclass neighbor samples in original space are far away from each other after projected, which changes the potential pattern information between samples and is not beneficial to maintain the manifold structure of data.

In order to solve the intraclass outlier problem of LSDA algorithm and maintain the local structure of interclass samples, a new algorithm named locality sensitive discriminant projection (LSDP) is proposed in this paper. In order to make the intraclass samples after projection more compact, LSDP not only minimizes the distance of intraclass neighbor samples after projection but also minimizes the distance of intraclass but non-neighbor samples after projection to avoid the outlier problem. We redefine the weights between samples according to the Euclidean distance between samples and label information. The weights between intraclass neighbor samples larger than the weights between intraclass non-neighbor samples, which make intraclass neighbor samples dominate the optimization so as to maintain the local structure of intraclass samples. At the same time, the intraclass non-neighbor samples are added in optimization to make intraclass samples have stronger compactness after projection so as to solve the problem of intraclass outliers. Compared with LSDA and ILSDA, LSDP algorithm not only solves the intraclass outlier problem of LSDA but also keeps the intraclass samples' relation on the original space. Compared with ILSDA that uses the intraclass scatter to solve outlier problem, LSDP uses the different weights of intraclass neighbor samples and intraclass non-neighbor samples to reflect the intraclass local structure to maintain the manifold of samples. In order to maintain the local structure of interclass neighbor samples, the weights of interclass samples are redefined so that the interclass neighbor samples that are close to each other in the original space keep close after projected, and those that are far away in the original space remains faraway after projection. For LSDA, ILSDA and MFA, compared with maximum the distances of interclass neighbor samples after projecting, LSDP algorithm keeps the interclass local structure on the original subspace. Then, we integrated the intraclass representation and interclass representation in an equation. The analysis of algorithm and the results of experiments show the effectiveness of LSDP.

The main contributions of this work are as follows.

1. LSDP takes into account not only the local neighbor relations of intraclass samples but also the nonlocal neighbor relations of intraclass samples to solve the classification difficulties caused by intraclass outliers.
2. A new method for calculating the weight of intraclass samples is proposed so that the weight of intraclass neighbor samples is greater than the weight of intraclass non-neighbor samples to maintain the local structure of the intraclass neighbor samples, enhance the tightness between the intraclass non-neighbor samples and other intraclass samples, and reduce the impact of outliers on the classification performance.
3. A new method for calculating the weight of the interclass neighbor samples is proposed, which may guarantee that the weight of the interclass neighbor samples increases with the Euclidean distance between them to maintain the local structure of the interclass neighbor samples after projection. Maintaining this local structure can keep the underlying patterns of samples in subspace and improve the classification performance.

The remainder of this paper is organized as follows. In Sec. 2, related works are introduced. In Sec. 3, the LSDP algorithm is proposed. In Sec. 4, experiments are conducted and analyzed. In Sec. 5, conclusions and future work are provided.

2 Related Work

Suppose we have n samples, each sample is a d-dimensional column vector, x_i represents the i'th samples, $x_i \in R^d$. The matrix $X = \{x_1, x_2, \ldots, x_i, \ldots, x_n\} \in R^{d\times n}$ represents the samples set, each column of X represents a sample. The matrix $V = \{v_1, v_2, \ldots, v_r\} \in R^{d\times r}$ is the transformation matrix, which transforms sample x_i to the low-dimensional representation y_i by $y_i = V^T x_i$, $y_i \in R^r$, $r \ll d$. The label of x_i is $l(x_i) \in \{1, 2, \ldots, C\}$, which indicates the class of x_i. There are C classes of samples, the number of i'th class is n_i, and all the samples' number can be represented as $n = \sum_{i=1}^{C} n_i$.

$N_k(x_i) = \{x_i^1, x_i^2, \ldots, x_i^k\}$ is defined to represent the set of k samples closest to x_i in all samples. Euclidean distance is used for distance measurement.

2.1 *Locality Preserving Projection*

LPP uses the local structure of samples to construct neighbor graph G. If x_i is in the k-nearest neighbor of x_j or x_j in the k-nearest neighbor of x_i, there has an edge between x_i and x_j. The weight matrix W characterizes similarity between samples, if there has an edge between x_i and x_j, the value of W_{ij} is taken 1. The value of W is defined as

$$W_{ij} = \begin{cases} 1, & x_i \in N_k(x_j) \text{ or } x_j \in N_k(x_i) \\ 0, & \text{otherwise} \end{cases}. \tag{1}$$

LPP maintains the sample's neighbor relationship to maintain the manifold structure of data. The objective function of LPP is defined by minimizing the distance between neighbor samples after projection:

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

After linear transformation, Eq. (2) can be rewritten as

$$\begin{aligned} &\frac{1}{2} \min \sum_{ij} \|y_i - y_j\|^2 W_{ij} \\ &= \frac{1}{2} \min \sum_{ij} (V^T x_i - V^T x_j)^T (V^T x_i - V^T x_j) W_{ij} \\ &= \min \operatorname{trace}[V^T X (D - W) X^T V]. \end{aligned} \tag{3}$$

In Eq. (3), "trace" denotes the trace of the matrix. $D = \operatorname{diag}(D_{ii})$ is a diagonal matrix whose main diagonal elements are D_{ii}. $D_{ii} = \sum_j W_{ij}$ is the i'th row (or column, because W is symmetric matrix) sums of weight matrix W.

Laplacian matrix is defined as $L = D - W$. Matrix D provides a natural measure on the data points, the bigger D_{ii} is, the more "important" is y_i.[32] We impose $V^T X D X^T V = I$ as a constraint. LPP minimizes the distance of neighbor samples after projection, which can be represented as

$$\arg\min \operatorname{trace}(V^T X L X^T V), \text{ s.t. } V^T X D X^T V = I. \tag{4}$$

According to Eqs. (1) and (4), if $x_i \in N_k(x_j)$ or $x_j \in N_k(x_i)$, then $W_{ij} = 1$. LPP seeks for the projection direction that can make the nearest neighbor sample closer after projection. If x_i and x_j are not in each other's k-nearest sample set, $W_{ij} = 0$, LPP does not consider these points.

LPP uses the neighbor relationship of samples to maintain the intrinsic manifold structure of the data. However, LPP ignores the influence of non-neighbor samples and does not use the label information of samples, which will weaken the effect of LPP in classification tasks.

2.2 *Locality Sensitive Discriminant Analysis*

LSDA improves LPP by using the label information of samples. Similar to LPP, LSDA selects each sample's k-nearest sample set according to the Euclidean distance between samples. For a general sample x_i, $N_k(x_i)$ is the nearest sample set of x_i. Then, according to the label information of samples, $N_k(x_i)$ can be divided into $N_k^w(x_i)$ and $N_k^b(x_i)$. $N_k^w(x_i)$ is the intraclass neighbor set of x_i, and $N_k^b(x_i)$ is the interclass neighbor set of x_i. The samples in $N_k^w(x_i)$ have same label with x_i and belong to $N_k(x_i)$, samples in $N_k^b(x_i)$ have different labels with x_i and belong to $N_k(x_i)$. LSDA divides nearest neighbor graph G in LPP into intraclass graph (G^w) and interclass graph (G^b) according to label information. There has an edge between x_i and x_j if $x_i \in N_k^w(x_j)$ or $x_j \in N_k^w(x_i)$ in intraclass graph G^w. There has an edge between x_i and x_j if $x_i \in N_k^b(x_j)$ or $x_j \in N_k^b(x_i)$ in interclass graph G^b. The objective of LSDP is to minimize the distance of intraclass neighbor samples and maximize the distance interclass neighbor samples, which can be expressed as

$$\min \sum_{ij} (y_i - y_j)^2 W_{ij}^w, \tag{5}$$

$$\max \sum_{ij} (y_i - y_j)^2 W_{ij}^b. \tag{6}$$

The weights between intraclass neighbor samples (W_{ij}^w) and interclass neighbor samples (W_{ij}^b) are defined as

$$W_{ij}^w = \begin{cases} 1, & x_i \in N_k^w(x_j) \text{ or } x_j \in N_k^w(x_i) \\ 0, & \text{otherwise} \end{cases}, \tag{7}$$

$$W_{ij}^b = \begin{cases} 1, & x_i \in N_k^b(x_j) \text{ or } x_j \in N_k^b(x_i) \\ 0, & \text{otherwise} \end{cases}. \tag{8}$$

$D^w = \operatorname{diag}(D_{ii}^w)$ is a diagonal matrix whose main diagonal elements are D_{ii}^w, $D_{ii}^w = \sum_j W_{ij}^w$. $D^b = \operatorname{diag}(D_{ii}^w)$ is a diagonal matrix whose main diagonal elements are D_{ii}^b, $D_{ii}^b = \sum_j W_{ij}^b$. The Laplacian matrix $L_b = D^b - W^b$. The matrix D^w provides a natural measure on the samples, we impose a constraint $V^T X D^w X^T V = I$.[36] After transformation of Eqs. (5) and (6), the objection of LSDP can be rewritten as Eq. (9). α is a parameter and $0 \le \alpha \le 1$

$$\begin{aligned} &\arg\max \operatorname{trace}\{V^T X[\alpha L_b + (1-\alpha) W^w] X^T V\}, \\ &\text{s.t. } V^T X D^w X^T V = I. \end{aligned} \tag{9}$$

LSDA minimizes the distance of intraclass neighbor samples to make the intraclass neighbor samples closer in subspace

and maximizes the distance of interclass neighbor samples to make the interclass neighbor samples far away from each other in subspace to enhance the discriminant ability. The intraclass neighbor graph and interclass neighbor graph are based on neighbor graph. If the samples' labels in k-nearest neighbor set $N_k(x_i)$ are all different with x_i, the intraclass neighbor set of x_i will be empty. In this case, x_i is an outlier point. On the optimization of LSDA, these outliers cannot be projected closer with the same class samples, which will affect the classification effect. Second, maximizing the distance of interclass neighbor samples will destroy the local neighbor structure of interclass neighbor samples. The neighbor samples of different classes in the original space will be farther away from each other in subspace, which violates the assumption of manifold learning to keep the local structure of data.

3 Locality Sensitive Discriminant Projection

3.1 *Intraclass Representation*

Inspired by the LSDA, LSDP is proposed to solve the outlier problem of LSDA to improve the performance of the algorithm in classification tasks. Meanwhile, the local structure of the data was maintained so that the data in subspace could reflect the essential characteristics of the data to maintain the manifold structure.

First, the outlier problem of LSDA algorithm is analyzed deeply. As shown in Fig. 1(a), there are two classes and each class has four samples of 2-D. LSDA algorithm first calculates the k-nearest neighbor samples of each sample, then the k-nearest neighbor samples set of each sample is divided into intraclass neighbor samples set and the interclass neighbor samples set according to the label information of samples. According to the situation shown in Fig. 1(a), if the value k of nearest neighbor samples is chosen 3, then the three neighbor samples of outliers in class 1 are the category of class 2, and the three neighbor samples of outliers in class 2 are the category of class 1. According to the solution method of LSDA algorithm, the 2-D samples in Fig. 1(a) are reduced to one-dimension (1-D), as shown in Fig. 1(b). The two outliers in Fig. 1(a) cannot maintain a close distance with the intraclass samples after projected because they have no edge connecting with the intraclass samples, which weakens the performance of LSDA in classification tasks.

The problem of intraclass outliers can be solved by expanding the number of nearest neighbor samples to relieve. If nearest neighbor parameter k value is $n-1$, where n is the number of all samples, the intraclass neighbor set will not be empty and intraclass outliers shown in Fig. 1(a) will not exist. But the method does not use the local structure of data effectively.

According to the analysis in Figs. 1(a) and 1(b), another method to solve the problem of intraclass outliers is to use the label information of data to make the intraclass samples closer to each other after projection, so as to avoid outliers. Intraclass outliers have no connection with the intraclass samples, which leads to the problem of outliers far away from the intraclass sample after projection. Therefore, LSDP we proposed takes not only intraclass neighbor samples but also intraclass non-neighbor samples into account. The intraclass neighbor samples are used to maintain the manifold structure, which is consistent with the role of intraclass neighbor samples in LSDA algorithm. The intraclass non-neighbor samples are used to increase the relationship between intraclass samples so that the intraclass samples in subspace have stronger compactness to avoid outlier problems. As shown in Fig. 1(c), by increasing the compactness of intraclass samples after projection, outliers are closer with the samples of the same class.

The weights of samples reflect the proportion in optimization. We redefine the weights between intraclass neighbor samples and intraclass non-neighbor samples as

$$W_{ij}^{w}=\begin{cases}K(x_i,x_j)[1+K(x_i,x_j)], & x_i\in N_k^w(x_j) \text{ or } x_j\in N_k^w(x_i)\\ K(x_i,x_j)[1-K(x_i,x_j)], & l(x_i)=l(x_j) \text{ and } x_i\notin N_k^w(x_j) \text{ and } x_j\notin N_k^w(x_i)\\ 0, & \text{otherwise}\end{cases}. \tag{10}$$

$K(x_i,x_j)=\exp(-\|x_i-x_j\|^2/t)$, $t>0$ is the heat kernel parameter that can change the value of $K(x_i,x_j)$. $\|x_i-x_j\|^2$ is the L_2-norm of (x_i-x_j), and $N_k^w(x_i)$ is the intraclass neighbor set of x_i, which has been defined in Sec. 2.

Then, we minimize the distance of intraclass samples as

$$\min\sum_{ij}\|y_i-y_j\|^2W_{ij}^{w}. \tag{11}$$

In Eq. (10), $K(x_i,x_j)[1+K(x_i,x_j)]$ is the intraclass neighbor samples' weight and $K(x_i,x_j)[1-K(x_i,x_j)]$ is the intraclass non-neighbor samples' weight. Since $1\le[1+\exp(-\|x_i-x_j\|^2/t)]\le 2$ and $0\le[1-\exp(-\|x_i-x_j\|^2/t)]\le 1$, so the weight of intraclass neighbor samples and intraclass non-neighbor samples can be controlled in certain ranges.[37] As shown in Eq. (11), intraclass neighbor samples are closer after projected than the intraclass non-neighbor

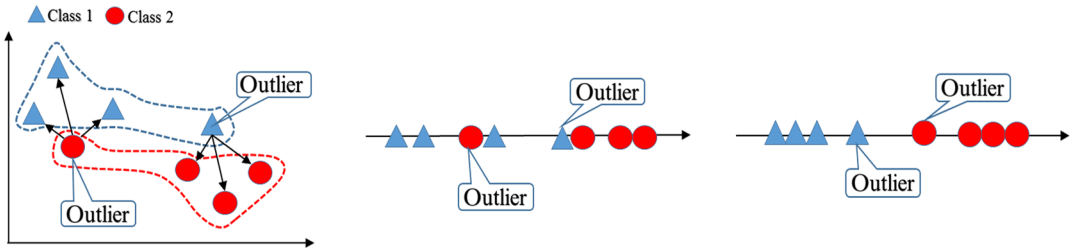

Fig. 1 (a) A 2-D toy dataset with outliers. (b) Low-dimensional (1-D) features extracted by LSDA. (c) Low-dimensional (1-D) features extracted by LSDP.

samples because for a constant distance between samples the weight of intraclass neighbor samples is larger than the weight of intraclass non-neighbor samples. The local structure of intraclass neighbor samples can be maintained, and the intraclass non-neighbor samples are used to improve the compactness of intraclass samples after projection.

3.2 *Interclass Representation*

The local structure of interclass samples preserves discriminant information and potential manifold information. So maintaining the local structure of interclass samples is also important. Figure 2(a) shows three different classes of 2-D samples, class 1, class 2, and class 3. If we only consider the discriminant information of interclass neighbor samples to make the interclass neighbor points far away in subspace, as shown in Fig. 2(b), we will destroy the neighbor relationship of interclass samples. As shown in Fig. 2(c), we want to maintain the local structure of points after projection, keeping the points close in original space closer in subspace. In Fig. 2(c), the class 1 sample that is close to the class 2 sample in original space is still close to the class 2 sample in subspace, and the class 1 sample that is far from the class 3 sample in the original space is still far from the class 3 sample in subspace.

In order to maintain the local relationship of interclass neighbor samples, we define the weight of interclass samples as

$$W_{ij}^{b}=\begin{cases}1-K(x_i,x_j), & x_i\in N_k^b(x_j) \text{ or } x_j\in N_k^b(x_i)\\ 0, & \text{otherwise}\end{cases}. \tag{12}$$

We maximize the distance of interclass neighbor samples after projection as

$$\max\sum_{ij}\|y_i-y_j\|^2W_{ij}^{b}. \tag{13}$$

In Eq. (12), $[1-K(x_i,x_j)]$ is a monotone increasing function about $\|x_i-x_j\|^2$. If two interclass neighbor samples in the original space have great Euclidean distance, the weight between two samples is large. If the two samples in the original space are closer to each other, the weight of them is small. By maximizing Eq. (13), the interclass neighbor samples that are far away in original space tend to have large distance in subspace.

3.3 *Optimal Linear Embedding*

In order to facilitate the solution, the above objective function [Eq. (11)] that minimizes the projected distance of similar samples and the objective function [Eq. (13)] that maximizes the projected distance of different samples are integrated together in this section.

First, after linear transformation $y_i=V^Tx_i$, the objective function [Eq. (11)] can be rewritten as

$$\begin{aligned}&\min\sum_{ij}\|y_i-y_j\|^2W_{ij}^{w}=2\min\sum_{ij}\|V^Tx_i-V^Tx_j\|^2W_{ij}^{w}\\&=2\min\sum_{ij}(V^Tx_i-V^Tx_j)^T(V^Tx_i-V^Tx_j)W_{ij}^{w}\\&=2\min\operatorname{trace}\left\{\left(\sum_i V^Tx_iD_{ii}^{w}x_i^TV-\sum_{ij}V^Tx_iW_{ij}^{w}x_j^TV\right)\right\}\\&=2\min\operatorname{trace}\{V^TX(D^w-W^w)X^TV\}.\end{aligned} \tag{14}$$

At the same time, the objective function [Eq. (13)] can be rewritten as Eq. (15) after linear transformation

$$\begin{aligned}&\min\sum_{ij}\|y_i-y_j\|^2W_{ij}^{b}\\&=2\max\sum_{ij}\|V^Tx_i-V^Tx_j\|^2W_{ij}^{b}\\&=2\max\sum_{ij}(V^Tx_i-V^Tx_j)^T(V^Tx_i-V^Tx_j)W_{ij}^{b}\\&=2\max\operatorname{trace}\left\{\left(\sum_i V^Tx_iD_{ii}^{b}x_i^TV-\sum_{ij}V^Tx_iW_{ij}^{b}x_j^TV\right)\right\}\\&=2\max\operatorname{trace}\{V^TX(D^b-W^b)X^TV\}\\&=2\max\operatorname{trace}(V^TXL^bX^TV).\end{aligned} \tag{15}$$

D^w, D^b, and L^b in Eqs. (14) and (15) are the same with LSDA and are defined in Sec. 2.2. The matrix D^w is a natural measure of samples. If D_{ii}^{w} is large, then it implies that the class containing x_i has a high density around x_i. Therefore, the bigger the value of D_{ii}^{w} is, the more "important" is x_i.[36] We impose a constraint as follows:

$$V^TXD^wX^TV=I. \tag{16}$$

Then make Eq. (16) as constraint conditions, object function [Eq. (14)] can be rewritten as

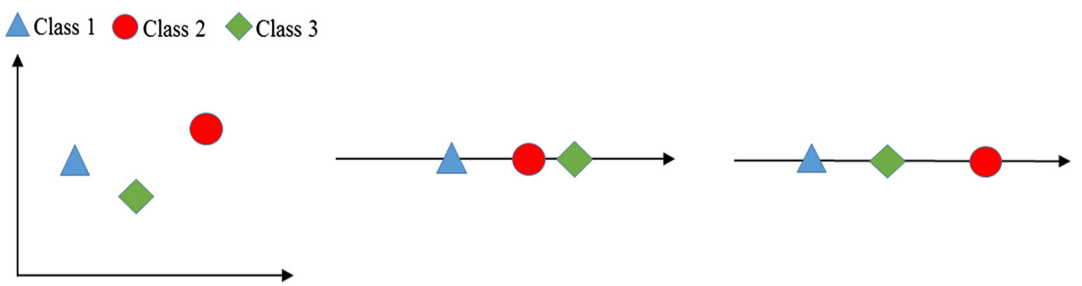

Fig. 2 (a) A 2-D toy dataset with three different class. (b) Low-dimensional (1-D) features extracted by only enhancing the discriminant ability. (c) Low-dimensional (1-D) features extracted by considering the local structure.

$$\begin{aligned}&\min \operatorname{trace}\{V^T X(D^w - W^w)X^T V\}\\&= \min \operatorname{trace}(I - V^T X W^w X^T V)\\&\Rightarrow \max \operatorname{trace}(V^T X W^w X^T V).\end{aligned} \tag{17}$$

Finally, we integrate the objective functions [Eqs. (15) and (17)], and the optimization problem of LSDP reduces to finding

$$\begin{aligned}&\max \operatorname{trace}\{V^T X[\alpha L^b + (1-\alpha)W^w]X^T V\},\\&\text{s.t. } V^T X D^w X^T V = I.\end{aligned} \tag{18}$$

α is a parameter and $0 \le \alpha \le 1$. We can change the value of α to enhance the adaptive ability of LSDP.

We use Lagrange multiplier method to obtain the optimal transformation matrix V

$$\frac{\partial}{\partial V}\{V^T X[\alpha L^b + (1-\alpha)W^w]X^T V - \lambda V^T X D^w X^T V\} = 0. \tag{19}$$

Equation (19) can be converted the eigensolver problem as

$$X[\alpha L^b + (1-\alpha)W^w]X^T V = \lambda X D^w X^T V. \tag{20}$$

Find the column vectors $v_1, v_2, \ldots, v_r$ to be the solutions of Eq. (20), ordered according to their eigenvalues, $\lambda_1 > \lambda_2 > \ldots > \lambda_r$ $(r \ll d)$. $V = \{v_1, v_2, \ldots, v_r\} \in R^{d\times r}$ is the optimal tansformation matrix.

We give the algorithm description of LSDP in Algorithm 1.

In the end, we use the O to analyze the computational complexity of LSDP. In step 1 of LSDP, we should calculate two samples' distance and choose each sample's k-nearest samples by sorting the distance between them. First, we should calculate the distance of d-dimensional samples with other n samples, which computational complexity is $O(n^2 d)$. Then, we sort the distances between each sample and the other n samples, and select k nearest neighbor samples. The computational complexity of these operations is $O(nd)$ through the bubble sorting algorithm. So the computational complexity of step 1 in LSDP is $O(n^2 d)$. In step 2, we should divide the each samples k-nearest samples by the label information, which takes $O(nk)$ computational complexity. In step 3, we take $O(n^2)$ to compute the weights of each sample with other n samples. In step 4, we should decompose the d^2 matrix in Eq. (20), which takes $O(d^3)$ computational complexity by eigen decomposition. The computational complexity in step 4 is much higher than other step because of $d \gg n$, which takes the main computational time. Step 5 is a multiplication of two matrices, which takes $O(ndr)$. Combining the above results, the computational complexity of LSDP is $O(d^3 + n^2 d + ndr + n^2)$, which indicates that LSDP is an algorithm with polynomial order.

Algorithm 1 Locality sensitive discriminant projection

Input: Dataset of n samples $X = \{x_1, x_2, \ldots, x_i, \ldots, x_n\} \in R^{d\times n}$ and the corresponding label of each sample set $\{l(x_1), l(x_2), \ldots, l(x_n)\}$

Output: The projection matrix $V \in R^{d\times r}$, which can transform any sample $x_i \in R^d$ in original space to $y_i \in R^r$ in subspace by $y_i = V^T x_i$

Step 1. Calculate k-nearest samples of any sample x_i and form the k-nearest neighbor samples set $N_k(x_i)$ of x_i.

Step 2. Divide $N_k(x_i)$ into $N_k^w(x_i)$ and $N_k^b(x_i)$ according to the sample's label information. The samples are divided into $N_k^w(x_i)$ if the label of samples in $N_k(x_i)$ is the same as the label of x_i. The samples are divided into $N_k^b(x_i)$ if the label of samples in $N_k(x_i)$ is different with the label of x_i.

Step 3. Calculate weight matrix W^w and W^b according to the weight calculation [Eqs. (10) and (12)].

Step 4. Find the optimal transformation matrix $V = \{v_1, v_2, \ldots, v_r\} \in R^{d\times r}$ by solving the eigensolver problem in Eq. (20), where the column vectors $v_1, v_2, \ldots, v_r$ be the solutions of Eq. (20), ordered according to their eigenvalues, $\lambda_1 > \lambda_2 > \ldots > \lambda_r (r \ll d)$.

Step 5. Use $y_i = V^T x_i$ to transform any samples x_i in the original space into y_i in subspace.

4 Experimental Results and Analysis

In order to evaluate the performance of the proposed LSDP algorithm, the experiments are carried out on the public face dataset ORL, Yale, and FERET compared with PCA, LDA,[21] LPP,[32] LSDA,[36] ILSDA,[41] MFA,[43] and DMMP.[42] First, the original data are reduced to the appropriate dimensions by these dimensionality reduction algorithms, and then the data after dimensionality reduction are classified by the classifier. Classifiers include nearest neighbor classifier and support vector machine and so on. For simplicity, we use nearest neighbor classifier and Euclidean distance to measure the similarity between samples. The accuracy rate of classification is the number of samples correctly classified by nearest neighbor classifier on the test set divided by the total number of samples in the test set. The classification accuracy on the test set is used to measure the performance of different dimensionality reduction algorithms. Before using LDA, LPP, MFA, and DMMP, the data are preprocessed with PCA to reduce the data to a certain dimension to avoid the problem of small size samples. Because LSDA, ILSDA, DMMP, and LSDP algorithms have no matrix inverse operation, they are not affected by the small sample problem. However, PCA preprocessing of LSDA, ILSDA, DMMP, and LSDP algorithms is still carried out, so as to make a fair comparison with other PCA preprocessing algorithms and reduce the noise of samples. The LSDP algorithm involves the problem of parameter selection, which is still unresolved in theory. In this paper, the different combinations of parameters are searched in detail through cross-validation. We give the running time of all algorithms on the Yale database and analyze the computational complexity of each algorithm on our personal computer. All experiments were programmed in Matlab2016A and executed on a computer with Intel Core i5-4200 CPU at 1.6 GHz and 8 GB physical memory.

4.1 *Experiments on the ORL Database*

ORL database[44] contains 40 persons. Each person has 10 images. These images were captured at different expressions (open or closed eyes, smiling or nonsmiling) and facial details (glasses or no glasses). These images were taken with a tolerance for some tilting and rotation of the face up to 20 deg. All images are grayscale and normalized to a

Fig. 3 Two persons in ORL, 10 images per person.

resolution of 64×64 pixels. Figure 3 shows 20 samples of two people on the ORL database, each person has 10 images. In the experiment on ORL database, p (= 4, 5, 6) samples of each person were randomly chosen as the training set and the rest were taken as the test set.

PCA was used for data preprocessing before using ILSDA and LSDP. The ILSDA and LSDP algorithm is initialized to 40 by PCA in this experiment. The parameters involved in ILSDA and LSDP were searched in detail on ORL database and was set as follows. The k-nearest neighbor parameter of ILSDA algorithm was set as 5, the heat kernel parameter t of ILSDA was set as 1, and the tradeoff parameter α of ILSDA was set as 500 according to Ref. 41 and fine tuning. The heat kernel parameter t in $K(x_i, x_j)$ of LSDP was set as d^2, and d is the standard derivation of the training set. The value of k-nearest neighbor parameter k in LSDP is set as 20. The tradeoff parameter α in LSDP was set as 0.01. The results of other algorithms are referenced from Ref. 42.

In the first experiment on ORL database, the performance of LSDP under different α was tested. Table 1 shows the relationship between the optimal recognition accuracy rates obtained by LSDP and the values of different α when $p = 4$, 5, 6. The values of α ranges from 0.1 to 0.001 with a wide range to show the corresponding accuracy rate in various cases. It can be seen from Table 1 that the value of α is not the larger the better or the smaller the better, α needs to be selected according to the actual effect. The optimal recognition accuracy rate was obtained by $\alpha = 0.01$ when $p = 4$, 5, 6. Figures 4–6 show the curves of accuracy rates versus dimensionality. It can be seen that LSDP algorithm is not sensitive to the value of α. At $p = 4$, 5, 6, the accuracy rates of LSDP versus different values of α is similar while the maximal accuracy rate was obtained at $\alpha = 0.01$. The character of not sensitive to the values of α reduces the application difficulty of LSDP in practical applications.

In the second experiment, we compare the performances of different algorithms when $p = 4$, 5, 6. Table 2 shows the optimal recognition accuracy rates of PCA, LDA, LPP, LSDA, ILSDA, MFA, DMMP, and LSDP at $p = 4$, 5, 6. According to Table 2, the accuracy rates of PCA when $p = 4$, 5, 6 are almost lower than that of LDA, LPP, LSDA, ILSDA, MFA, DMMP, and LSDP. Because PCA does not

Table 1 Optimal recognition accuracy rates (%) of LSDP versus different α on the ORL database.

Train samples	$\alpha = 0.1$	$\alpha = 0.05$	$\alpha = 0.01$	$\alpha = 0.001$
$p = 4$	90.83	93.75	94.58	94.58
$p = 5$	96.00	97.00	97.50	96.50
$p = 6$	98.75	98.75	99.38	99.38

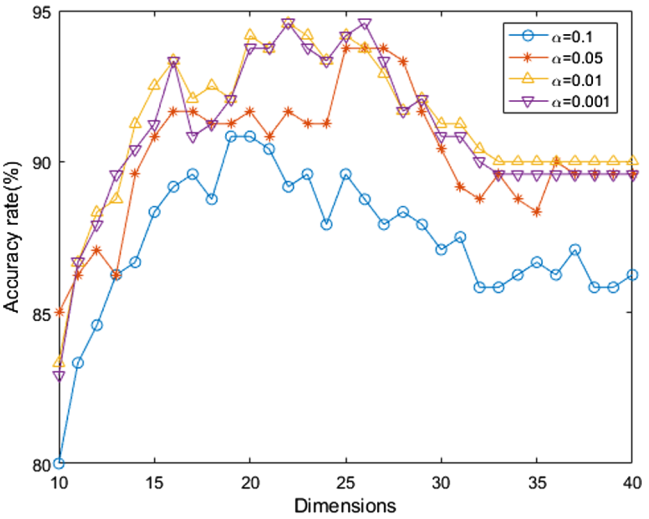

Fig.4 Accuracy rates versus dimensionality on ORL database when $p = 4$.

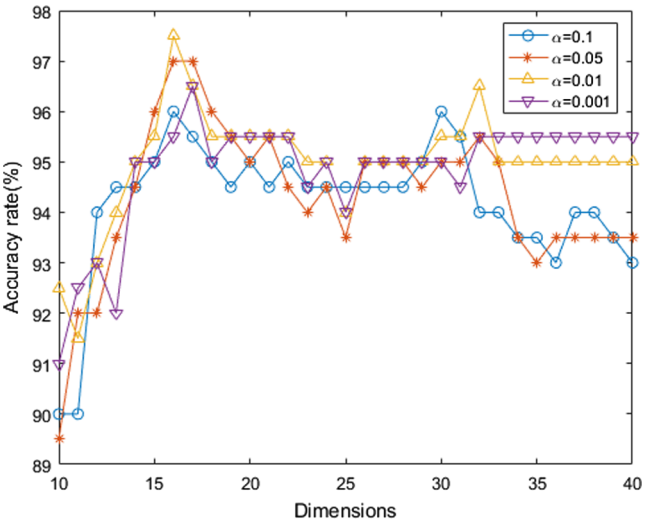

Fig. 5 Accuracy rates versus dimensionality on ORL database when $p = 5$.

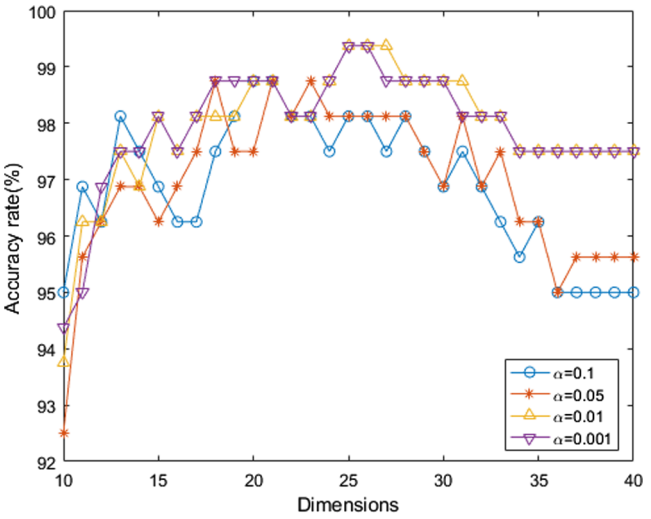

Fig. 6 Accuracy rates versus dimensionality on ORL database when $p = 6$.

Table 2 Optimal recognition accuracy rates (%) of different algorithms on the ORL database.

Method	$p = 4$	$p = 5$	$p = 6$
PCA[a]	89.17	90.00	95.63
LDA[a]	89.17	90.00	96.25
LPP[a]	89.17	91.00	95.00
LSDA[a]	90.42	92.50	96.88
MFA[a]	90.00	90.00	96.88
DMMP[a]	90.42	93.50	98.12
ILSDA	94.58	97.00	99.38
LSDP	**94.58**	**97.50**	**99.38**

[a]The results are provided by Ref. 42.
Note: Bold character represents that it is the best result.

use the label information of samples, it performs worse in classification tasks. The accuracy rates of manifold learning algorithms such as LPP, LSDA, ILSDA, DMMP, and LSDP is almost higher than the algorithms based on the global structure of data, such as PCA and LDA. This shows that the local structure of data plays an important role in ORL dataset. The accuracy rate of LSDA, ILSDA, DMMP, and LSDP is higher than that of LPP when $p = 4, 5, 6$ because LPP is an unsupervised algorithm that fails to utilize the label information of data. Compared with MFA algorithm, LSDP algorithm not only retains the local structure of intraclass samples but also retains the local structure of interclass samples so that the projected data can better reflect the essential manifold structure of the data. Therefore, when $p = 4, 5, 6$, the classification accuracy rate of LSDP algorithm is higher than that of MFA algorithm. Compared with DMMP, which is proposed in recent years, LSDP also gets the better recognition accuracy. The DMMP algorithm uses local structure of data while the weights of neighbor samples are constant value, which cannot reflect the relation of neighbor samples well. The accuracy rates of ILSDA on the ORL database are similar with LSDP when $p = 4, 5, 6$. ILSDA algorithm uses the intraclass scatter matrix in LDA to add the compactness of intraclass samples, which shows that the algorithm performance can be improved by alleviating the outlier problem. Compared with ILSDA, LSDP uses the weights of neighbor samples and non-neighbor samples to solve the intraclass outlier problem and maintain the intraclass local structure, which also achieves better accuracy rates on the ORL database.

Figure 7 is the 19 nearest neighbor images of an image when $p = 4$ on a random training set for further analysis. As shown in Fig. 7, the first image is as a comparison, the rest 19 images are the 19 nearest neighbor samples of the first image, sorted by Euclidean distance with the first image increasing. As we can see, the neighbor samples of the first image. The classes of rest 19 images are different with the first image, which shows the outlier problem existing in the ORL database. As we can see, the neighbor samples of the first image in Fig. 7 have the similar face orientation, which leads to the closer Euclidean distance. The intraclass outlier problem caused by face orientation is common in face database. Compared with LSDA algorithm, LSDP solves the outlier problem by increasing the compactness of intraclass samples in subspace. The experimental results in Table 2 show that the accuracy of LSDP is higher than that of PCA, LDA, LPP, LSDA, MFA, and DMMP when $p = 4, 5, 6$, which indicates the effectiveness of LSDP.

4.2 *Experiments on the Yale Database*

Yale database[45] contains 165 images of 15 different people, each person has 11 images taken under different facial expressions, such as happy, sad, surprised, wearing glasses

Fig. 7 Nineteen nearest neighbor images of the first image in a random split on ORL database when $p = 4$.

Fig. 8 Two persons in Yale, 11 images per person.

or not, and different lighting directions. The images on the Yale database were grayscale and normalized to the resolution of 64×64 pixels. Some images are shown in Fig. 8. On the Yale database, p (= 4, 5, 6) samples of each person were randomly chosen as the training set and the rest were taken as the test set.

The parameters of LSDP algorithm are determined by cross-validation. We set the dimension preprocessed by PCA as 120 before using ILSDA and LSDP. On the Yale database, the k-nearest neighbor parameter k of LSDP was set as 25. The tradeoff parameter α is set as 0.01. The heat kernel parameter t of LSDP is set as d^2, which is the same as those in the experiments on ORL database. The k-nearest neighbor parameter of ILSDA was set as 5, the heat kernel parameter t of ILSDA was set as 1, and the tradeoff parameter α of ILSDA was set as 500 according to Ref. 41.

In the first experiment on the Yale database, we compared the optimal recognition accuracy rates of different algorithms when $p = 4, 5, 6$ to evaluate the performance of algorithms. The optimal recognition accuracy rates of each algorithm when $p = 4, 5, 6$ are shown in Table 3. According to Table 3, the accuracy rates of PCA are always lower than other algorithms because it does not use the label information of samples, which is important in classification task. The accuracy rates of LDA are higher than PCA because it uses the label information to enhance the performance in classification task. While the unsupervised LPP algorithm achieved better performance than PCA and LDA because it used the local structure of samples, which indicates the importance of local structure. At the same time, the accuracy rates of LSDA, MFA, and ILSDA are lower than LPP because to maximize the distances of interclass neighbor samples will destroy the local structure of the samples. The DMMP algorithm achieved better accuracy rates than LDA because it used the local structure of samples. The accuracy rates of LSDP algorithm are always higher than ILSDA algorithm when $p = 4, 5, 6$. Compared with ILSDA, LSDP maintains the local structure of intraclass samples by using the weights of intraclass samples. Since ILSDA integrates intraclass scatter that contains the global information of samples, it the local structure of intraclass samples. LSDP is superior to LSDA algorithm when $p = 4$, and LSDP is superior to MFA algorithm when $p = 5$. The accuracy rates of LSDP does not always higher than LSDA and MFA when $p = 4, 5$ because LSDP keeping the interclass local structure enhances the generalization ability of algorithm rather than the discriminating ability of algorithm. With the increase of p, the accuracy of the LSDP increases continuously. Increasing the number of training samples can improve the generalization ability of the LSDP algorithm on the new samples, so as to improve the recognition accuracy. When $p = 6$, LSDP achieves the maximal recognition accuracy rate. In order to further analyze the LSDP, we show the nearest neighbor samples of samples on the Yale database when $p = 6$.

Table 3 Optimal recognition accuracy rates (%) of different algorithms on the Yale database.

Method	$p = 4$	$p = 5$	$p = 6$
PCA[a]	91.43	92.22	92.00
LDA[a]	93.33	96.67	97.33
LPP[a]	95.24	97.78	98.67
LSDA[a]	92.38	97.78	96.00
MFA[a]	95.24	95.56	96.00
DMMP[a]	**97.14**	**100**	**98.67**
ILSDA	91.43	94.44	96.00
LSDP	93.33	96.67	**98.67**

[a]The results are provided by Ref. 42.
Note: Bold character represents that it is the best result.

As shown in Fig. 9, on the training set randomly split by setting $p = 6$, the 19 nearest neighbor images of the first image have different classes with the first image. This is the outlier problem described in Fig. 1, indicating the existence of an intraclass outlier problem on Yale database. As shown in Fig. 9, the first image is close to its 19 nearest neighbor images in pixel intensity, which results in a smaller distance between them when calculating the nearest neighbor samples. This kind of outlier problem caused by illumination is popular in practical application.

At the same time, it can be seen from Table 3 that the LSDP obtains the maximal recognition accuracy rate when $p = 6$. This reason is that LSDP algorithm increases the intraclass compactness after projected to solve the outlier problem shown in Fig. 9.

In the second experiment on the Yale database, we compare the running time of each algorithm and analyze their computational complexity. The 40 running times of each

Fig. 9 19 nearest neighbor images of the first image in a random split on Yale database when $p = 6$.

Table 4 The 40 times running time (s) of different algorithms on the Yale database.

Method	$p = 4$	$p = 5$	$p = 6$
PCA	0.266	0.352	0.409
LDA	0.258	0.381	0.543
LPP	0.500	0.648	0.757
LSDA	0.673	0.878	1.005
MFA	0.347	0.348	0.395
DMMP	0.552	0.719	0.842
ILSDA	0.231	0.288	0.360
LSDP	0.541	0.680	0.865

algorithm when $p = 4, 5, 6$ on the Yale database are given in Table 4. The main computational complexity of these algorithms is eigen decomposition, which consumes $O(d^3)$. As we can see, the running times of ILSDA are lower than other algorithms because ILSDA uses two steps to decompose the matrix according to the rank of the matrix. This consideration of computational complexity on ILSDA algorithm reduces the computational complexity of eigen decomposition. The running time of PCA and LDA is lower than other algorithms that need to compute the k-nearest neighbor samples. Calculating the nearest neighbor samples in the LPP, LSDA, MFA, DMMP, and LSDP algorithms takes more time than calculating the global structure in PCA and LDA. The MFA algorithm constructs the intraclass graph and interclass graph that reduces the scale in calculating k-nearest neighbor samples and leads to the lower running time. The LPP need not divide the neighbor samples according to label information so it takes lower running time than LSDA, DMMP, and LSDP. The LPP, LSDA, DMMP, and LSDP take the similar running time on the Yale database because they all need the neighbor samples' calculating and neighbor samples' dividing.

The LSDP algorithm solves the outlier problem caused by the illumination on the Yale database and obtains the highest recognition accuracy with the increase of the number of samples. According to the computational complexity analyzed above, the LSDP algorithm does not increase the running time compared with LSDA, which indicates the effectiveness of the LSDP algorithm.

4.3 *Experiments on the FERET Database*

FERET database[46] contains 14,126 images from 1199 persons. The subset of FERET was used in experiments, which contains 200 persons, each with seven images. The original images are grayscale and normalized to a resolution of 64×64 pixels. Some processed images are shown in Fig. 10.

We set $p = 3$ in the following experiments on FERET database, that is, three images for each person were used as the training set, and the remaining four images were used as the test set according to Ref. 42. The parameters of ILSDA and LSDP are determined by cross-validation. Before using ILSDA and LSDP algorithm, PCA is used to preprocess the original data to reduce to 120-dimensions. After a detailed searching of parameter combinations, LSDP algorithm needs to choose three parameters, the k-nearest neighbor parameter k of LSDP was set as 10. The tradeoff parameter α is set as 0.01. The heat kernel parameter t of LSDP is set as $0.5 \times d^2$, d is the standard derivation of the training set. The k-nearest neighbor parameter of ILSDA algorithm was set as 10, the heat kernel parameter t of ILSDA was set as 1, the tradeoff parameter α of ILSDA was set as 100 according to the literature,[41] and fine tuning. The results of PCA, LDA, LPP, LSDA, MFA, and DMMP on FERET are referenced from Ref. 42.

In the first experiment on the FERET database, we analyzed the influence of heat kernel parameter t on LSDP. Table 5 shows the relationship between the accuracy rate of LSDP and the value of t. From Table 5, the accuracy rate of LSDP algorithm on FERET database changes greatly with the value of t. The accuracy rate of LSDP is 28.88% when $t = 0.1 \times d^2$, whereas the maximal accuracy rate of LSDP on FERET database is 50.25% when $t = 0.5 \times d^2$. This shows that although the heat kernel parameter t increases the flexibility of the algorithm, LSDP algorithm is sensitive to the value of t. Therefore, the value of heat kernel parameter t needs to be carefully selected in practical application.

In the second experiment on the FERET database, we analyzed the influence of the k-nearest neighbor parameter k on the algorithm. Table 6 shows the relationship between k-nearest neighbor parameter k and the accuracy of LPP, LSDA, ILSDA, DMMP, and LSDP on FERET database. It can be seen from Table 6 that the accuracy of ILSDA and LSDP algorithm on the FERET database changes little with the different values of k, and LSDP achieves the maximal accuracy when $k = 10$. The accuracy rate of ILSDA does

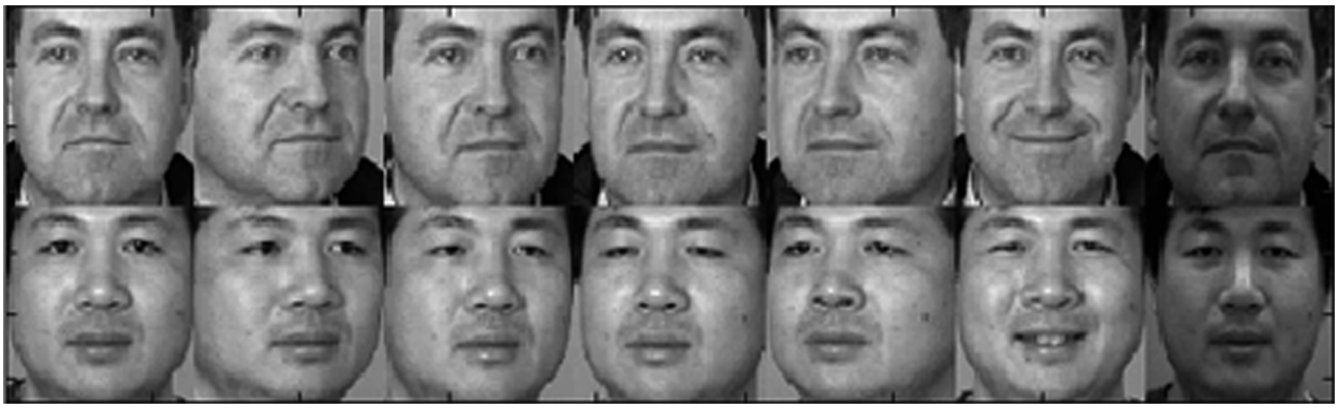

Fig. 10 Two persons in FERET, seven images per person.

Table 5 Optimal recognition accuracy rates (%) of LSDP versus t on FERET database when $p = 3$.

Parameter t	$t = 0.1 \times d^2$	$t = 0.5 \times d^2$	$t = 1 \times d^2$	$t = 1.5 \times d^2$
LSDP	28.88	50.25	46.75	46.00
Dimensions	86	47	60	67

Table 6 Optimal recognition accuracy rates (%) of LPP, LSDA, ILSDA, DMMP, and LSDP versus parameter k on the FERET database when $p = 3$.

Method	$k = 5$	$k = 10$	$k = 15$	$k = 20$
LPP[a]	34.50	34.63	34.88	34.63
LSDA[a]	32.12	32.50	33.62	34.13
DMMP[a]	45.00	45.50	47.88	48.25
ILSDA	48.75	48.75	48.75	48.75
LSDP	50.00	50.25	49.13	48.13

[a]The results are provided by Ref. 42.

not change much with the k because ILSDA has used the intraclass scatter, which will reduce the importance of local structure in algorithm. The accuracy of LSDA increases with the value of k, indicating that LSDA algorithm is seriously affected by outliers on the FERET database, and it needs to increase parameter k to alleviate the outlier problem. Figure 11 shows that when $p = 3$, there are outliers on the training set. Figure 11 is the 19 nearest neighbor images of an image when $p = 3$ on the training set. As shown in Fig. 11, the first image is as a comparison, and the other 19 images are the 19 nearest neighbor samples of the first image. Only the 19th image in Fig. 11 has the same class as the first image in Fig. 11. According to the analysis in Sec. 3.1, if the k-nearest neighbor parameter k of LSDA is small, such as less than 18, then all the neighbor samples of the first image have different classes with the first image in Fig. 11, which leads to the outlier problem described in Sec. 3.1.

The third experiment on the FERET database compared to the accuracy rate of different algorithms Table 7 shows the accuracy rates and the corresponding dimension of PCA, LDA, LPP, LSDA, ILSDA, MFA, DMMP, and LSDP when $p = 3$ on FERET database. The results of PCA, LDA, LPP, LSDA, DMMP, and MFA are referenced from Ref. 42. As can be seen from Table 7, the accuracy rate of PCA, LDA, ILSDA, MFA, DMMP, and LSDP is higher than that of LPP and LSDA. The reason is that the training set when $p = 3$ on FERET makes the algorithm more susceptible to outlier problems. Therefore, the accuracy rate of the manifold learning algorithms LPP and LSDA based on the local structure of data is lower than those algorithms based on the global structure of data. MFA algorithm independently constructs the intraclass graph and interclass graph, which reduces the outlier problem caused by dividing the neighbor graph into the intraclass neighbor graph and interclass neighbor graph. Therefore, the accuracy rate of MFA is

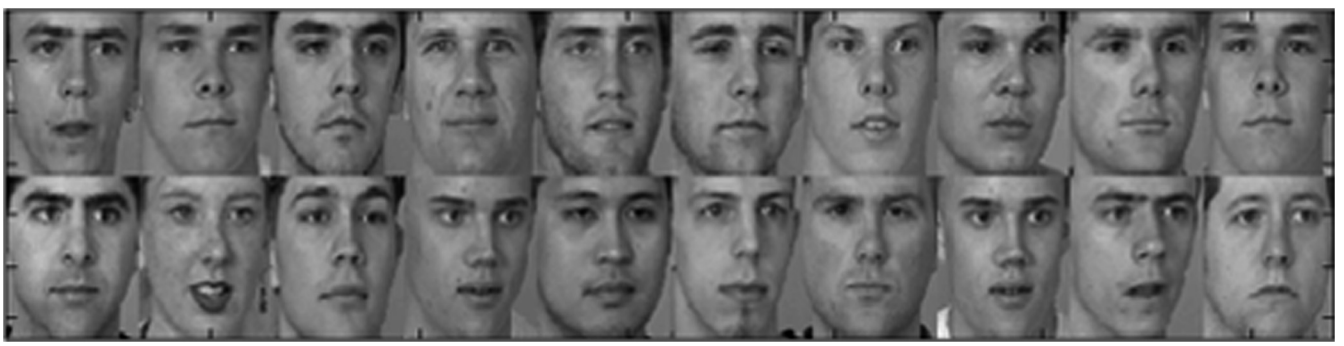

Fig. 11 Nineteen nearest neighbor images of the first image in a random split on FERET database when $p = 3$.

Table 7 Optimal recognition accuracy rates (%) of different algorithms on the FERET database when $p = 3$.

Method	PCA[a]	LDA[a]	LPP[a]	MFA[a]	LSDA[a]	DMMP[a]	ILSDA	LSDP
Accuracy rate	42.5	46.25	34.88	44.5	34.13	48.25	48.75	**50.25**
Dimensions	40	60	70	90	40	40	71	**47**

[a]The results are provided by Ref. 42.
Note: Bold character represents that it is the best result.

higher than that of LSDA. ILSDA and LSDP algorithms avoid the outlier problem, so its accuracy rate is higher than LSDA.

LSDP algorithm not only maintains the manifold structure between with-class samples but also maintains the local neighbor relationship of interclass neighbor samples, which makes it easier to find the intrinsic manifold of data, so it has a higher accuracy than ILSDA and MFA. ILSDA solves the outlier problem by minimizing the intraclass scatter and achieves a better performance than LSDA. While ILSDA algorithm ignores the local structure of interclass neighbor samples, which will lead the poor performance compared with LSDP. The LSDP maintains the interclass local structure, which will reveal the underlying pattern information in subspace. Meanwhile, LSDP obtains a higher accuracy rate than DMMP, which is the state-of-the-art method proposed in recent years because LDSP maintains the intraclass and interclass local structure simultaneously. The LSDP obtains maximal accuracy rate 50.25% on FERET database than other algorithms in Table 7, which indicates the effectiveness of LSDP algorithm.

5 Conclusion and Future Work

In this paper, an algorithm called LSDP is proposed for feature extraction in face recognition. LSDP solves the problem of outliers within the same class by increasing the compactness of the projected samples within the same class. At the same time, LSDP maintains the local structure to retain more information and maintains the manifold structure of data. Experiments and analysis on ORL, Yale, and FERET datasets verify the effectiveness of LSDP algorithm. The algorithm LSDP is more efficient than the directly related LPP, LSDA, ILSDA, MFA, and DMMP. LSDP algorithm proposed in this paper needs three parameters to improve the flexibility of the algorithm. While currently there is no effective method to choose the parameters in theory, the setting of parameters is mainly by cross-validation. In the future, work can be further carried out to find effective parameter optimization methods.

Acknowledgments

This work was supported by research funds of CCNU from the colleges' basic research (innovation funding projects) (Grant No. 2018CXZZ040).

References

1. Y. Yi et al., "Joint graph optimization and projection learning for dimensionality reduction," *Pattern Recognit.* **92**, 258–273 (2019).
2. C. Jin, S. W. Jin, and L. N. Qin, "Attribute selection method based on a hybrid BPNN and PSO algorithm," *Appl. Soft Comput.* **12**(8), 2147–2155 (2012).
3. R. Nie et al., "Facial feature extraction using frequency map series in PCNN," *J. Sens.* **2016**, 1–9 (2016).
4. C. Yu, R. Nie, and D. Zhou, "A regularized locality projection-based sparsity discriminant analysis for face recognition," *Int. J. Pattern Recognit. Artif. Intell.* **32**(5), 1856006 (2018).
5. N. Abudarham, L. Shkiller, and G. Yovel, "Critical features for face recognition," *Cognition* **182**, 73–83 (2019).
6. J. Gui et al., "Locality preserving discriminant projections for face and palmprint recognition," *Neurocomputing* **73**(13–15), 2696–2707 (2010).
7. B. Chen et al., "Max-margin discriminant projection via data augmentation," *IEEE Trans. Knowl. Data Eng.* **27**(7), 1964–1976 (2015).
8. Y. Yi et al., "Non-negative matrix factorization with locality constrained adaptive graph," *IEEE Trans. Circuits and Syst. Video Technol.* (2019).
9. C. Jin and S. W. Jin, "Automatic image annotation using feature selection based on improving quantum particle swarm optimization," *Signal Process.* **109**, 172–181 (2015).
10. X. Y. Jing et al., "Face feature extraction and recognition based on discriminant subclass-center manifold preserving projection," *Pattern Recognit. Lett.* **33**(6), 709–717 (2012).
11. Y. Yi et al., "Adaptive multiple graph regularized semi-supervised extreme learning machine," *Soft Comput.* **22**(11), 3545–3562 (2018).
12. Y. Zhu, S. Chen, and Q. Tian, "Spatial regularization in subspace learning for face recognition: implicit vs. explicit," *Neurocomputing* **173**, 1554–1564 (2016).
13. J. Wang et al., "Locality constrained graph optimization for dimensionality reduction," *Neurocomputing* **245**, 55–67 (2017).
14. P. Huang et al., "Collaborative representation based local discriminant projection for feature extraction," *Digital Signal Process.* **76**, 84–93 (2018).
15. M. Pang et al., "Robust heterogeneous discriminative analysis for face recognition with single sample per person," *Pattern Recognit.* **89**, 91–107 (2019).
16. Y. Yi et al., "Semi-supervised ridge regression with adaptive graph-based label propagation," *Appl. Sci.* **8**(12), 2636 (2018).
17. G. F. Lu et al., "Matrix exponential based discriminant locality preserving projections for feature extraction," *Neural Networks* **97**, 127–136 (2018).
18. C. Ornek and E. Vura, "Nonlinear supervised dimensionality reduction via smooth regular embeddings," *Pattern Recognit.* **87**, 55–66 (2019).
19. D. Zhang et al., "A new linear discriminant analysis algorithm based on L1-norm maximization and locality preserving projection," *Pattern Anal. Appl.* **21**(3), 685–701 (2018).
20. J. Li, "Unsupervised robust discriminative manifold embedding with self-expressiveness," *Neural Networks* **113**, 102–115 (2019).
21. V. P. Kshirsagar, M. R. Baviskar, and M. E. Gaikwad, "Face recognition using Eigenfaces," in *IEEE Comput. Soc. Conf. Comput. Res. and Dev.*, pp. 586–591 (1991).
22. M. A. Turk and A. P. Pentland, "Eigenfaces for recognition," *J. Cognit. Neurosci.* **3**(1), 71–86 (1991).
23. P. N. Belhumeur, J. P. Hespanha, and D. J. Kriegman, "Eigenfaces vs. fisherfaces: recognition using class specific linear projection," *IEEE Trans. Pattern Anal. Mach. Intell.* **19**(7), 711–720 (1997).
24. Q. Gao et al., "Enhanced fisher discriminant criterion for image recognition," *Pattern Recognit.* **45**(10), 3717–3724 (2012).
25. H. F. Li, T. Jiang, and K. S. Zhang, "Efficient and robust feature extraction by maximum margin criterion," *IEEE Trans. Neural Networks* **17**(1), 157–165 (2006).
26. Y. Cui and L. Y. Fan, "A novel supervised dimensionality reduction algorithm: graph-based fisher analysis," *Pattern Recognit.* **45**(4), 1471–1481 (2012).
27. J. B. Tenenbaum, V. de Silva, and J. C. Langford, "A global geometric framework for nonlinear dimensionality reduction," *Science* **290**(5500), 2319–2323 (2000).
28. S. T. Roweis and L. K. Saul, "Nonlinear dimensionality reduction by locally linear embedding," *Science* **290**(5500), 2323–2326 (2000).
29. M. Belkin and P. Niyogi, "Laplacian eigenmaps for dimensionality reduction and data representation," *Neural Comput.* **15**(6), 1373–1396 (2003).
30. T. J. Chin and D. Suter, "Out-of-sample extrapolation of learned manifolds," *IEEE Trans. Pattern Anal. Mach. Intell.* **30**(9), 1547–1556 (2008).

31. X. He et al., "Neighborhood preserving embedding," in *10th IEEE Int. Conf. Comput. Vision*, pp. 1208–1213 (2005).
32. X. F. He et al., "Face recognition using Laplacian faces," *IEEE Trans. Pattern Anal. Mach. Intell.* **27**(3), 328–340 (2005).
33. D. Cai et al., "Orthogonal Laplacianfaces for face recognition," *IEEE Trans. Image Process.* **15**(11), 3608–3614 (2006).
34. W. K. Wong and H. T. Zhao, "Supervised optimal locality preserving projection," *Pattern Recognit.* **45**(1), 186–197 (2012).
35. L. Yang et al., "Null space discriminant locality preserving projections for face recognition," *Neurocomputing* **71**(16–18), 3644–3649 (2008).
36. D. Cai et al., "Locality sensitive discriminant analysis," in *20th Int. Joint Conf. Artif. Intell.*, pp. 1713–1726 (2007).
37. H. T. Zhao et al., "Local structure based supervised feature extraction," *Pattern Recognit.* **39**(8), 1546–1550 (2006).
38. Q. Gao et al., "Stable locality sensitive discriminant analysis for image recognition," *Neural Networks* **54**, 49–56 (2014).
39. P. Huang et al., "Discriminant similarity and variance preserving projection for feature extraction," *Neurocomputing* **139**, 180–188 (2014).
40. Q. Hua et al., "Local similarity and diversity preserving discriminant projection for face and handwriting digits recognition," *Neurocomputing* **86**, 150–157 (2012).
41. Y. Yi et al., "An improved locality sensitive discriminant analysis approach for feature extraction," *Multimedia Tools Appl.* **74**(1), 85–104 (2015).
42. Z. Yang et al., "Discriminant maximum margin projections for face recognition," *Multimedia Tools Appl.* 1–19 (2018).
43. S. Yan et al., "Graph embedding and extensions: a general framework for dimensionality reduction," *IEEE Trans. Pattern Anal. Mach. Intell.* **29**(1), 40–51 (2007).
44. "ORL face database," http://www.cam-orl.co.uk, AT&T Laboratories, Cambridge.
45. Center for Computational Vision and Control at Yale University, "Yale face database," http://cvc.yale.edu/projects/yalefaces/yalefaces.html.
46. National Institute of Standards and Technologe, "FERET face database," https://www.nist.gov/itl/iad/image-group/color-feret-database.

Yi-Kang Wei is a master's candidate in the School of Computers, Central China Normal University, Wuhan, Hubei, China. His main research interests include digital image processing, dimensionality reduction, and face recognition.

Cong Jin received her PhD from the Institute for Image Recognition and Artificial Intelligence, Huazhong University of Science and Technology, Wuhan, China. She is a professor of the School of Computers, Central China Normal University, Wuhan, Hubei, China. Her main research interests include digital image processing, pattern recognition, and intelligence information processing.