

Research Article

An Efficient Feature Extraction Method, Global Between Maximum and Local Within Minimum, and Its Applications

Lei Wang,^{1,2} Jiangshe Zhang,^{1,2} and Fei Zang^{1,2}

¹ School of Science, Xi'an Jiaotong University, Xi'an 710049, China

² State Key Laboratory for Manufacturing Systems Engineering, Xi'an Jiaotong University, Xi'an 710049, China

Correspondence should be addressed to Lei Wang, wangleisxcc@gmail.com

Received 28 March 2011; Revised 16 April 2011; Accepted 18 April 2011

Academic Editor: Jyh Horng Chou

Copyright © 2011 Lei Wang et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Feature extraction plays an important role in preprocessing procedure in dealing with small sample size problems. Considering the fact that LDA, LPP, and many other existing methods are confined to one case of the data set. To solve this problem, we propose an efficient method in this paper, named global between maximum and local within minimum. It not only considers the global structure of the data set, but also makes the best of the local geometry of the data set through dividing the data set into four domains. This method preserves relations of the nearest neighborhood, as well as demonstrates an excellent performance in classification. Superiority of the proposed method in this paper is manifested in many experiments on data visualization, face representative, and face recognition.

1. Introduction

Nowadays with the continual development of information technology, the amount of data has largely expanded, such as in the domain of pattern recognition, artificial intelligence, and computer vision. Because the dimension of the samples of data set is a lot greater than the number of the obtained samples of data set, it results in "the curse of dimensionality" [1]. Feature extraction method plays an important role in dealing with small sample size (SSS) problems. It represents original high dimensional data in the low-dimensional space through capturing some important data structure and information and is a common preprocessing procedure in multivariate statistical data analysis. At present, feature extraction methods have successfully been applied in many domains such as text classification [2], remote sensing image analysis [3], microarray data analysis [4], and face recognition [5, 6].

Generally speaking, feature extraction methods can be divided into the following three kinds with regard to applied tools. Firstly, it is based on algebra properties and takes advantage of generalized eigen decomposition in matrix theory and extracts some features that contained more discriminative information through discovering algebra structure of samples of data set. Secondly, on the basis of geometry characteristics, it aims to find optimal projection directions in some cases. Thirdly, in the process of dealing with SSS problems, traditional statistical methods, based on the number of the samples of data set face some challenges, thus, we must take advantage of intersectant knowledge from statistical inference and information science to obtain informative features we expect.

From the viewpoint of the data structure, feature extraction methods have two ways: globally and locally. The former concerns that the data set itself or the class in the data set is regarded as the whole and extracts the important features wholly of the corresponding data set. There are some typical methods such as principle component analysis (PCA) [7, 8], linear discriminant analysis (LDA) [7, 9], and maximum margin criterion (MMC) [10]. Although intended results could be got in practice, those methods do not make use of the inner structure of data set. Meanwhile the latter makes use of neighborhood structure of each sample in the data set and maintains the local information hidden in the extracted features. Locality preserving projections (LPP) [11, 12], neighborhood preserving embedding (NPE) [13], and average neighborhood margin maximum (ANMM) [14] suppose the neighborhood of the sample lies in the submanifold space of the data set and preserve such property. DLPP [15], PRLPP [16], and ILPP [17] are based on local information which be proposed to address the small sample size (SSS) problem. In [18], an entropy regularization term is incorporated into the objective function for controlling the uniformity level of the edge weights in graph. DLPP/MMC [19] seeks to maximize the difference between local class based on maximum margin criterion. Those methods only use the local geometry or global information and do not find entirely the intrinsical structure of the training set.

In this paper, we combine global character with local structure of the data set and in order to perform linear projection and propose a new feature extraction method called global between maximum, meanwhile, local within minimum (GBMLWM). As for a fixed sample from the data set, the others are divided into four parts: the same class with the fixed sample within or without its nearest neighborhood and the different class from the fixed sample within or without its nearest neighborhood, see Figure 1. In the course of local within minimum, we make use of three domains: domain I, domain II, and domain III. And in the process of global between maximum, similarly to LDA, our methods by maximum between-class scatter. So it not only overcomes the disadvantages of the global method, but also takes full advantage of the local method. It is worthwhile highlighting some properties of GBMLWM algorithm from some perspectives as the followings:

- (1) GBMLWM method shares excellent properties with LDA and MMC. In this paper, we maintain global merit in the process of global between maximum. Similar to LDA, we first keep all samples in the data set away from the class centroid, and then let the samples, labeled the same class with the fixed sample and beyond its nearest neighborhood close to its class centroid. So, GBMLWM is a way to supervise, and it is feasible to take apart between class of the data set and keep close within class of the data set.
- (2) For making use of local structure of the data set, GBMLWM inherits several important advantages of local methods, that is, LPP. In this paper, we maintain the neighborhood of the sample in the data set and keep the samples labeled the

same class with the fixed sample in its nearest neighborhood approach. At the same time, let the samples different from the fixed sample away from it. So, GBMLWM maintains the submanifold space of the fixed sample.

- (3) As connection to PCA, LDA, MMC, LPP, and ANMM, we could derive those methods from GBMLWM framework by imposing some conditions, that is to say, those methods are the special case of GBMLWM. Visual and classification experiments have also indicated that proposed method in the paper is superior to the above methods.

The rest of this paper is organized as follows. Section 2 briefly reviews global and local methods, that is, PCA, LDA, MMC, LPP, and ANMM. The GBMLWM algorithm is put forward in Section 3, and its relationship with the above methods is also discussed in this section. The experimental outcomes are presented in Section 4. The conclusion appears in the Section 5.

2. Brief Review of Global and Local Methods

Suppose that $\mathbf{X} = [\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n] \in \mathbf{R}^{m \times n}$ is a set of m -dimensional samples of size n , and it is composed of C_i , $i = 1, \dots, C$, where each class contains n_i samples, $\sum_{i=1}^C n_i = n$, and let \mathbf{x}_j^i a m -dimension column vector which denotes the j th sample from the i th class. Generally speaking, the aim of the linear feature extraction or dimensionality reduction is to find an optimal linear transformation $\mathbf{W} \in \mathbf{R}^{m \times d}$ ($d \ll m$) from the original high-dimensional space to the goal low-dimensional space $\mathbf{y}_i = \mathbf{W}^T \mathbf{x}_i$, so that those transformed data in terms of different optimal criteria best represent different information such as that of algebra and geometry structure.

2.1. Principle Component Analysis

PCA attempts to seek an optimal projection direction so that covariance of the data set is maximized, or average cost of projection is minimized after transformation. The objective function of PCA is defined as follows:

$$\max_{\mathbf{W}} \sum_{i=1}^n (\mathbf{y}_i - \mathbf{m}_y)^2, \quad (2.1)$$

where $\mathbf{y}_i = \mathbf{W}^T \mathbf{x}_i$, \mathbf{m}_y is the mean of $\{\mathbf{y}_i\}_{i=1}^n$. Applying algebra knowledge, (2.1) may be rewritten as

$$\max_{\mathbf{W}} \mathbf{W}^T \mathbf{S}_t \mathbf{W}, \quad (2.2)$$

where

$$\mathbf{S}_t = \frac{1}{n} \sum_{i=1}^n (\mathbf{x}_i - \mathbf{m}_x)(\mathbf{x}_i - \mathbf{m}_x)^T \quad (2.3)$$

is the sample covariance matrix. \mathbf{m}_x is the mean of the all samples. The optimal $\mathbf{W} = [\mathbf{w}_1, \mathbf{w}_2, \dots, \mathbf{w}_d]$ is the eigenvectors of \mathbf{S}_t corresponding to the first d largest eigenvalues.

2.2. Linear Discriminant Analysis

The purpose of LDA is to discriminate and classify, and it seeks an optimal discriminative subspace by maximizing between-class scatter matrices, meanwhile, minimizing within-class scatter matrices. LDA's objective is to find a set of vectors \mathbf{W} according to

$$\mathbf{W} = \arg \max_{\mathbf{W}} \frac{|\mathbf{W}^T \mathbf{S}_b \mathbf{W}|}{|\mathbf{W}^T \mathbf{S}_w \mathbf{W}|}, \quad (2.4)$$

where

$$\begin{aligned} \mathbf{S}_b &= \frac{1}{n} \sum_{i=1}^C n_i (\mathbf{m}_x^i - \mathbf{m}_x) (\mathbf{m}_x^i - \mathbf{m}_x)^T, \\ \mathbf{S}_w &= \frac{1}{n} \sum_{i=1}^C \sum_{j=1}^{n_i} (\mathbf{x}_j^i - \mathbf{m}_x^i) (\mathbf{x}_j^i - \mathbf{m}_x^i)^T, \end{aligned} \quad (2.5)$$

respectively, represent the between-class scatter matrix and the within-class scatter matrix. \mathbf{m}_x^i is the mean of the i th class. The projection directions \mathbf{W} are the generalized eigenvectors $\mathbf{w}_1, \mathbf{w}_2, \dots$, and \mathbf{w}_d solving $\mathbf{S}_b \mathbf{w} = \lambda \mathbf{S}_w \mathbf{w}$ associated with the first d largest eigenvalues.

2.3. Maximum Margin Criterion

MMC keeps similarity or dissimilarity information of the high-dimensional space as much as possible after dimensionality reduction by employing the overall variance and measuring the average margin between different classes. MMC's projection directions matrix is as follows

$$\mathbf{W} = \arg \max_{\mathbf{W}} (\mathbf{W}^T (\mathbf{S}_b - \mathbf{S}_w) \mathbf{W}), \quad (2.6)$$

where $\mathbf{S}_b, \mathbf{S}_w$ are defined as (2.5).

2.4. Locality Preserving Projection

PCA, LDA and MMC aim to preserve global structure of the data set, while LPP is to preserve the local structure of the data set. LPP models the local submanifold structure by maintaining the neighborhood relations of the fore and aft transformed samples in data set. With the same mathematical notations as above, the objective function of LPP is defined as follows:

$$\min_{\mathbf{W}} \frac{\mathbf{W}^T \mathbf{X} \mathbf{L} \mathbf{X}^T \mathbf{W}}{\mathbf{W}^T \mathbf{X} \mathbf{D} \mathbf{X}^T \mathbf{W}}, \quad (2.7)$$

where \mathbf{D} is a diagonal matrix, that is, $\mathbf{D}(i, i) = \sum_j \mathbf{S}\mathbf{L}_{ij}$, $i = 1, \dots, n$, $\mathbf{L} = \mathbf{D} - \mathbf{S}\mathbf{L}$ is the Laplacian matrix. And $\mathbf{S}\mathbf{L} = (\mathbf{S}\mathbf{L}_{ij})_{n \times n}$ is a similarity matrix, defined as follows:

$$\mathbf{S}\mathbf{L}(i, j) = \begin{cases} \mathbf{S}(i, j), & \mathbf{x}_i \in \mathbf{N}_j \text{ or } \mathbf{x}_j \in \mathbf{N}_i \\ 0, & \text{otherwise,} \end{cases} \quad (2.8)$$

where $\mathbf{S}(i, j) = \exp(-\|\mathbf{x}_i - \mathbf{x}_j\|^2/t)$, for $i, j = 1, \dots, n$, t is a kernel parameter, \mathbf{N}_i is the set of nearest neighborhood of \mathbf{x}_i . The optimal \mathbf{W} is given by the d eigenvectors corresponding to minimum eigenvalue solution to the following generalized eigenvalue problem:

$$\mathbf{X}\mathbf{L}\mathbf{X}^T \mathbf{w} = \lambda \mathbf{X}\mathbf{D}\mathbf{X}^T \mathbf{w}. \quad (2.9)$$

2.5. Average Neighborhood Margin Maximum

Different from PCA and LDA, ANMM aims to obtain effective discriminating information by using average local neighborhood margin maximum. For each sample, ANMM aims at pulling the neighborhood samples with the same label towards it as near as possible, meanwhile, pushing the neighborhood samples with different labels away from it as far as possible. ANMM's solutions as follows:

$$\mathbf{W} = \arg \max_{\mathbf{W}} [\mathbf{W}^T (\mathbf{A} - \mathbf{B}) \mathbf{W}], \quad (2.10)$$

where \mathbf{A} is called the scatterness matrix, \mathbf{B} is called the compactness matrix

$$\begin{aligned} \mathbf{A} &= \sum_{i,j} \frac{(\mathbf{x}_i - \mathbf{x}_j)(\mathbf{x}_i - \mathbf{x}_j)^T}{|\mathbf{N}_i^e|}, \\ \mathbf{B} &= \sum_{i,j} \frac{(\mathbf{x}_i - \mathbf{x}_j)(\mathbf{x}_i - \mathbf{x}_j)^T}{|\mathbf{N}_i^o|}, \end{aligned} \quad (2.11)$$

and \mathbf{N}_i^e , \mathbf{N}_i^o , respectively, is ξ the nearest heterogenous and homogenous neighborhood of the \mathbf{x}_i , $|\cdot|$ is the cardinality of a set. Here, we can regard ANMM as the local version of the MMC.

3. Global between Maximum and Local Within Minimum

In this section, we present our algorithm—global between maximum, simultaneously local within minimum (GBMLWM). It profits from global and local methods. GBMLWM algorithm preserves not only the local neighborhood of submanifold structure, but also the global information of the data set. To state our proposed algorithm, we first give four domains about

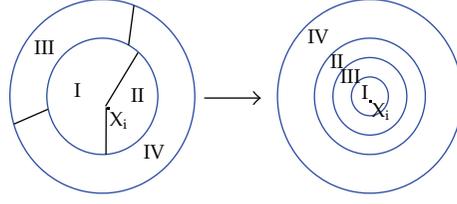


Figure 1: Here are four domains into which other samples except x_i in the data set are divided. The left figure shows four domains in the original high-dimensional space, and the right depicts four domains in the low-dimensional space.

x_i as follows:

Domain I: those samples are a subset of the nearest neighborhood of x_i and labeled the same class with x_i .

Domain II: those samples are also a subset of the nearest neighborhood of x_i , but labeled the different class from x_i .

Domain III: those samples labeled the same class with x_i , but do not lie in the nearest neighborhood of x_i .

Domain IV: those samples do not lie in the nearest neighborhood of x_i and also are labeled as the different class from x_i .

Figure 1 shows us an intuition about the above four domains. The nearest neighborhood of x_i consists of domain I and II. The samples labeled the same class with x_i lie in domain I and III, and the samples labeled the different class from x_i lie in II and IV.

3.1. Global between Maximum

The purpose of classification and feature extraction is to make the samples labeled as different class apart from each other. We first operate those points in domain II and IV via maximizing global and local between-class scatter. That is to say, our aim is not only to make the data globally separable, but also to maximize the distance between different classes in the nearest neighborhood. Thus, our objective functions are defined as follows:

$$\begin{aligned}
 J_4(\mathbf{W}) &= \frac{1}{n} \sum_{i=1}^C n_i (\mathbf{m}_y^i - \mathbf{m}_y) (\mathbf{m}_y^i - \mathbf{m}_y)^T \\
 &= \mathbf{W}^T \left(\frac{1}{n} \sum_{i=1}^C n_i (\mathbf{m}_x^i - \mathbf{m}_x) (\mathbf{m}_x^i - \mathbf{m}_x)^T \right) \mathbf{W} \\
 &= \mathbf{W}^T \mathbf{S}_t \mathbf{W}.
 \end{aligned} \tag{3.1}$$

3.2. Local Within Minimum

As for classification, maximizing between class is not adequate, and compacting within-class scatter is also required. So, we now make the samples from domain I close to x_i itself, the

samples from II away from \mathbf{x}_i and the samples from domain II close to their own class centroid.

$$\begin{aligned}
\mathbf{J}_1(\mathbf{W}) &= \sum_{i,j} (\mathbf{y}_i - \mathbf{y}_j)^2 \mathbf{S}_{ij} \\
&= \mathbf{W}^T \left(\sum_{i,j} (\mathbf{x}_i - \mathbf{x}_j)(\mathbf{x}_i - \mathbf{x}_j)^T \mathbf{S}_{ij} \right) \mathbf{W} \\
&= \mathbf{W}^T \mathbf{X} \mathbf{L}_1 \mathbf{X}^T \mathbf{W} \\
\mathbf{J}_2(\mathbf{W}) &= \sum_{i,j} (\mathbf{y}_i - \mathbf{y}_j)^2 \mathbf{S}_{ij}, \\
&= \mathbf{W}^T \left(\sum_{i,j} (\mathbf{x}_i - \mathbf{x}_j)(\mathbf{x}_i - \mathbf{x}_j)^T \mathbf{S}_{ij} \right) \mathbf{W} \\
&= \mathbf{W}^T \mathbf{X} \mathbf{L}_2 \mathbf{X}^T \mathbf{W} \\
\mathbf{J}_3(\mathbf{W}) &= \sum_{i=1}^C \sum_{\mathbf{y}_j \in C_i \cap \text{III}} (\mathbf{y}_j - \mathbf{m}_y^i)(\mathbf{y}_j - \mathbf{m}_y^i)^T, \\
&= \mathbf{W}^T \left(\sum_{i=1}^C \sum_{\mathbf{x}_j \in C_i \cap \text{III}} (\mathbf{x}_j - \mathbf{m}_x^i)(\mathbf{x}_j - \mathbf{m}_x^i)^T \right) \mathbf{W} \\
&= \mathbf{W}^T \mathbf{S} \mathbf{L}_w \mathbf{X}^T \mathbf{W},
\end{aligned} \tag{3.2}$$

where

$$\mathbf{L}_1 = \mathbf{D}_1 - \mathbf{S}_1 \tag{3.3}$$

and $\mathbf{D}_1(i, i) = \sum_j \mathbf{S}_1(i, j)$, $\mathbf{S}_1(i, j) = \mathbf{S}(i, j)$ if $\mathbf{x}_i, \mathbf{x}_j$ lie in corresponding to oneself domain I; otherwise, $\mathbf{S}_1(i, j) = 0$

$$\mathbf{L}_2 = \mathbf{D}_2 - \mathbf{S}_2 \tag{3.4}$$

and $\mathbf{D}_2(i, i) = \sum_j \mathbf{S}_2(i, j)$, $\mathbf{S}_2(i, j) = \mathbf{S}(i, j)$ if $\mathbf{x}_i, \mathbf{x}_j$ lie in domain II, otherwise, $\mathbf{S}_2(i, j) = 0$.

$$\mathbf{S} \mathbf{L}_w = \sum_{i=1}^C \sum_{\mathbf{x}_j \in C_i \cap \text{III}} (\mathbf{x}_j - \mathbf{m}_x^i)(\mathbf{x}_j - \mathbf{m}_x^i)^T. \tag{3.5}$$

3.3. GBMLBM Algorithm

In the previous description, nearest neighborhood of \mathbf{x}_i is indicated as K nearest neighborhood based on Euclidean distance between two samples from the data set. Our objective function is defined as follows:

$$\begin{aligned} J(\mathbf{W}) &= (J_2(\mathbf{W}) + J_4(\mathbf{W})) - (J_1(\mathbf{W}) + J_3(\mathbf{W})) \\ &= \mathbf{W}^T \left[(\mathbf{S}_t - \mathbf{S}L_w) - (\mathbf{X}L_2\mathbf{X}^T - \mathbf{X}L_1\mathbf{X}^T) \right] \mathbf{W} \\ &= \mathbf{W}^T \mathbf{M} \mathbf{W}, \end{aligned} \quad (3.6)$$

where

$$\mathbf{M} = (\mathbf{S}_t - \mathbf{S}L_w) - (\mathbf{X}L_2\mathbf{X}^T - \mathbf{X}L_1\mathbf{X}^T). \quad (3.7)$$

And then our optimal projection directions \mathbf{W} are solutions to the following optimization problem:

$$\begin{aligned} \max_{\mathbf{W}} J(\mathbf{W}), \\ \mathbf{W}^T \mathbf{W} = \mathbf{I}. \end{aligned} \quad (3.8)$$

So, $\mathbf{W} = [\mathbf{w}_1, \dots, \mathbf{w}_d]$ is the eigenvectors of $\mathbf{M}\mathbf{w} = \lambda\mathbf{w}$ corresponding to the first d largest eigenvalues. It is obvious that the GBMLWM algorithm is fairly straightforward instead of computing inverse matrix, and thus it absolutely avoids the SSS problem. Now, the algorithm procedure of GBMLBM is formally summarized as follows:

- (1) as for each sample $\mathbf{x}_i, i = 1, \dots, n$, dividing the samples from data set except $\mathbf{x}_i, i = 1, \dots, n$ into four domains: I, II, III, and VI;
- (2) computing $\mathbf{S}_t, \mathbf{L}_1, \mathbf{L}_2, \mathbf{S}L_w$, according to (2.3), (3.3), (3.4), and (3.5), respectively;
- (3) and then, we can obtain matrix \mathbf{M} according to (3.7);
- (4) computing the generalized eigenvectors of $\mathbf{M}\mathbf{w} = \lambda\mathbf{w}$, and the optimal projection matrix $\mathbf{W} = [\mathbf{w}_1, \dots, \mathbf{w}_d]$ corresponding to the d largest eigenvalues, where d is the rank of matrix \mathbf{M} . For a testing sample \mathbf{x} , its image in the lower dimensional space is given by

$$\mathbf{x} \mapsto \mathbf{y} = \mathbf{W}^T \mathbf{x}. \quad (3.9)$$

3.4. Discussion

Here, we find those methods limited to global structure or local geometry of the data set are special case of GBMLWM algorithm. PCA regards the data set as a whole domain and demands all the samples away from the total mean of the data set. Thus, we see that PCA is an unsupervised version special case of GBMLWM algorithm. Both MMC and LDA divide

Table 1: Estimate approximately computational complexity for the six different algorithms, where n denotes the total number of training samples, k and C are the size of neighborhood and the number of class, respectively.

Method	PCA	LDA	MMC	LPP	ANMM	GBMLWM
Computational Complexity	$O(n)$	$O(n)$	$O(n)$	$O(n^2 + kn)$	$O(n^2 + kn)$	$O(n^2 + kn + kC)$

samples except \mathbf{x}_i in the data set into two domains: one is composed of the samples labeled as the same class with \mathbf{x}_i , called within-class \mathbf{S}_w ; the other contains the samples labeled the different class from the \mathbf{x}_i , called between-class \mathbf{S}_b . They, respectively, correspond to the domains $I \cup III$ and $II \cup VI$, as illustrated in Figure 1. The local methods, such as LPP and ANMM, are different from the above methods based on global structure. LPP and ANMM divide the whole data set into two domains according to the nearest neighborhood of \mathbf{x}_i . LPP is operated in $I \cup II$, while ANMM method in the domain I and II , as depicted in Figure 1. Those local methods do not utilize the global information of the data set and are local version special case of the algorithm proposed in this paper. The superiority of the GBMLWM algorithm is manifested in the data experiments in the following section.

Training cost is the amount of computations required to find the optimal projection vectors and the sample feature vectors of the training set for comparison. We compare the training cost of the methods based on their computational complexities. Here, we suppose that each class has the same number of training samples. If we regard each column vector as a computational cell and do not consider the computational complexity of eigen-analysis, we estimate approximately computational complexity for six different algorithms which include based-local methods and based-global techniques. Table 1 gives the analysis of computational complexity for the six different algorithms. From Table 1, we can see that our method has the largest training cost. However, in practice, the size of neighborhood and the number of class are often not large enough to cause much more computation of our algorithm. The computational complexity of GBMLWM also shows that our algorithm not only considers the global information, but also utilizes the local geometry. That makes our algorithm efficiently reflect the intrinsic structure of the training set. The following experimental results also manifest this point.

4. Experiments

In this section, we will carry out several experiments to show the effectiveness of the proposed GBWMLW method for data visualization, face representative, and recognition. Here, we will compare the global methods, that is, PCA, LDA, MMC, and local methods, that is, LPP, NPE, and ANMM, with our proposed method on the following four databases: MNIST digit, Yale, ORL, and UMIST database. In the processing of the PCA, we only maintain the $N - C$ dimensions to ensure scatter matrix nonsingular. In testing phases, the size of neighborhood k is determined by 5-fold cross validation in all experiments, and the nearest neighbor (NN) rule is used in classification. In using the LPP and GBMLWM algorithms, the weight of two samples is computed with Gaussian kernel, and the kernel parameter is selected as follows: we firstly compute the pairwise distance among all the training samples, then, t is made equal to the half median of those pairwise distance.

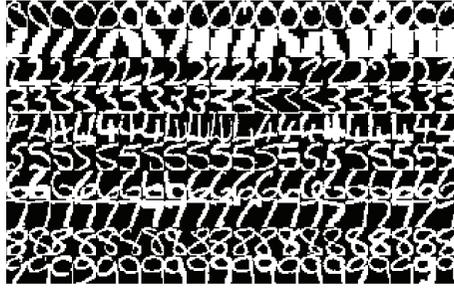


Figure 2: All the samples of handwritten digits from number 0 to 9 used in our data visualization experiment.

4.1. Data Visualization

In this subsection, we first use a publicly available handwritten digits to illustrate data visualization. MNIST database [20] has 10 digits, and each digit contains 39 samples. The number of total samples is 390 which each image has the size 20×16 . Here, we only select 20 samples from each digit. So the size of the training set is 320×200 , and each image is represented lexicographically as a high-dimensional vector of the length 320. Figure 2 shows all the samples of the ten digits. For visualization, we project the data set in 2-D space by all seven subspace learning methods. And the experiment results are depicted in Figure 3. With the exception of LDA and GBMLWM, the samples from the different digits seem to heavily overlap. Compared with GBMLWM algorithm, LDA makes the samples from the same class become a point. Although this phenomenon is helpful for classification, it has a poor generalization ability since it does not exhibit the case in each object oneself. GBMLWM algorithm not only separates each digit, but also shows what is hidden in each digit. When the number of the nearest neighbor of x_i reduces from $K = 15$ to $K = 2$, the samples from the same object become more and more compacted. That also verifies that LDA is a special case of GBMLWM.

4.2. Yale Database

This experiment aims to demonstrate the ability of capturing the important information on Yale face database [21], called face representative. The Yale face database contains 165 gray scale images of 15 individuals. There are 11 images per subject, one per different facial expression or configuration: center-light, with/without glasses, happy, left/right light, normal, sad, sleepy, surprised and wink. All images from the Yale database were cropped and the cropped images normalized to the 32×32 pixels with 256 gray level per pixel. Some samples from the Yale database are shown in Figure 4. Here, the training set is composed of all the samples from this database. And the most significant 10 eigenfaces obtained from the Yale face database through using the seven subspace learning methods are shown in Figure 5. From the Figure 5, we obviously see that our algorithm captures more basic information of the face than other methods.

4.3. UMIST Database

The UMIST database [22] contains 564 images of 20 individuals, each covering a range of poses from profile to frontal views. Subjects cover a range of race, sex, and appearance.

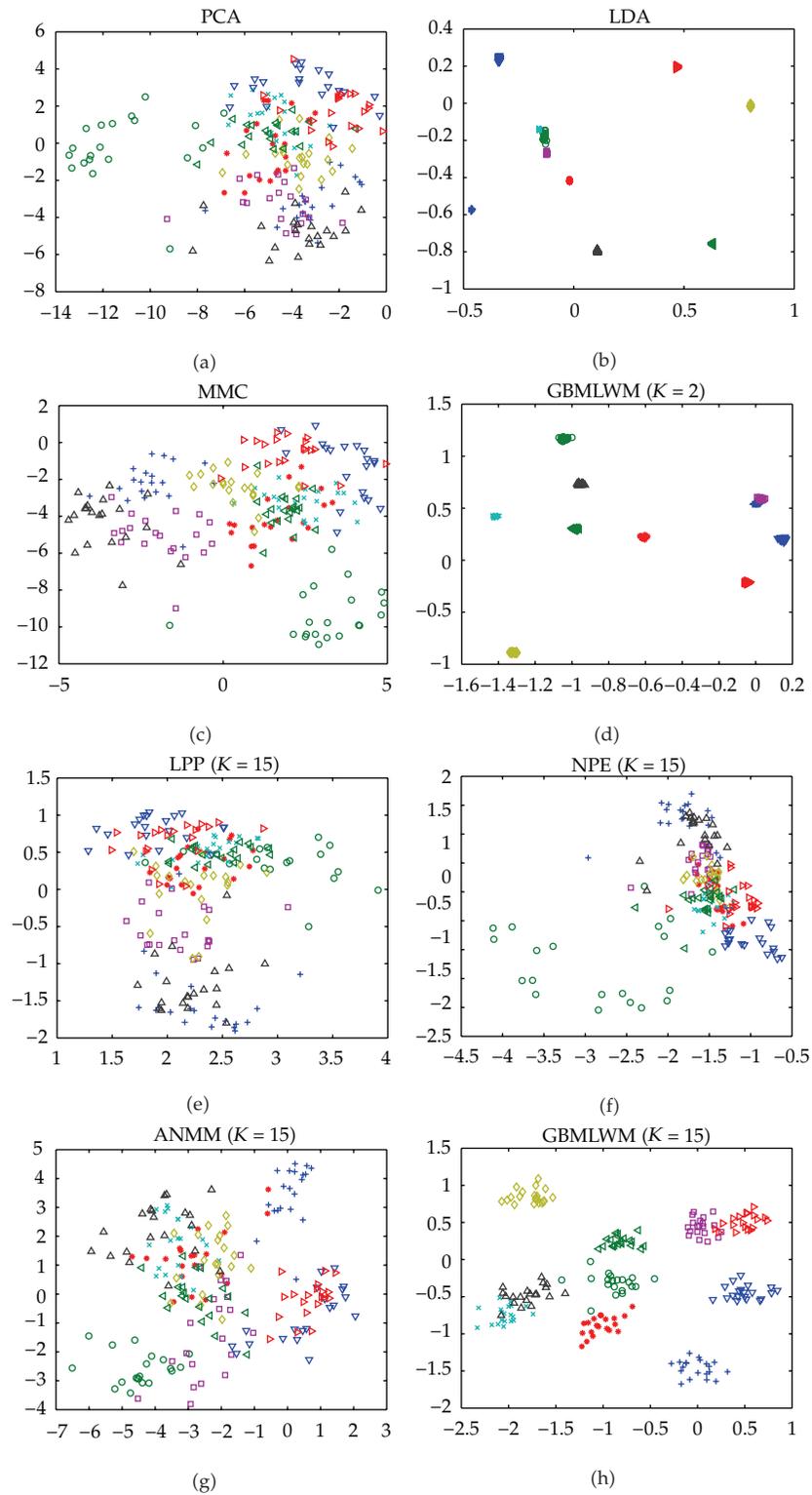


Figure 3: Two-dimensional projections of the handwritten digits, respectively, by using seven related subspace learning methods. “+” denotes 0, “o” denotes 1, “•” denotes 2, “x” denotes 3, “□” denotes 4, “◊” denotes 5, “Δ” denotes 6, “▽” denotes 7, “◁” denotes 8, “▷” denotes 9.



Figure 4: Some face samples from the Yale database.



Figure 5: The most significant 10 eigenfaces obtained from the Yale face database through using the seven subspace learning methods: PCA, LDA, MMC, LPP, NPE, ANMM, and GBMLWM from top to bottom.

We use a cropped version of the UMIST database that is publicly available at S. Roweis' Web page. All the cropped images normalized to the 64×64 pixels with 256 gray level per pixel. Figure 6 shows some images of an individual. We randomly select three, four, five, and six images of each individual for training, and the rest for testing. We repeat these trails ten times and compute the average results. The maximal average recognition rates of seven subspace learning methods are presented in Table 2. From Table 2, we find that GBMLWM algorithm's highest accuracy, respectively, are 79.88%, 86.10%, 91.85%, and 93.80% on the different training sets and corresponding testing sets. The improvements are significant. Furthermore, the dimensions of the four GBMLWM subspaces corresponding to the maximal recognition rates are remarkably low, and they are 15, 13, 11, and 18, respectively.

4.4. ORL Database

In the ORL face database [23], there are 40 distinct subjects, each of which contains ten different images. So there are 400 images in all. For some subjects, the images are taken at different times, varying the lighting, facial expressions and facial details. All the images are taken against a dark homogeneous background with the subjects in an upright, frontal position. All images from the ORL database are cropped, and the cropped images normalized to the 32×32 pixels with 256 gray level per pixel. Same samples from this database are showed in Figure 7. In this experiment, four training sets, respectively, correspond to the numbers of samples from each subject three, four, five, and six. And other samples, respectively, form the testing sets. We repeat these trails ten times and compute the average results. The recognition rates versus the reduced dimensions are shown in Figure 8. The best average recognition rates of seven subspace learning methods are presented in Table 3. It can be seen that GBMLWM algorithm's recognition rates remarkably outperform the other methods in all the four training subsets with the highest accuracy of 90.07%, 94.67%, 97.20%, and 97.31%, respectively. The standard deviations of the GBMLWM corresponding to the best results are 0.03, 0.02, 0.02, and 0.02.



Figure 6: Some face samples from the Yale database.



Figure 7: Some face samples from the ORL database.

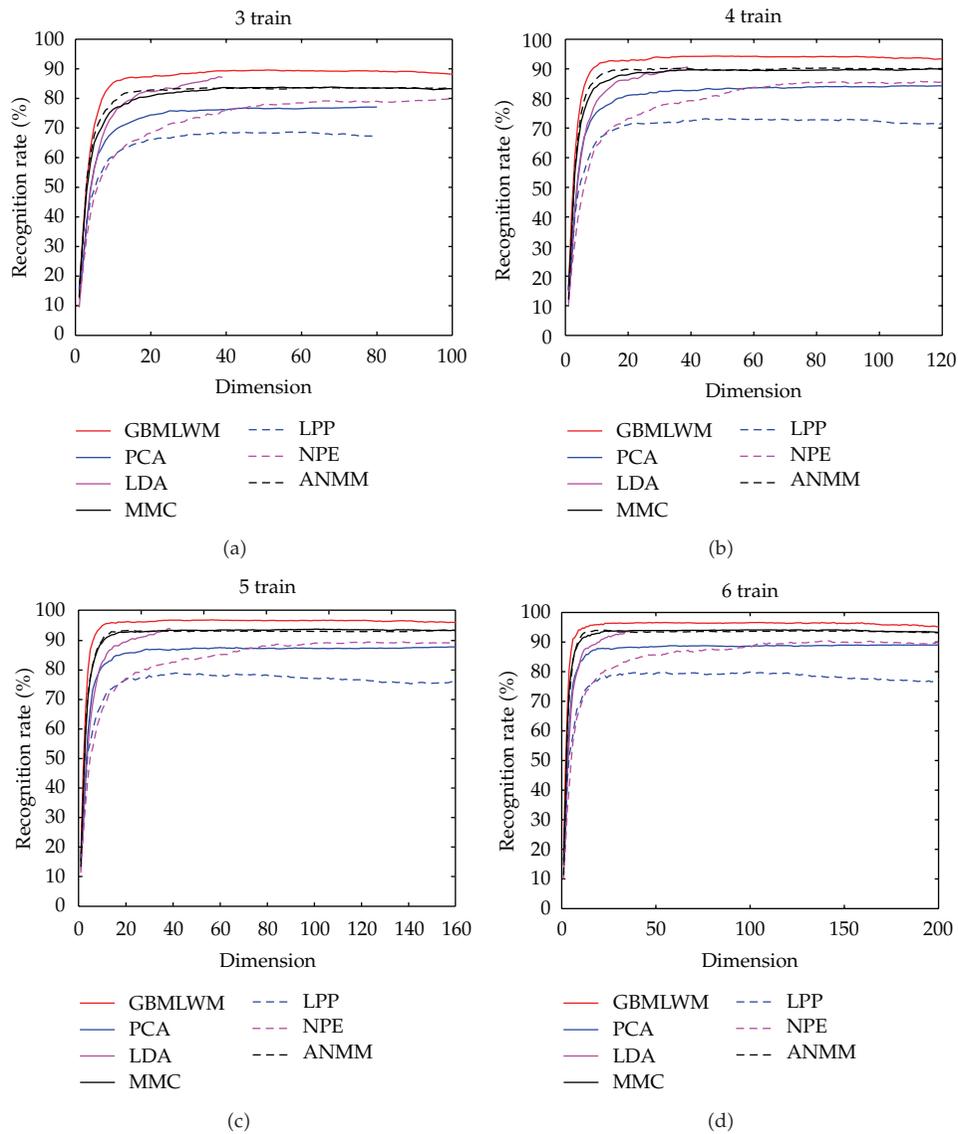


Figure 8: Average recognition rates of seven subspace learning methods and the different samples from each object versus the reduced dimensions on the ORL database.

Table 2: Recognition accuracy (%) of different algorithms and the numbers in the bracket corresponding to dimensions on the UMIST database.

Size	PCA	LDA	MMC	LPP	NPE	ANMM	GBMLWM
3	77.26 (50)	77.55 (17)	80.06 (44)	71.61 (32)	70.60 (38)	77.90 (17)	79.88 (15)
4	76.08 (67)	82.73 (19)	83.94 (67)	75.43 (50)	75.74 (60)	81.60 (22)	86.10 (13)
5	82.69 (87)	87.73 (19)	90.59 (77)	81.28 (77)	81.43 (80)	88.48 (20)	91.85 (11)
6	85.72 (68)	89.52 (13)	92.81 (96)	85.74 (77)	85.82 (100)	91.91 (13)	93.80 (18)

Table 3: The best recognition accuracy (%) of different algorithms and the numbers in the bracket corresponding to standard deviation on the ORL database.

Size	PCA	LDA	MMC	LPP	NPE	ANMM	GBMLWM
3	77.43 (0.03)	87.57 (0.03)	84.61 (0.02)	70.04 (0.04)	80.86 (0.03)	84.54 (0.03)	90.07 (0.03)
4	84.71 (0.02)	91.08 (0.02)	90.96 (0.02)	74.33 (0.03)	87.04 (0.02)	91.25 (0.02)	94.64 (0.02)
5	88.20 (0.02)	93.95 (0.02)	93.95 (0.02)	80.15 (0.03)	90.45 (0.02)	94.15 (0.02)	97.20 (0.02)
6	89.44 (0.03)	94.37 (0.02)	94.56 (0.02)	81.69 (0.03)	91.56 (0.03)	95.12 (0.03)	97.31 (0.02)

5. Conclusions

In this paper, we have proposed a new linear projection method, called GBMLWM. It is an efficient linear subspace learning method with the supervised and unsupervised character. Similar to PCA, LDA, and MMC, we consider the global character of the data set. At the same time, similar to LPP, NPE, and ANMM, we also make the best use of the local geometry structure of the data set. We have pointed out that the existing linear subspace learning methods are a special case of our GBMLWM algorithm. A large number of experiments demonstrate that the method which we propose is obviously superior to other existing methods, such as LDA and LPP.

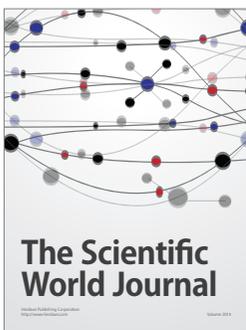
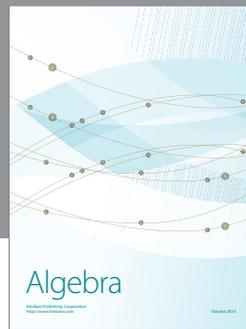
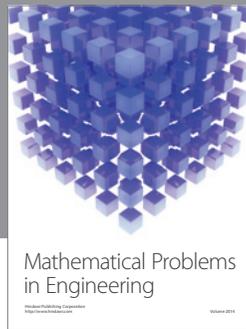
Acknowledgments

The authors would like to express their gratitude to the anonymous referees as well as the Editor and Associate Editor for their valuable comments which lead to substantial improvements of the paper. This work was supported by the National Basic Research Program of China (973 Program) (Grant No. 2007CB311002), the National Natural Science Foundation of China (Grant Nos. 60675013, 61075006), the Research Fund for the Doctoral Program of Higher Education of China (No. 20100201120048).

References

- [1] V. Vapnik, *The Nature of Statistical Learning Theory*, Statistics for Engineering and Information Science, Springer, New York, NY, USA, Second edition, 2000.
- [2] S. Chakrabarti, S. Roy, and M. Soundalgekar, "Fast and accurate text classification via multiple linear discriminant projections," in *Proceedings of the 28th international conference on Very Large Data Bases*, 2002.
- [3] B. Mojaradi, H. Moghaddam, M. Valadan Zoej, and R. Duin, "Dimensionality reduction of hyperspectral data via spectral feature extraction," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 47, no. 7, pp. 2091–2105, 2009.

- [4] M. Zhu and A. Martinez, "Using the information embedded in the testing sample to break the limits caused by the small sample size in microarray-based classification," *BMC Bioinformatics*, vol. 9, article 280, 2008.
- [5] M. Turk and A. Pentland, "Eigenfaces for recognition," *Journal of Cognitive Neuroscience*, vol. 3, no. 1, pp. 71–86, 1991.
- [6] Z. Li, D. Lin, and X. Tang, "Nonparametric discriminant analysis for face recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 31, no. 4, pp. 755–761, 2009.
- [7] R. O. Duda, P. E. Hart, and D. G. Stork, *Pattern Classification*, Wiley-Interscience, New York, NY, USA, 2nd edition, 2001.
- [8] I. T. Jolliffe, *Principal Component Analysis*, Springer Series in Statistics, Springer, New York, NY, USA, 2nd edition, 2002.
- [9] K. Fukunaga, *Introduction to Statistical Pattern Recognition*, Computer Science and Scientific Computing, Academic Press, Boston, Mass, USA, 2nd edition, 1990.
- [10] H. Li, T. Jiang, and K. Zhang, "Efficient and robust feature extraction by maximum margin criterion," *IEEE Transactions on Neural Networks*, vol. 17, no. 1, pp. 157–165, 2006.
- [11] X. He, P. Yan, and Y. Niyogi, "Locality preserving projections," in *Proceedings of the Conference on Advances in Neural Information Processing Systems*, vol. 1, 2003.
- [12] X. He, S. Yan, Y. Hu, P. Niyogi, and H. J. Zhang, "Face recognition using Laplacianfaces," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 27, no. 3, pp. 328–340, 2005.
- [13] X. He, D. Cai, S. Yan, and H. J. Zhang, "Neighborhood preserving embedding," in *Proceedings of the IEEE International Conference on Computer Vision (ICCV '05)*, vol. 2, pp. 1208–1213, 2005.
- [14] F. Wang, X. Wang, D. Zhang, C. Zhang, and T. Li, "MarginFace: a novel face recognition method by average neighborhood margin maximization," *Pattern Recognition*, vol. 42, no. 11, pp. 2863–2875, 2009.
- [15] J. Chen, B. Li, and B. Yuan, "Face recognition using direct LPP algorithm," in *Proceedings of the 9th International Conference on Signal Processing (ICSP '08)*, pp. 1457–1460, 2008.
- [16] J. Lu and Y. P. Tan, "Regularized locality preserving projections and its extensions for face recognition," *IEEE Transactions on Systems, Man, and Cybernetics, Part B*, vol. 40, no. 3, pp. 958–963, 2010.
- [17] Y. Xu, A. Zhong, J. Yang, and D. Zhang, "LPP solution schemes for use with face recognition," *Pattern Recognition*, vol. 43, no. 12, pp. 4165–4176, 2010.
- [18] L. Zhang, L. Qiao, and S. Chen, "Graph-optimized locality preserving projections," *Pattern Recognition*, vol. 43, no. 6, pp. 1993–2002, 2010.
- [19] G.-F. Lu, Z. Lin, and Z. Jin, "Face recognition using discriminant locality preserving projections based on maximum margin criterion," *Pattern Recognition*, vol. 43, no. 10, pp. 3572–3579, 2010.
- [20] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," *Proceedings of the IEEE*, vol. 86, no. 11, pp. 2278–2324, 1998.
- [21] P. N. Belhumeur, J. P. Hespanha, and D. J. Kriegman, "Eigenfaces vs. fisherfaces: recognition using class specific linear projection," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 19, no. 7, pp. 711–720, 1997.
- [22] D. Graham and N. Allinson, "Characterizing virtual eigensignatures for general purpose face recognition," in *Face Recognition: From Theory to Applications*, vol. 163, pp. 116–456, 1998.
- [23] F. Samaria and A. Harter, "Parameterisation of a stochastic model for human face identification," in *Proceedings of the IEEE Workshop on Applications of Computer Vision*, pp. 138–142, Sarasota, Fla, USA, December 1994.



Hindawi

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

