Neurocomputing 86 (2012) 150–157

Contents lists available at SciVerse ScienceDirect

Neurocomputing

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

Local similarity and diversity preserving discriminant projection for face and handwriting digits recognition

Qiang Hua [a,b,*], Lijie Bai [b], Xizhao Wang [b], Yuchao Liu [b]

[a] Department of Computer Science and Engineering, Nanjing University of Aeronautics & Astronautics, Nanjing 210016, PR China
[b] College of Mathematics and Computer Science, Hebei University, Baoding 071002, PR China

ARTICLE INFO

Article history:
Received 13 October 2011
Received in revised form
24 December 2011
Accepted 19 January 2012
Communicated by D. Wang
Available online 3 March 2012

Keywords:
Dimensionality reduction
Face recognition
Subspace
Diversity
Locality preserving projection

ABSTRACT

In this paper, a novel supervised subspace learning algorithm, named local similarity and diversity preserving discriminant projection (LSDDP), is presented. LSDDP defines two weighted adjacency graphs, namely similarity graph and diversity graph. LSDDP constructs the similarity scatter and diversity scatter with the weights, which are adjustable according to the global supervisor and the local semi-supervisor information of the data. Thus LSDDP could utilize both the similarity and diversity information of the data simultaneously for dimensionality reduction. After characterizing the similarity scatter and diversity scatter, a concise feature extraction criterion arised via minimizing the difference between them and the optimal projection is obtained by performing the eigen-decomposition. Thus our method successfully addresses the SSS problem without losing any discriminating information. Finally the proposed model is verified by the face and handwriting digits recognition experiments. The experimental results on Yale, ORL and CMU-PIE face database and the USPS handwriting digits database indicate the effectiveness of our method.

1. Introduction

With the rapid development of data acquisition technology, the number of the high-dimensional data is increasing in the pattern recognition filed. This presents many challenges to the development of the pattern recognition. The large quantities of features may even degrade the performance of the classifier and the curse of high dimensionality limits many practical technologies. Dimensionality reduction [1–3] as an effective technique to overcome the curse of dimensionality has been attracting much attention and is quite desirable not only in the aspect of enhancing the performance of the classifiers but also in terms of data storage and the computational complexity [4].

Over the past few decades, a large volume of dimensionality reduction methods have been proposed and most of them have been successfully applied to many applications such as face recognition and so on [5,6]. These methods can be roughly categorized into two classes: supervised and unsupervised. Up to now, PCA [7] is the most popular unsupervised method. PCA constructs a low-dimensional representation of the original data via minimizing the reconstruction error. Since PCA ignores the label information, it has little to do with the classification task. Unlike PCA, LDA [8] is one of the most canonical supervised methods. LDA can find an optimal projection by maximizing the ratio of the trace of the between-class scatter to the trace of the within-class scatter. Due to utilizing the available class information, LDA is more effective than PCA in many applications. It is noteworthy to point out that both PCA and LDA share a common characteristic, that is, only the global linear Euclidean structure of data are taken into account and no attention are paid to the local structure of the data. As a result, once the data points are resided on a nonlinear manifold, both PCA and LDA fail to explore the essential structure of the data.

Recently many researches have indicated that nature images especially the face images possibly reside on a nonlinear low-dimensional sub-manifold. A lot of researchers are attracted to straightforwardly find the inherent nonlinear structure of the data and then many manifold learning algorithms are proposed. Among them, locally preserving projection (LPP) [9], as a linear approximation to Laplacian Eigenmap [10,11], is one of most representative. Unlike PCA and LDA, LPP seeks to preserve the local information by preserving the neighborhood of the data and has been successfully applied into many practical applications such as face recognition and yielded impressive result. Compared with other manifold learning methods, LPP possesses an obvious advantage that the map of LPP is explicit and is easy to compute.

Now some improvements to the original LPP have been developed. To boost the ability of local preserving, which has a close relation to the recognition ability, Cai et al. proposed an orthogonal LPP algorithm (OLPP) [12]. In [13], Yang et al. presented a UDP algorithm. UDP characterizes both the local scatter

* Corresponding author at: Department of Computer Science and Engineering, Nanjing University of Aeronautics & Astronautics, Nanjing 210016, PR China. Tel.: +86 3125079638.
E-mail addresses: huaq@mail.hbu.cn, huaq@nuaa.edu.cn (Q. Hua).

0925-2312/$ - see front matter © 2012 Elsevier B.V. All rights reserved.
doi:10.1016/j.neucom.2012.01.031

and the non-local scatter, which makes UDP more powerful. O-LPP [14], a novel subspace learning approach, was proposed to respond to the SSS problem. Many other techniques are closely related to or explicitly build on original LPP [15,16]. However these aforementioned methods lose the sight of the class information, which is very important to recognition tasks. To make full use the available label information many supervised edition of LPP have been presented. SLPP [17] was developed to improve the discriminating capability of LPP. LDP [18] incorporates the local structure and class information. Null space discriminant locality preserving projection [19] was proposed by Yang et al. to overcome the SSS problem of DLPP [20]. Based on OLPP, Zhu and Zhu presented an ODLPP [21] method. Li and Wang developed an orthogonal discriminant projection (ODP) [22], which is a modified method for UDP. Gui et al. introduced a novel supervised LPP algorithm, namely LPDP [23], by adding the criterion of maximum margin criterion (MMC) into the objective function of LPP. LPDP could preserve both global discriminant and local structure of the data. This remarkable property endows LPDP with stronger recognition ability.

One common limitation of the above mentioned algorithms is that they only pay attention on the local similarity information and neglect the diversity information of the sample. However impairing the diversity information of the sample may arise the over learning problem. From the perspective of statistic, Gao et al. introduced the diversity information between two points and proposed S-LSDP [24], a robust technique for the classification task, which overcomes the over learning problem successfully. Unfortunately there are still some limitations existed in S-LSDP. First, the weighted adjacency diversity graph is constructed without the guide of the class information. Second, the similarity matrix defined in S-LSDP makes the neighborhood graph of the input data disconnected. Some researches show that the manifold structure of the data also is very important as well as the class information for the classification task. Finally, in many practical applications, especially in the face recognition, S-LSDP often encounters the SSS problem.

In this paper a novel supervised dimensionality reduction method, namely LSDDP, is proposed to address the shortcomings of S-LSDP and further boost the discriminating capability of LPP. In LSDDP, both the weighted adjacency similarity graph and weighted adjacency diversity graph take the local structure information and the label information into account, thus LSDDP can preserve the main manifold structure of the data and has more powerful discriminating ability. Furthermore, using the two weighted adjacency graph LSDDP defines two matrixes: namely similarity scatter matrix and diversity scatter matrix. After charactering the similarity scatter matrix and the diversity scatter matrix, a concise feature extraction criterion is raised via minimizing the difference between the similarity scatter and the diversity scatter. Then the optimal projection is obtained by solving an eigen-equation, where the SSS problem is successfully overcome.

The paper is organized as following: in Section 2 we will simply review S-LSDP. The new technique will be introduced in Section 3. The experimental results for applying to the face and handwriting digits recognition will be offered in Section 4. Followed by, the brief conclusions about this paper are given in Section 5.

2. Related work on S-LSDP

In order to address the over-fitting problem of SLPP [17], Gao et al. proposed the S-LSDP method. In S-LSDP two weighted graph are constructed: weighted adjacency similarity graph and weighted adjacency diversity graph. The former, constructed using the class information, measures the similarity of the points and the latter measures the diversity of the points.

Let $X=[X_1,X_2,\cdots,X_N]$ denote the sample set and the dimensionality of each element is D. $G_S=(V, E, S)$ and $G_d=(V, E, B)$, denote the weighted adjacency similarity graph and weighted adjacency diversity graph, respectively, where V is the set of vertices, E is the set of edges connecting the vertices; S is a weighted matrix with elements characterizing the similarity of two points and B is a weighted matrix with the elements characterizing the diversity of two points.

The element of weighted matrix S is defined as following:

$$S_{ij}=\begin{cases}\exp\left(\frac{-\|X_i-X_j\|^2}{t}\right) & \begin{array}{l}\text{if } X_i \text{ is among } k \text{ nearest neighbors of } X_j\\ \text{or } X_j \text{ is among } k \text{ nearest neighbors of } X_i\\ \text{and } \tau_i=\tau_j\end{array}\\ 0 & \text{otherwise}\end{cases} \tag{1}$$

where τ_i is the class label of X_i, and $t\in(0,+\infty)$.

The element of weighted matrix B can be calculated as follows:

$$B_{ij}=\begin{cases}\exp\left(\frac{-b}{\|X_i-X_j\|^2}\right) & \begin{array}{l}\text{if } X_i \text{ is among } k1 \text{ nearest neighbors of } X_j\\ \text{or } X_j \text{ is among } k1 \text{ nearest neighbors of } X_i\end{array}\\ 0 & \text{otherwise}\end{cases} \tag{2}$$

where $b\in(0,+\infty)$. B_{ij} measures the contribution of X_i relative to X_j to the diversity information.

The aim of S-LSDP is to find a discriminative projection that can preserve the similarity and the diversity information at the same time. The feature extraction criterion of S-LSDP can be expressed as follows:

$$J(W)=\arg\min_{W^TW=I}\frac{W^TG_LW}{W^TG_NW} \tag{3}$$

where $G_L=XLX^T$ is the weighted similarity scatter matrix, $G_N=X\overline{L}X^T$ denotes the weighted diversity scatter matrix of the pattern, where $L=D-S$ and $\overline{L}=\overline{D}-B$ are the Laplacian matrix, $D_{ii}=\sum_j S_{ij}$,$\overline{D}_{ii}=\sum_j B_{ij}$. W is the transformation matrix, which is obtained by solving the generalized eigen-equation $G_LW=\lambda G_NW$, and W subjects to:

$$W_i^TW_j=\begin{cases}0 & i\neq j\\ 1 & i=j\end{cases}. \tag{4}$$

3. Local similarity and diversity preserving discriminant projection

As discussed in the Section 1, the weighted adjacency diversity graph in S-LSDP is constructed without considering the class information, which will inevitably makes S-LSDP cannot achieve satisfactory result. In addition, S-LSDP often suffers from the SSS problem in practical application. To overcome the problem, PCA is first introduced to project the original data into a lower dimensional feature space, and then S-LSDP is applied in the PCA feature subspace. But this may lead to some useful information be thrown away in the PCA step. In order to address the limitation mentioned above and further enhance the recognition ability of original LPP, LSDDP is put forward in this paper. LSDDP adjusts the weights of the adjacency similarity graph and the adjacency diversity graph according to the label information and the local structure. In LSDDP, instead of generalized eigen-decomposition, the optimal projection is found by performing eigen-decomposition. Thus our method will not suffer from the SSS problem.

The weighted matrix S defined in (1) will distort the neighborhood relationship of the input data, which contradicts to the idea of LPP. In other words the manifold structure of the data, which is also very important for classification, is distorted. In order to better preserve the manifold structure of the data and explore the available label information of the data, the elements of the weighted matrix S can be redefined as following [18]:

$$S_{ij}=\begin{cases}\exp(-\|X_i-X_j\|^2/t)(1+\exp(-\|X_i-X_j\|^2/t)) & \begin{array}{l}\text{if } X_i \text{ is among } k \text{ nearest neighbors of } X_j\\ \text{or } X_j \text{ is among } k \text{ nearest neighbors of } X_i\\ \text{and } \tau_i=\tau_j\end{array}\\ \exp(-\|X_i-X_j\|^2/t)(1-\exp(-\|X_i-X_j\|^2/t)) & \begin{array}{l}\text{if } X_i \text{ is among } k \text{ nearest neighbors of } X_j\\ \text{or } X_j \text{ is among } k \text{ nearest neighbors of } X_i\\ \text{and } \tau_i\neq\tau_j\end{array}\\ 0 & \text{otherwise}\end{cases} \tag{5}$$

where τ_i is the class label of X_i, and $t\in(0,+\infty)$. S_{ij} measures the discriminating similarity between two points. The advantages of the discriminating similarity [18] can be summarized as follows:

Property 1. When the Euclidean distance is equal; the weight of the points in homogeneous class is larger than that in heterogeneous class. That is to say the points in same labels is much similar than that in different class. This is favorable for recognition.

Property 2. The discriminating similarity has the ability of neighborhood preserving, thus the main geometric structure of the data set can be largely preserved.

Property 3. The value of the discriminating similarity decreases toward to 0 with the increase of the Euclidean distance. This endows the discriminating similarity with the ability of preventing noise, i.e. the more distant points from homogeneous class the less similar to each other.

Let $Y_i=W^TX_i$ be the image of X_i in the low dimensional space, where W is the transformation matrix. In order to preserve the local information of the data, we impose the following objective:

$$\min\sum_{ij}\|Y_i-Y_j\|^2S_{ij} \tag{6}$$

Substituting $Y_i=W^TX_i$ into Eq. (6), then the Eq. (6) can be rewritten as follows:

$$\begin{aligned}\sum_{ij}\|Y_i-Y_j\|^2S_{ij}&=\sum_{ij}\|W^TX_i-W^TX_j\|^2S_{ij}\\ &=2tr\left(\sum_{ij}W^TX_iS_{ij}X_i{}^TW-\sum_{ij}W^TX_iS_{ij}X_j{}^TW\right)\\ &=2tr(W^TXDX^TW-W^TXSX^TW)\\ &=2tr(W^TX(D-S)X^TW)\\ &=2tr(W^TXLX^TW)\\ &=2tr(W^TG_LW)\end{aligned} \tag{7}$$

where D is a diagonal matrix, its elements is the row (or column since S is symmetric) sum of the S, i.e. $D_{ii}=\sum_j S_{ij}$, $L=D-S$ is the Laplacian matrix. $G_L=XLX^T$ is the weighted similarity scatter matrix.

Generally we assume that two points are close to each other in space if they are similar to each other. On the contrary, if the two points are far away from each other, they convey more diversity information. The goal of minimizing Eq. (7) is to ensure that if two points are close to each other in original space they are also close to each other in feature space, i.e. preserve the similarity information as much as possible. However, it is apt to over fit the training data, which will degrade the performance of the algorithm. While preserving the diversity information can efficiently avoid the over-learning problem. Furthermore, to a point, the more distinctive information it conveys the more important to represent its pattern and keeping the representative during the projection is expected to make algorithm more robust and more discriminating. For the purpose of making full use the class information and better preserving the manifold structure of the data, we give the definition of discriminating diversity in Eq. (8)

$$B_{ij}=\begin{cases}\exp(-b/(\|X_i-X_j\|^2))(1-\exp(-b/(\|X_i-X_j\|^2))) & \begin{array}{l}\text{if } X_i \text{ is among } k1 \text{ nearest neighbors of } X_j\\ \text{or } X_j \text{ is among } k1 \text{ nearest neighbors of } X_i\\ \text{and } \tau_i=\tau_j\end{array}\\ \exp(-b/(\|X_i-X_j\|^2))(1+\exp(-b/(\|X_i-X_j\|^2))) & \begin{array}{l}\text{if } X_i \text{ is among } k1 \text{ nearest neighbors of } X_j\\ \text{or } X_j \text{ is among } k1 \text{ nearest neighbors of } X_j\\ \text{and } \tau_i\neq\tau_j\end{array}\\ 0 & \text{otherwise}\end{cases} \tag{8}$$

where $b\in(0,\ +\infty)$, and τ_i is the class label of X_i.

Fig. 1 is the plot of B_{ij} as a function of $b/(\|X_i-X_j\|^2)$, where $f1$, $f2$ and $f3$, respectively, denote the cases that (1) X_i is among $k1$ neighbors of X_j or X_j is among $k1$ neighbors of X_i, and they have same label; (2) X_i is among $k1$ neighbors of X_j or X_j is among $k1$ neighbors of X_i, and they have different labels; (3) the other case.

The discriminating diversity integrates both the local structure and the class information. By contrasting with the element of weighted matrix defined in Eq. (2), the discriminating diversity has such advantages as these:

Property 1. From Fig.1 we can clearly see that when the Euclidean distance is equal, the inter-diversity is larger than the intra-diversity. That means the points with larger weight will be more possible to be in different class. On the contrary, the points with small weight may have same labels. This is a good property for classification task.

Property 2. The discriminating diversity combines local structure and class information. What is more, the discriminating diversity can prevent the local neighborhood relationship from being forcefully distorted and the main geometric structure of the data can be largely preserved.

Property 3. Fig. 1 has shown that the inter-class discriminating diversity weight increases with the increase of the Euclidean distance. This could make the points in heterogeneous class are mapped far from each other in the feature space. The character of the intra-class diversity can prevent the data in homogeneous class from being mapped too far to each other. Thus the margin between different classes is larger than that in S-LSDP.This implies that the discriminating diversity can augment the margin between different classes. The obvious advantages of discriminating diversity make the proposed method more suitable to classification tasks.

Because of these advantages of discriminating diversity, LSDDP can be expected to have more discriminating power than S-LSDP.

For the purposes of preserving the diversity of the pattern, we impose the following objective:

$$\max \sum_{ij} \|Y_i - Y_j\|^2 B_{ij} \tag{9}$$

Maximizing Eq. (9) attempts to ensure that if two points are far away in original space they are also far way in the feature space, i.e. preserve the diversity information as much as possible.

Substituting $Y_i = W^T X_i$ into the Eq. (9), we can see that:

$$\begin{aligned}
&\sum_{ij} \|Y_i - Y_j\|^2 B_{ij} \\
&= \sum_{ij} \|W^T X_i - W^T X_j\|^2 B_{ij} \\
&= 2tr\left(\sum_{ij} W^T X_i B_{ij} X_i^T W - \sum_{ij} W^T X_i B_{ij} X_j^T W\right) \\
&= 2tr(W^T X \overline{D} X^T W - W^T X B X^T W) \\
&= 2tr(W^T X (\overline{D} - B) X^T W) \\
&= 2tr(W^T X \overline{L} X^T W) \\
&= 2tr(W^T G_N W)
\end{aligned} \tag{10}$$

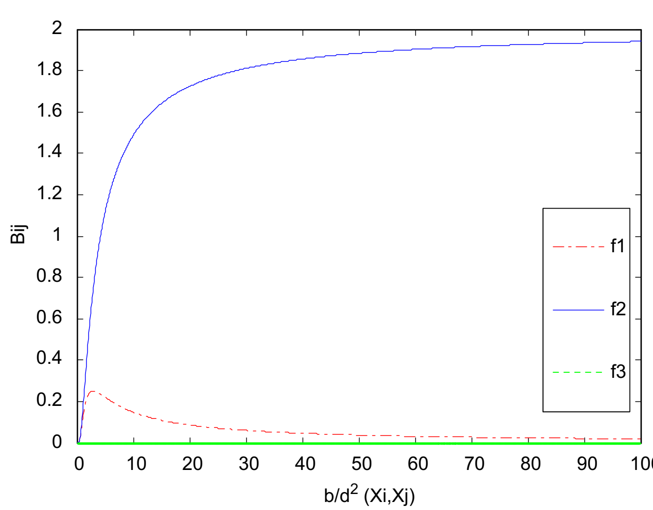

Fig. 1. Plot of B_{ij} as a function of $b/d^2(X_i, X_j)$, where $d^2(X_i, X_j)$ denotes $\|X_i - X_j\|^2$.

where $\overline{D}$ is a diagonal matrix, i.e. $\overline{D}_{ii} = \sum_j B_{ij}$. $\overline{L} = \overline{D} - B$ is the Laplacian matrix. $G_N = X\overline{L}X^T$ denotes the weighted diversity scatter matrix of the pattern.

The proposed method is attempt to learn a linear transformation that could deal with the similarity and diversity information simultaneously, that is to say the transformation should minimize Eq. (7) and maximize Eq. (9) at same the time. In S-LSDP the objective function is intend on minimizing the ratio of the similarity scatter matrix to the diversity scatter matrix. However, a ratio form often makes algorithms suffer from the SSS problem in many applications, especially in face recognition. In order to avoid the SSS problem, we change the objective function into a form of difference between the similarity scatter matrix and the diversity scatter matrix the in the proposed method, since the objective function $\min G_N^{-1} G_L$ and the objective function $\min (G_L - G_N)$ have the same motivation that it is to minimize G_L and maximize G_N at the same time. The objective function of the proposed method is expressed as following:

$$J = \min_{W^T W = I} (W^T G_L W - W^T G_N W) = \min_{W^T W = I} W^T (G_L - G_N) W \tag{11}$$

Eq. (11) can be solved by Lagrangian multiplier method

$$\frac{\partial}{\partial W} W^T (G_L - G_N) W - \lambda (W^T W - I) = 0 \tag{12}$$

where λ is the Lagrangian multiplier. Thus we can get

$$(G_L - G_N) W = \lambda W \tag{13}$$

Then we can obtain the optimal transformation matrix W, which consists of the eigenvectors corresponding to the first d non-zero smallest eigenvalues of $G_L - G_N$.

After obtaining the optimal transformation matrix, we use a classifier to classify a new sample. For a new sample X^* we can easily obtain its image using the transformation matrix, denoted by $Y^* = WX^*$.

To summarize, the classification has four steps as follows:

Step1: For each data point X_i, identify its k nearest neighbors and $k1$ nearest neighbors by KNN algorithm, then construct weighted adjacency similarity graph and weighted adjacency diversity graph and calculate the weights of the two graphs, respectively, based on Eqs. (5) and (8).

Step 2: Calculate the weighted similarity scatter matrix and the weighted diversity scatter matrix G_L and G_N.

Table 1
Average recognition accurate rate(%) and the corresponding reduced dimensions (shown in parentheses) for the five methods on Yale face database.

Method	PCA	LPP	LDP	SLSDP	LSDDP
4 train	68.95 (35)	78.67 (31)	90 (15)	75.90 (11)	**93.52** (33)
5 train	69.56 (34)	79.22 (45)	94.56 (38)	80.22 (13)	**95.44** (33)
6 train	70.13 (59)	80 (40)	96.13 (56)	83.33 (22)	**97.87** (31)

Fig. 2. Samples of one person from the Yale face database.

Step 3: Optimize the objective function in Eq. (11). This can be done by solving the eigenvalue problem in Eq. (13) and obtain the optimal transformation matrix W.

Step 4: Map the given query using the transformation matrix W and then predict its class label using a given classifier.

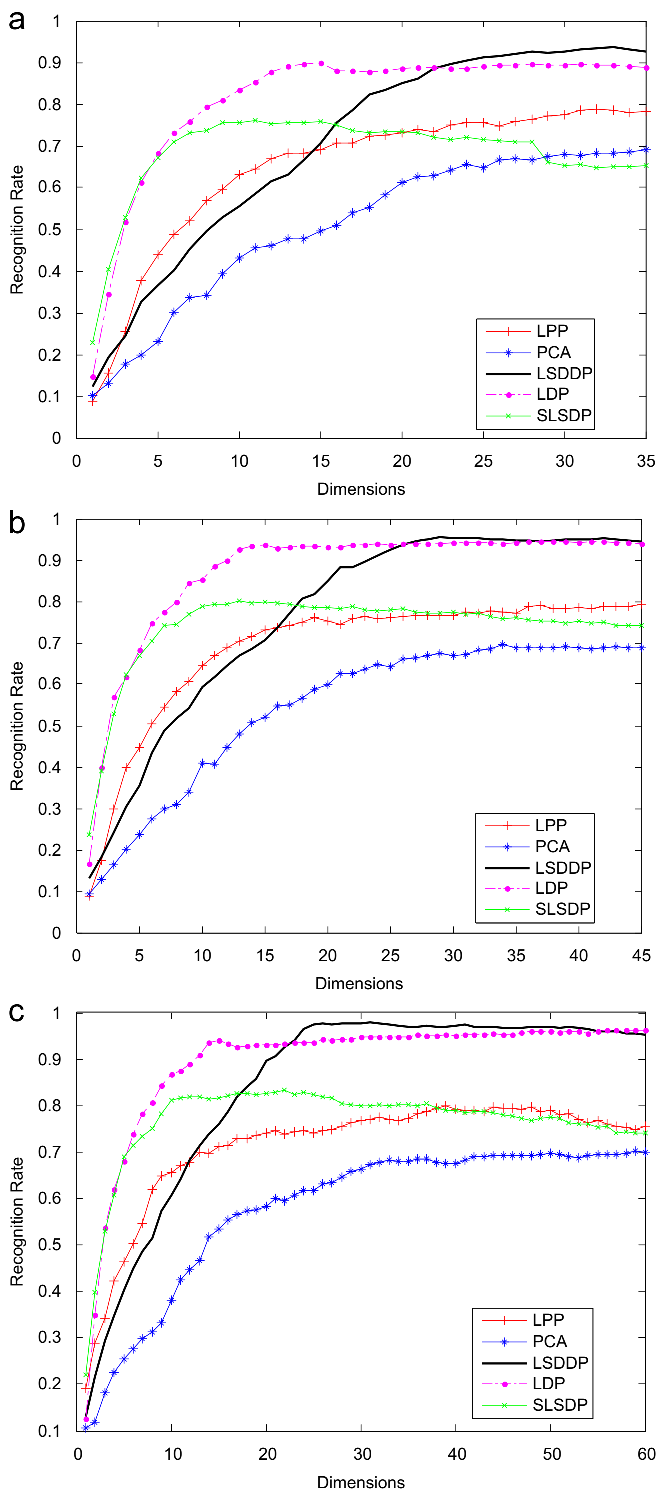

Fig. 3. Performance curves for the five methods on the Yale database. (a) Four images for training, (b) five images for training, (c) six images for training.

4. Experimental results

In this section we will carry out a set of experiments to show the effectiveness of our method for face and handwriting digits recognition. The Yale, ORL and CMU-PIE face databases and USPS handwriting digits are applied to evaluate the performance of the proposed method. In the experiments we also compared LSDDP with LPP, PCA, LDP and S-LSDP. What deserves mention is that LPP, LDP, S-LSDP and the proposed method all involve the parameter selection problem. In this paper the optimal parameters for the above mentioned methods are set by exhaustive search, since theoretically the parameter selection problem remains unsolved. Finally, in our experiment, the nearest neighbor classifier is employed to perform classification for its simplicity. The Euclidean metric is used as our distance measure. All the experiments were implemented by MATLAB 7.0

4.1. Experiments on Yale database

The Yale face database was sponsored by the Yale Center for Computer Vision and Control. The face database contains 165 grayscale images of 15 individuals (each person has 11 different images). The images demonstrate variations in lighting condition (left-light, center-lighter and right-light), facial expression (normal, sad, sleepy, surprised and wink), and with/without glasses. In our experiment, each image was resized to 32×32 pixel to save time. Fig. 2 shows the images of one person in Yale face database.

In the first experiment we randomly selected p images (p varying from 4 to 6) from each person as training set and the rest images were used as testing set. The experiment repeated 10 times. The best average recognition rate and the corresponding reduced dimension for the five methods are listed in Table 1. From Table 1 we can see that the maximal recognition rate of all the method increase with the increasing of the train number. Fig. 3 shows the average recognition rate curves for the five methods. Fig. 3 indicates that with the increase of the dimensions, our method outperformed others, while PCA is always performed the worst in all case. The main reason may be that our method can deal with the similarity and diversity information simultaneously. Furthermore, the strong structure preserving and discriminating ability make the proposed method more suitable for the recognition tasks. PCA is an unsupervised method and fail to explore the intrinsic structure of the data, so it always performed the worst.

Table 2
Average accurate rate(%) and the corresponding reduced dimensions (shown in parentheses) for the five methods on ORL face database.

Method	PCA	LPP	LDP	SLSDP	LSDDP
4 train	67.38 (100)	65.33 (35)	81.46 (46)	77.75 (52)	**90.50** (45)
5 train	70.95 (154)	70.45 (57)	85.00 (42)	80.30 (50)	**94.25** (29)
6 train	72.00 (59)	76.56 (51)	89.94 (43)	84.31 (40)	**96.94** (51)

Fig. 4. Samples of one person from the ORL face database.

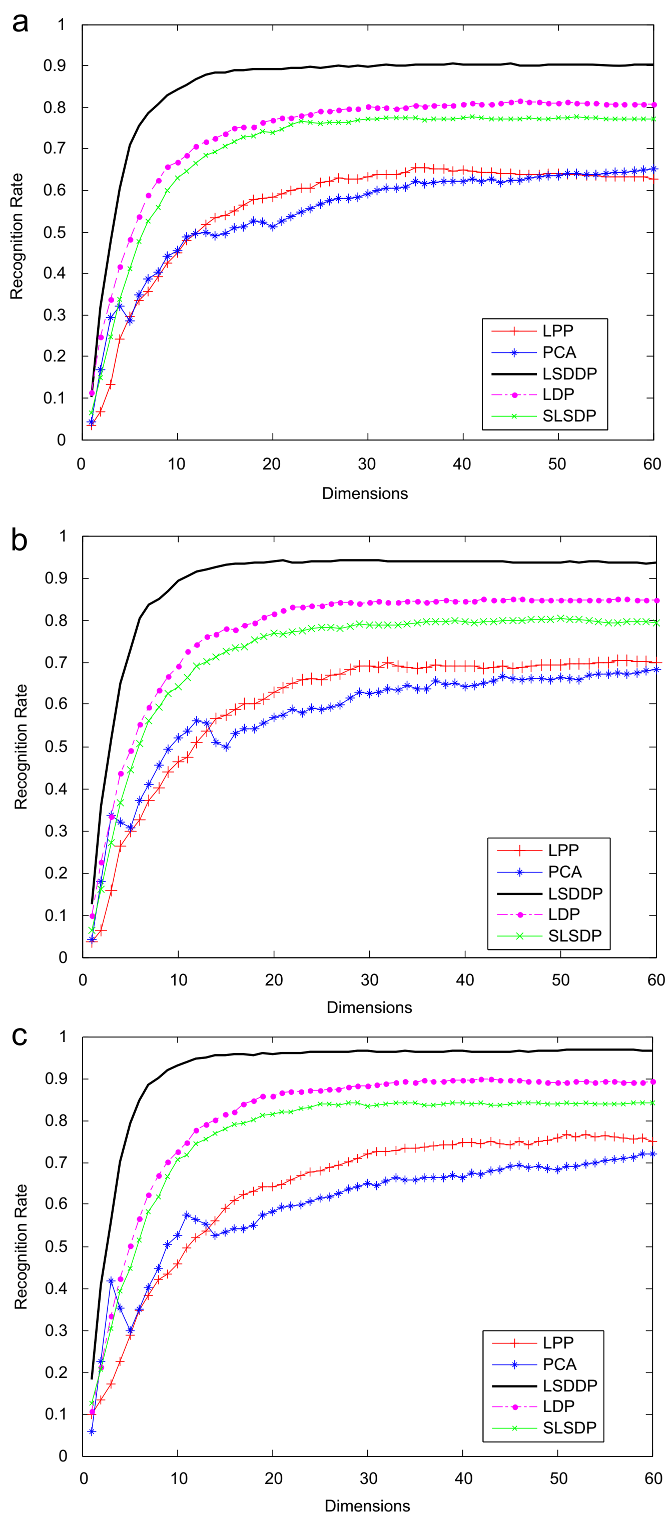

Fig. 5. Performance curves for the five methods on the ORL face database. (a) Four images for training, (b) five images for training, (c) six images for training.

4.2. Experiments on ORL database

In this subsection we verify the performance of LSDDP on the ORL face database. The ORL face database contains 400 face images of 40 distinct subjects. For some subjects, the images were taken at different times, varying the lighting, facial expressions (open/closed eyes, smiling/not smiling) and facial details (glasses/no glasses). All the images were taken against a dark homogeneous background with the subjects in an upright, frontal position (with tolerance for some side movement). All the images were in grayscale and resized to 32×32 pixel to save time. Fig. 4 shows the sample images of one person.

In this experiment we also randomly selected p images (p varying from 4 to 6) from each person as training set and the rest images as testing set. The experiment repeated 10 times independent. The best average recognition rate and the corresponding reduced dimension for the five methods are listed in Table 2. The performance curves for the five methods are shown in Fig. 5. From Table 2 and Fig. 5 we can see that the supervised methods are more powerful than unsupervised ones. As discussed above, the good property of the discriminating similarity and discriminating diversity make the proposed method more suitable for the recognition than other methods.

4.3. Experiment on CMU-PIE face database

The CMU-PIE face database includes 68 subjects with 41,368 face images as a whole. We chose the Pose29 subset for our experiment. The subset contains 1632 grayscale images of 68 individuals (each person has 24 different images). The images demonstrate variations in lighting condition and facial expression. All the images were in grayscale and resized to 64×64 pixel to save time. Fig. 6 shows the sample images of one person.

In the experiment we randomly choose 12 images from each individual for training and the left 12 images for testing. The best mean recognition rates for 10 times are listed in Table 3. The recognition rate of each method over the variation of the dimensions is plotted in Fig. 7. As we can see, in spite of the variation on the lighting conditions and poses, our algorithm outperforms other competitors. One possible reason is that LSDDP can capture the local discriminating diversity and the local structure information very well, thus our algorithm does not suffer from the over fitting and local structure distorted problem. S-LSDP also could deal with the similarity and diversity information at the same time, but it fails to preserve the neighborhood relationship of the data.

Table 3
Average recognition rata and the corresponding reduced dimensions (shown in parentheses) for the five methods on CMU-PIE face database.

Method	PCA	LPP	LDP	S-LSDP	LSDDP
Recognition rate(%)	71.91	83.91	88.69	89.67	**92.75**
Dimensions	197	159	157	101	**91**

Fig. 6. Sample images of one person from CMU-PIE face database.

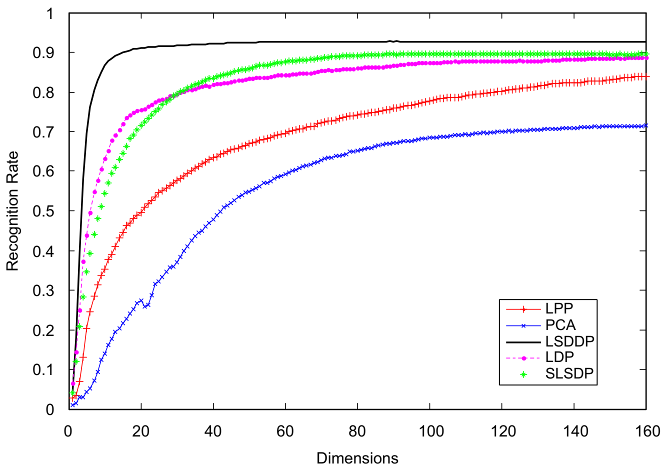

Fig. 7. Performance curves for the five methods on the CMU-PIE face database.

Fig. 8. Sample digital images "0" from USPS handwriting database.

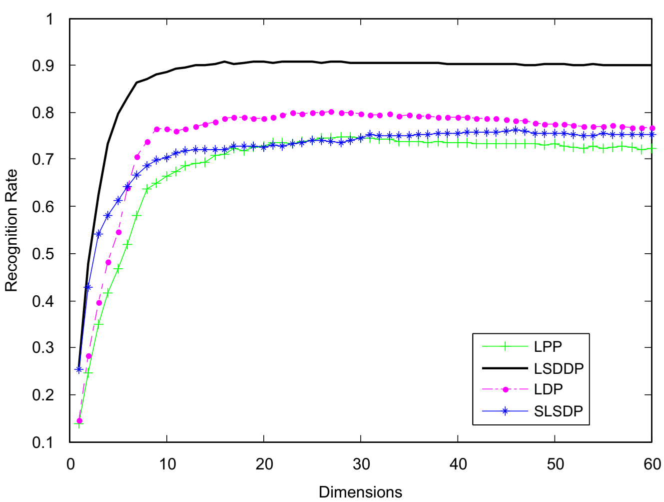

Fig. 9. Performance curves for the four methods on the USPS database.

4.4. *Experiment on USPS handwriting digits database*

The USPS handwriting digital data include 10 classes from "0" to "9". Each class has 1100 examples. In our experiment, we random select 100 images from each class. 50 images are used to training and the left are used to test. Each image is transformed to a vector with 256 dimensions. Fig. 8 displays a subset of digital "0" from original USPS handwriting digital database.

In this experiment, we directly applied LPP, LDP and S-LSDP to the data without taking a PCA preprocessing, and compare with our method. The experiment repeated 10 times. Fig. 9 shows the best recognition rate curves for the four methods. The best average recognition rate and the corresponding reduced dimension for the four methods are listed in Table 4.

From all the figures and tables we can clearly see that the proposed method is always outperformed than other methods. This is because our method has more manifold structure preserving power as well as more discriminating power.

Table 4
Average recognition accurate rate and the corresponding reduced dimensions for the four methods on USPS database.

Method	LPP	LDP	SLSDP	LSDDP
Recognition rate(%)	74.74	80.18	76.16	**90.68**
Dimensions	28	27	46	**24**

Table 5
Computational time (repeated 10 echos independent) on Yale and ORL.

Method	Yale (s)			ORL (s)		
	4 train	5 train	6 train	4 train	5 train	6 train
PCA	12.00	13.05	13.09	55.31	58.51	57.82
LPP	8.72	12.50	12.74	56.07	58.20	60.14
LDP	9.22	12.85	13.68	58.15	65.14	66.30
S-LSDP	5.93	10.88	13.51	55.41	54.69	58.91
LSDDP	135.88	144.66	161.23	141.44	143.55	145.22

Besides the recognition rates, the time complexity is also important for real applications. We list the computational time of all the methods obtained from Yale and ORL datasets in Table 5. From Table 5 we can see that the drawback of LSDDP is much time consuming than the other method. The *no free lunch theorem* tell us there is lack of inherent superiority of any classifier, in other words, LSDDP sacrificed the computation efficiency for the superior generalization performance.

5. Conclusion

In this paper, we presented a novel algorithm which is based on the spectral graph, namely LSDDP, for dimensionality reduction. Two contributions were made in this paper. (1) Combined with local structure information, the class information is introduced to redefine the weights of the similarity graph and the diversity graph, then LSDDP can prevents the main geometric structure of the data and has more discriminating power. These merits make our method more robust and suitable for classification tasks. (2) Instead of the generalized eigen-decomposition, the optimized transformation matrix can be computed by solving an eigen-equation in our method. Thus LSDDP does not suffer from the SSS problem. The new technique was applied to the Yale, ORL and CMU-PIE face datasets as well as the USPS handwriting digits database. The experimental results show that the proposed method has the distinctly effective generalization performance in classification. However, our method is much more time-consuming than other methods, how to improve is our future work.

Acknowledgment

The authors want to thank the reviewers for their helpful comments, and thank Professor Songcan Chen for his suggestions to improve the paper. This work is supported in part by the NSF of China (Nos. 61170040, 60903088, 60903089), and by the Scientific Research Foundation of Education Department of Hebei Province (Nos. 2009312, 2009410).

References

[1] S.C. Yan, D. Xu, B.Y. Zhang, et al., Graph embedding and extension: a general framework for dimensionality reduction, IEEE Trans. Pattern Anal. Mach. Intell. 29 (1) (2007) 40–51.

[2] S.T. Roweis, L.K. Saul, Nonlinear dimensionality reduction by locally linear embedding, Science 290 (5500) (2000) 2323–2326.
[3] J.B. Tenenbaum, V. de Silva, J.C. Langford, A global geometric framework for nonlinear dimensionality reduction, Science 290 (5500) (2000) 2319–2323.
[4] Q. He, Z.X. Xie, Q.H. Hu, et al., Neighborhood based sample and feature selection for SVM classification learning, Neurocomputing 74 (2011) 1585–1594.
[5] Y. Xu, F.X. Song, G. Feng, Y.N. Zhao, A novel local preserving projection scheme for use with face recognition, Expert Syst. Appl. 37 (2010) 6718–6721.
[6] F. Zang, J.S. Zhang, Discriminative learning by sparse representation for classification, Neurocomputing 74 (2011) 2179–2183.
[7] M.A. Turk, A.P. Pentland, Face recognition using eigenfaces, in: Proceedings of the Conference on Computer Vision and Pattern Recognition. Hawaii, USA, 1991, pp. 586–591.
[8] S. Mikat, G. Fitsch. Fisher discriminant analysis with kernels, in: Proceedings of the 1999 IEEE Signal Processing Society Workshop, 1999, pp. 41–48.
[9] X.F. He, D. Cai, S.C. Yan, et al., Neighborhood preserving embedding, in: Proceedings of the Tenth IEEE International Conference on Computer Vision2 (ICCV 2005), 2005, pp. 1208–1213.
[10] M. Belkin, P. Niyogi, Laplacian eigenmaps and spectral techniques for embedding and clustering, Neural Inf. Process. Syst. 14 (2002) 585–591.
[11] M. Belkin, P. Niyogi, Laplacian eigenmaps for dimensionality reduction and data representation, Neural Comput. 15 (6) (2003) 1373–1396.
[12] D. Cai, X. He, J.W. Han., et al., Orthogonal Laplacianfaces for face recognition, IEEE Trans. Image Process. 15 (11) (2006) 3609–3614.
[13] J. Yang, D. Zhang, J.Y. Yang, et al., Globally maximizing, locally minimizing: unsupervised discriminant projection with application to face and palm biometrics, IEEE Trans. Pattern Anal. Mach. Intell. 29 (4) (2007) 650–664.
[14] Y. Chen, X.H. Xu, J.H. Lai, Optimal locality preserving projection for face recognition, Neurocomputing 74 (2011) 3941–3945.
[15] L.S. Qiao, L.M. Zhang, S.C. Chen, An empirical study of two typical locality preserving linear discriminant analysis methods, Neurocomputing 73 (10–12) (2010) 1587–1594.
[16] B. Yang, S.C. Chen, Sample-dependent graph construction with application to dimensionality reduction, Neurocomputing 74 (1–3) (2010) 301–314.
[17] X.F. He, S.C. Yan, Y.X. Hu., et al., Face recognition using Laplacianfaces, IEEE Trans. Pattern Anal. Mach. Intell. 27 (3) (2005) 328–340.
[18] H.T. Zhao, S.Y. Sun, Z.L. Jing, et al., Local structure based on supervised feature extraction, Pattern Recognition 39 (8) (2006) 1546–1550.
[19] L.P. Yang, W.G. Gong, X.H. Gu, et al., Null space discriminat locality preserving projections, Neurocomputing 71 (16–18) (2008) 3644–3649.
[20] W.W. Yu, X.L. Teng, C.Q. Liu, Face recognition using discriminant locality preserving proections, Image Vision Comput. 24 (3) (2006) 239–248.
[21] L. Zhu, S. Zhu, Face recognition based on orthogonal discriminant locality preserving projections, Neurocomputing 70 (9) (2007) 1543–1546.
[22] B. Li, C. Wang, D.S. Huang, Supervised feature extraction based on orthogonal discriminant projection, Neurocomputing 73 (1–3) (2009) 191–196.
[23] J. Gui, W. Jia, L. Zhu, et al., Locality preserving discriminant projection for face and palmprint recognition, Neurocomputing 73 (13–15) (2010) 2696–2707.
[24] Q.X. Gao, D.Y. Xie, H. Xiu, et al., Supervised feature extraction based on information fusion of local structure and diversity information, Acta Autom. Sin. 36 (2010) 1108–1114.

Qiang Hua. He received the B.Sc. and M.Sc. degrees from Hebei University, Baoding, China, in 1996 and 1999, respectively. He is currently a candidate doctor with Nanjing University of Aeronautics & Astronautics, Nanjing, China, and he is also working as an associate professor with Hebei University. His main research interests are focused on pattern recognition and neural networks.

Lijie Bai. Received her B.Sc degree in mathematical from Hengshui Collage, Hengshui, China, in 2009. From 2009 to now, she is pursuing her M.Sc degree in Hebei university, China. Her main research interest is subspace learning.

Xizhao Wang. Ph. D. supervisor. He is a professor of the College of Mathematics and Computer Science, Hebei University, China. He received his Ph.D. degree in computer science from Harbin Institute of Technology in 1998. He is an editor-in-chief for the International Journal of Machine Learning and Cybernetics, and associated editor for Information Sciences, IEEE Transactions on SMC Part (B), and International Journal of Pattern Recognition and Artificial Intelligence. His main research interests include machine learning and computational intelligence, neural networks, fuzzy measures and fuzzy integral.

Yuchao Liu. Received his B.Sc degree in mathematical from Handan Collage, Handan, China, in 2008. From 2009 to now, he is pursuing his M.Sc degree in Hebei university, China. His main research interest is subspace learning.