Orthogonal Locality Minimizing Globality Maximizing Projections for Feature Extraction

Feiping Nie,* Shiming Xiang, Yangqiu Song, and Changshui Zhang

Department of Automation,

Tsinghua University,

Beijing 100084, China

(Dated: November 21, 2008)

Abstract

Locality Preserving Projections(LPP) is a recently developed linear feature extraction algorithm, which has been frequently used in the task of face recognition and other applications. However, LPP does not satisfy the shift invariance property, which should be satisfied by a linear feature extraction algorithm. In this paper, we analyze the reason and derive the shift invariant LPP algorithm. Based on the analysis on the geometrical meaning of the shift invariant LPP algorithm, we propose two novel algorithms to minimize the locality and maximize the globality under an orthogonal projection matrix. Experimental results on face recognition are presented to demonstrate the effectiveness of the proposed algorithms.

Keywords: Locality Preserving Projections; subspace learning; shift invariant; feature extraction; face recognition.

PACS numbers:

*Electronic address: `e-mail:feipingnie@gmail.com`

I. INTRODUCTION

Linear feature extraction and dimensionality reduction techniques are very important approaches to deal with high-dimensional data, such as texts, images and videos. In the past decades, many supervised algorithms have been proposed for the purpose of classification. Linear Discriminant Analysis(LDA) [1] is one of the most popular ones. It has been successfully applied in many classification tasks such as face recognition. Recently, an algorithm called Locality Preserving Projections(LPP) [2] is developed, and is becoming a frequently used linear feature extraction technique for many applications.

For a linear feature extraction algorithm, data shifted by an arbitrary constant vector should not influence the learned projection matrix. However, the original LPP algorithm proposed in [2] does not satisfy this shift invariance property. We analyze the reason from the derivation of LPP. LPP is derived from Laplacian eigenmap [3], while in Laplacian eigenmap, there is an implicit relationship which is not considered in the derivation of the original LPP algorithm. Taking the relationship into account, we derive the shift invariant LPP algorithm that satisfies the shift invariance property.

The geometrical meaning of the derived shift invariant LPP is also revealed in this paper. We reveal that it is to minimize the sum of the Euclidean distances between data pairs which are local to each other, while the weighted covariance matrix of data is fixed to a constant matrix. Yang et al. recently proposed a unsupervised discriminant projection algorithm [4], which is to solve the following optimization problem

$$\mathbf{W} = \arg \min_{\mathbf{W}^T\mathbf{X}\mathbf{L}_t\mathbf{X}^T\mathbf{W}=\mathbf{I}} tr(\mathbf{W}^T\mathbf{X}\mathbf{L}\mathbf{X}^T\mathbf{W}) \tag{1}$$

where $\mathbf{X}$ is the data matrix, $\mathbf{L}$ is a Laplacian matrix and $\mathbf{L}_t = \mathbf{I} - \frac{1}{n}\mathbf{1}\mathbf{1}^T$ is the centralization matrix. The algorithm is said to has extended LPP to obtain a globally maximizing and locally minimizing projection. In fact, from the geometrical meaning of the shift invariant LPP we know that, although the shift invariant LPP is derived from the motivation of locality preserving, it preserves the local and global property simultaneously. We will see in Section III that the optimization problem (1) is just a simplified case of the optimization problem (11) in the shift invariant LPP. Similar to LPP, the learned projection matrix is not orthogonal.

Recently, orthogonal method has been attracting great interests as the orthogonality is of desirable property and often demonstrates good performance empirically [5–9]. An orthogonal LPP algorithm is proposed in [10], which is to solve the following optimization problem

$$\mathbf{W} = \arg\min_{\mathbf{W}^T\mathbf{W}=\mathbf{I}} tr(\mathbf{W}^T\mathbf{X}\mathbf{L}\mathbf{X}^T\mathbf{W}) \tag{2}$$

From the optimization problem we can see that the algorithm only performs a locality minimizing projection and does not take the global information into account, which might be not sufficient to gain the discriminative power.

Cai et al. also propose another orthogonal LPP algorithm recently [11]. They use a step-by-step procedure to obtain a set of orthogonal projections $\{\boldsymbol{w}_1, \boldsymbol{w}_2, ..., \boldsymbol{w}_m\}$. After calculating the first $k-1$ projections $\{\boldsymbol{w}_1, \boldsymbol{w}_2, ..., \boldsymbol{w}_{k-1}\}$, the k-th projection $\boldsymbol{w}_k$ is calculated by solving the following optimization problem

$$\boldsymbol{w}_k = \arg\min_{\mathbf{W}_{k-1}^T\boldsymbol{w}_k=\mathbf{0}} \frac{\boldsymbol{w}_k^T\mathbf{X}\mathbf{L}\mathbf{X}^T\boldsymbol{w}_k}{\boldsymbol{w}_k^T\mathbf{X}\mathbf{D}\mathbf{X}^T\boldsymbol{w}_k} \tag{3}$$

where $\mathbf{D}$ is the degree matrix on graph and $\mathbf{W}_{k-1} = [\boldsymbol{w}_1, \boldsymbol{w}_2, ..., \boldsymbol{w}_{k-1}]$. The step-by-step procedure makes the algorithm computationally more expensive and makes the objective w.r.t $\mathbf{W}$ to be optimized not clear. The shift invariance property is also not taken into account in their algorithm.

Inspired by the geometrical meaning of the shift invariant LPP, in this paper we propose two novel algorithms to minimize the locality and maximize the globality simultaneously under an orthogonal projection matrix. Experiments on face recognition are presented and the experimental results demonstrate that the proposed algorithms are effective for feature extraction.

The rest of this paper is organized as follows: In Section 2, we revisit the derivation of LPP, and the shift invariant LPP is derived in Section 3. In Section 4, we reveal the geometrical meaning of the derived shift invariant LPP. In Section 5, we propose two novel algorithms to perform locality minimizing globality maximizing projections. In Section 6, we present the experiments on face recognition to verify the effectiveness of the proposed algorithms. Finally, we conclude this paper in Section 7.

II. LOCALITY PRESERVING PROJECTIONS REVISITED

LPP is a linear feature extraction algorithm [2] derived from Laplacian eigenmap [3]. Given n data points $\boldsymbol{x}_i \in \mathbb{R}^d (i = 1, 2, ..., n)$, in order to discover the corresponding $\boldsymbol{y}_i \in$

$\mathbb{R}^m$ in the low-dimensional manifold for $\boldsymbol{x}_i$, Laplacian eigenmap is to solve the following optimization problem:

$$\mathbf{Y} = \arg \min_{\mathbf{YDY}^T=\mathbf{I}} tr(\mathbf{YLY}^T) \tag{4}$$

where $\mathbf{Y} = [\boldsymbol{y}_1, \boldsymbol{y}_2, ..., \boldsymbol{y}_n] \in \mathbb{R}^{m\times n}$, $\mathbf{D}$ is a diagonal matrix, $\mathbf{D}_{ii} = \sum_j \mathbf{A}_{ij}$ and $\mathbf{L} = \mathbf{D} - \mathbf{A}$ is a Laplacian matrix defined on a graph constructed by the given data. The affinity matrix $\mathbf{A}$ could be defined by

$$\mathbf{A}_{ij} = \begin{cases} e^{-\frac{\|\boldsymbol{x}_i - \boldsymbol{x}_j\|^2}{t}} & \boldsymbol{x}_i \textit{ and } \boldsymbol{x}_j \textit{ are neighbors} \\ \\ 0 & \textit{otherwise} \end{cases} \tag{5}$$

where t is the parameter of the heat kernel, and $\|\cdot\|$ denotes the 2-norm of vector, i.e., $\|\boldsymbol{x}\|^2 = \boldsymbol{x}^T\boldsymbol{x}$.

The map from $\boldsymbol{x} \in \mathbb{R}^d$ to $\boldsymbol{y} \in \mathbb{R}^m$ learned by Laplacian eigenmap is nonlinear, the aim of LPP is to find a linear map to approximate this nonlinear map. Denote data matrix $\mathbf{X} = [\boldsymbol{x}_1, \boldsymbol{x}_2, ..., \boldsymbol{x}_n] \in \mathbb{R}^{d\times n}$, LPP assumes the map from $\boldsymbol{x} \in \mathbb{R}^d$ to $\boldsymbol{y} \in \mathbb{R}^m$ is linear, and let

$$\mathbf{Y} = \mathbf{W}^T\mathbf{X} \tag{6}$$

where $\mathbf{W} = [\boldsymbol{w}_1, \boldsymbol{w}_2, ..., \boldsymbol{w}_m] \in \mathbb{R}^{d\times m}$ is a projection matrix.

Imposing the linear relationship (6) on the optimization problem (4), LPP is to solve another optimization problem as follows:

$$\mathbf{W} = \arg \min_{\mathbf{W}^T\mathbf{XDX}^T\mathbf{W}=\mathbf{I}} tr(\mathbf{W}^T\mathbf{XLX}^T\mathbf{W}) \tag{7}$$

where $tr(\cdot)$ denotes the trace operator of matrix.

The solution to this optimization problem is finally reduced to solving the following eigen-decomposition problem.

$$\mathbf{XLX}^T\mathbf{W} = \mathbf{XDX}^T\mathbf{W}\boldsymbol{\Lambda} \tag{8}$$

where $\boldsymbol{\Lambda}$ is the eigenvalue matrix and $\mathbf{W}$ is the corresponding eigenvector matrix of $(\mathbf{XDX}^T)^{-1}\mathbf{XLX}^T$.

III. SHIFT INVARIANT LPP

For a linear feature extraction algorithm, data shifted with an arbitrary constant vector should not influence the result of the learned projection matrix $\mathbf{W}$. That is to say, when data are shifted by $\tilde{\boldsymbol{x}} = \boldsymbol{x} - \boldsymbol{c}$, where $\boldsymbol{c}$ is an arbitrary constant vector, the learned projection matrix $\mathbf{W}$ should not be changed. The shift invarint property can be well understood. When the given data shifted with an arbitrary constant vector, the Euclidean distances between the data pairs are unchanged, and thus the structure of the data is unchanged. therefore, the learned projection matrix should also remain unchanged.

We will see that two popular linear feature extraction methods, Principal Component Analysis(PCA) and Linear Discriminant Analysis(LDA), both satisfy this property.

When data are shifted by $\tilde{\boldsymbol{x}} = \boldsymbol{x} - \boldsymbol{c}$, the data matrix becomes $\tilde{\mathbf{X}} = \mathbf{X} - \boldsymbol{c}\mathbf{1}^T$, where $\mathbf{1} = [1, 1, ..., 1]^T$. Then we have the following theorem:

Theorem 1: If $\mathbf{L}$ is a Laplacian matrix, then we have $\tilde{\mathbf{X}}\mathbf{L}\tilde{\mathbf{X}}^T = \mathbf{X}\mathbf{L}\mathbf{X}^T$, where $\tilde{\mathbf{X}} = \mathbf{X} - \boldsymbol{c}\mathbf{1}^T$.

Proof. According to the definition of Laplacian matrix [12], we know $\mathbf{L}\mathbf{1} = \mathbf{0}$. Then $\tilde{\mathbf{X}}\mathbf{L}\tilde{\mathbf{X}}^T = (\mathbf{X} - \boldsymbol{c}\mathbf{1}^T)\mathbf{L}(\mathbf{X} - \boldsymbol{c}\mathbf{1}^T)^T = \mathbf{X}\mathbf{L}\mathbf{X}^T - \mathbf{X}\mathbf{L}\mathbf{1}\boldsymbol{c}^T - \boldsymbol{c}\mathbf{1}^T\mathbf{L}\mathbf{X} + \boldsymbol{c}\mathbf{1}^T\mathbf{L}\mathbf{1}\boldsymbol{c}^T = \mathbf{X}\mathbf{L}\mathbf{X}^T$. $\square$

It is well known that the projection matrix $\mathbf{W}$ in PCA is formed by the eigenvectors of the total scatter matrix $\mathbf{S}_t$, and the projection matrix $\mathbf{W}$ in LDA is formed by the eigenvectors of $\mathbf{S}_w^{-1}\mathbf{S}_b$, where $\mathbf{S}_w$ is the within-class scatter matrix and $\mathbf{S}_b$ is the between-class scatter matrix. From the view of graph embedding we know that $\mathbf{S}_t = \mathbf{X}\mathbf{L}_t\mathbf{X}^T$, $\mathbf{S}_w = \mathbf{X}\mathbf{L}_w\mathbf{X}^T$ and $\mathbf{S}_b = \mathbf{X}\mathbf{L}_b\mathbf{X}^T$, where $\mathbf{L}_t$, $\mathbf{L}_w$ and $\mathbf{L}_b$ are all Laplacian matrices [13]. According to Theorem 1, if $\mathbf{X}$ is substituted by $\mathbf{X} - \boldsymbol{c}\mathbf{1}^T$, the matrices $\mathbf{S}_t$, $\mathbf{S}_w$ and $\mathbf{S}_b$ are all unchanged, and thus the learned $\mathbf{W}$ in LDA and in PCA are unchanged. Therefore, PCA and LDA both satisfy the shift invariance property.

LPP is also a linear feature extraction algorithm. However, the original LPP algorithm proposed in [2] does not satisfy this property. The reason is that LPP, which is to solve the optimization problem (7), is derived from the optimization problem (4), while the solution to (4) implicitly satisfies the following relationship:

$$\mathbf{Y}\mathbf{D}\mathbf{1} = \mathbf{0} \tag{9}$$

imposing the Equation (6) on the the optimization problem (4) can not guarantee that the implicit relationship (9) is satisfied.

In order to solve this problem and guarantee that the implicit relationship (9) is satisfied in the solution to the derived problem of LPP, we replace the linear constraint in Equation

(6) by another linear constraint as follows:

$$\mathbf{Y} = \mathbf{W}^T\mathbf{X}(\mathbf{I} - \frac{1}{\mathbf{1}^T\mathbf{D}\mathbf{1}}\mathbf{D}\mathbf{1}\mathbf{1}^T) \tag{10}$$

where $\mathbf{I}$ denotes an $n \times n$ identity matrix and $\mathbf{1} = [1, 1, ..., 1]^T$.

Obviously, Equation (10) is also a linear map from $\boldsymbol{x} \in \mathbb{R}^d$ to $\boldsymbol{y} \in \mathbb{R}^m$ and it satisfies the relationship (9) in all cases.

Imposing the Equation (10) insteads of Equation (6) on the the optimization problem (4), we can obtain a shift invariant LPP algorithm, which is to solve the following optimization problem

$$\mathbf{W} = \arg \min_{\mathbf{W}^T\mathbf{X}\mathbf{L}_d\mathbf{X}^T\mathbf{W}=\mathbf{I}} tr(\mathbf{W}^T\mathbf{X}\mathbf{L}\mathbf{X}^T\mathbf{W}) \tag{11}$$

where $\mathbf{L}_d = \mathbf{D} - \frac{\mathbf{1}}{\mathbf{1}^T\mathbf{D}\mathbf{1}}\mathbf{D}\mathbf{1}\mathbf{1}^T\mathbf{D}$.

Theorem 2: $\mathbf{L}_d$ is a Laplacian matrix.

Proof. First, we prove that $\mathbf{L}_d$ is positive semi-definite. For any vector $\boldsymbol{x} \in \mathbb{R}^n$, we have

$$\begin{aligned}
\boldsymbol{x}^T\mathbf{L}_d\boldsymbol{x} &= \boldsymbol{x}^T\mathbf{D}\boldsymbol{x} - \frac{\mathbf{1}}{\mathbf{1}^T\mathbf{D}\mathbf{1}}\boldsymbol{x}^T\mathbf{D}\mathbf{1}\mathbf{1}^T\mathbf{D}\boldsymbol{x} \\
&= \sum_{i=1}^{n}\mathbf{D}_{ii}\boldsymbol{x}_i^2 - \frac{\mathbf{1}}{\sum_{i=1}^{n}\mathbf{D}_{ii}}(\sum_{i=1}^{n}\mathbf{D}_{ii}\boldsymbol{x}_i)^2 \\
&= \frac{\mathbf{1}}{\sum_{i=1}^{n}\mathbf{D}_{ii}}\sum_{i,j=1}^{n}\mathbf{D}_{ii}\mathbf{D}_{jj}(\boldsymbol{x}_i - \boldsymbol{x}_j)^2 \\
&\geq 0
\end{aligned} \tag{12}$$

where $\boldsymbol{x}_i$ is the i-th element in $\boldsymbol{x}$. So $\mathbf{L}_d$ is positive semi-definite.

On the other hand, we can see that $\mathbf{L}_d\mathbf{1} = \mathbf{D}\mathbf{1} - \frac{\mathbf{1}}{\mathbf{1}^T\mathbf{D}\mathbf{1}}\mathbf{D}\mathbf{1}\mathbf{1}^T\mathbf{D}\mathbf{1} = \mathbf{0}$. Therefore, $\mathbf{L}_d$ is a Laplacian matrix. $\square$

As $\mathbf{L}$ and $\mathbf{L}_d$ both are Laplacian matrices, therefore, similarly to PCA and LDA, the shift invariance property is naturally satisfied in problem (11).

IV. GEOMETRIC MEANING OF THE SHIFT INVARIANT LPP

In this section, we reveal the geometric meaning of the derived shift invariant LPP algorithm and then based on the geometric meaning, we propose two novel feature extraction algorithms in the next section.

Denote $p_i = \frac{\mathbf{D}_{ii}}{\sum_i \mathbf{D}_{ii}}$, $\bar{\boldsymbol{x}} = \sum_i p_i \boldsymbol{x}_i$ and $\alpha = \sum_i \mathbf{D}_{ii}$. With some algebraic operations, one can easily verify the following two equations:

$$\mathbf{X}\mathbf{L}\mathbf{X}^T = \frac{1}{2}\sum_{i,j} \mathbf{A}_{ij}(\boldsymbol{x}_i - \boldsymbol{x}_j)(\boldsymbol{x}_i - \boldsymbol{x}_j)^T \tag{13}$$

$$\mathbf{X}\mathbf{L}_d\mathbf{X}^T = \alpha \sum_{i=1}^{n} p_i(\boldsymbol{x}_i - \bar{\boldsymbol{x}})(\boldsymbol{x}_i - \bar{\boldsymbol{x}})^T \tag{14}$$

According to Equation (13) and (14), we have

$$tr(\mathbf{X}\mathbf{L}\mathbf{X}^T) = \frac{1}{2}\sum_{i,j} \mathbf{A}_{ij}\|\boldsymbol{x}_i - \boldsymbol{x}_j\|^2 \tag{15}$$

$$tr(\mathbf{X}\mathbf{L}_d\mathbf{X}^T) = \alpha \sum_{i=1}^{n} p_i\|\boldsymbol{x}_i - \bar{\boldsymbol{x}}\|^2 \tag{16}$$

Under a projection matrix $\mathbf{W} \in \mathbb{R}^{d\times m}$, data point $\boldsymbol{x} \in \mathbb{R}^d$ is transformed into $\boldsymbol{y} \in \mathbb{R}^m$ by $\boldsymbol{y} = \mathbf{W}^T\boldsymbol{x}$. It is not difficult to verify the following two equations:

$$tr(\mathbf{W}^T\mathbf{X}\mathbf{L}\mathbf{X}^T\mathbf{W}) = \frac{1}{2}\sum_{i,j} \mathbf{A}_{ij}\|\boldsymbol{y}_i - \boldsymbol{y}_j\|^2 \tag{17}$$

$$\mathbf{W}^T\mathbf{X}\mathbf{L}_d\mathbf{X}^T\mathbf{W} = \alpha \sum_{i=1}^{n} p_i(\boldsymbol{y}_i - \bar{\boldsymbol{y}})(\boldsymbol{y}_i - \bar{\boldsymbol{y}})^T \tag{18}$$

Note that problem (11) is determined by Equation (17) and (18). According to Equation (17) and (18), the geometric meaning of the shift invariant LPP algorithm becomes clear. The algorithm tries to find a projection matrix $\mathbf{W}$, such that under this projection matrix, the sum of the Euclidean distances between data pairs which are local to each other is minimized, while the weighted covariance matrix is fixed to a constant matrix($\mathbf{W}^T\mathbf{X}\mathbf{L}_d\mathbf{X}^T\mathbf{W} = \mathbf{I}$). It is worthy noting that in the original LPP, the meaning of $\mathbf{W}^T\mathbf{X}\mathbf{D}\mathbf{X}^T\mathbf{W}$ is not the weighted covariance matrix anymore.

V. LOCALITY MINIMIZING GLOBALITY MAXIMIZING PROJECTIONS

From the previous analysis we know, LPP is to minimize $tr(\mathbf{W}^T\mathbf{X}\mathbf{L}\mathbf{X}^T\mathbf{W})$, while minimizing $tr(\mathbf{W}^T\mathbf{X}\mathbf{L}\mathbf{X}^T\mathbf{W})$ is equivalent to minimize the distances of data pairs in locality. To gain more discriminative power, it is desirable to minimize the locality and maximize the globality simultaneously.

According to Equation (18), we have

$$tr(\mathbf{W}^T\mathbf{X}\mathbf{L}_d\mathbf{X}^T\mathbf{W}) = \alpha\sum_{i=1}^{n} p_i\|\boldsymbol{y}_i - \bar{\boldsymbol{y}}\|^2 \tag{19}$$

so $tr(\mathbf{W}^T\mathbf{X}\mathbf{L}_d\mathbf{X}^T\mathbf{W})$ is the weighted variance of data, which can be seen as the global information in data. Therefore, it is desirable to minimize $tr(\mathbf{W}^T\mathbf{X}\mathbf{L}\mathbf{X}^T\mathbf{W})$ and maximize $tr(\mathbf{W}^T\mathbf{X}\mathbf{L}_d\mathbf{X}^T\mathbf{W})$ simultaneously. To this end, it is reasonable to select the following two criteria:

$$\mathcal{M}_1(\mathbf{W}) = tr(\mathbf{W}^T\mathbf{X}(\lambda\mathbf{L} - \mathbf{L}_d)\mathbf{X}^T\mathbf{W}) \tag{20}$$

$$\mathcal{M}_2(\mathbf{W}) = \frac{tr(\mathbf{W}^T\mathbf{X}\mathbf{L}\mathbf{X}^T\mathbf{W})}{tr(\mathbf{W}^T\mathbf{X}\mathbf{L}_d\mathbf{X}^T\mathbf{W})} \tag{21}$$

where λ is a user predefined constant. It is easy to see that the above two criteria both satisfy the shift invariance property.

In order to obtain an orthogonal projection matrix $\mathbf{W}$, we further add a constraint as $\mathbf{W}^T\mathbf{W} = \mathbf{I}$, then solve the following two optimization problems:

$$\mathbf{W} = \arg\min_{\mathbf{W}^T\mathbf{W}=\mathbf{I}} tr(\mathbf{W}^T\mathbf{X}(\lambda\mathbf{L} - \mathbf{L}_d)\mathbf{X}^T\mathbf{W}) \tag{22}$$

$$\mathbf{W} = \arg\min_{\mathbf{W}^T\mathbf{W}=\mathbf{I}} \frac{tr(\mathbf{W}^T\mathbf{X}\mathbf{L}\mathbf{X}^T\mathbf{W})}{tr(\mathbf{W}^T\mathbf{X}\mathbf{L}_d\mathbf{X}^T\mathbf{W})} \tag{23}$$

The solution to problem (22) can be easily obtained by eigen-decomposition of $\mathbf{X}(\lambda\mathbf{L} - \mathbf{L}_d)\mathbf{X}^T$ as follows

$$\mathbf{X}(\lambda\mathbf{L} - \mathbf{L}_d)\mathbf{X}^T\mathbf{W} = \mathbf{W}\boldsymbol{\Lambda} \tag{24}$$

where $\boldsymbol{\Lambda}$ is the eigenvalue matrix of $\mathbf{X}(\lambda\mathbf{L} - \mathbf{L}_d)\mathbf{X}^T$ and $\mathbf{W}$ is the corresponding eigenvector matrix.

Solving problem (23) is a little intractable. However, fortunately, the solution to it can also be efficiently obtained by iterative procedure [14, 15].

It is worthy noting that solving the optimization problems in (22)-(23) can be served as a general graph based feature extraction framework. Different constructing of the Laplacian matrices $\mathbf{L}$ and $\mathbf{L}_d$ will lead to different unsupervised, semi-supervised or supervised feature extraction algorithms, and the corresponding kernelization and tensorization extensions can also be easily derived from this framework [13].

VI. EXPERIMENTAL RESULTS

In this section, we evaluate the performance of the proposed algorithms for face recognition. The algorithms corresponding to optimization problems (22) and (23) are denoted by LMGMP1 and LMGMP2, respectively. In the experiments, λ in Equation (20) is set to 2. We compare the proposed algorithms with LDA, LPP and shift invariant LPP(SI-LPP).

LMGMP1, LMGMP2, LPP, SI-LPP construct the same graph structure based on the label information. PCA is used as a preprocessing step before performing the algorithms.

In each experiment, we randomly select several samples per class for training and the remaining samples are used for testing. The classification is based on 1-nearest neighbor classifier. The average accuracy rates and the standard deviations are recorded over 50 random splits.

A. AT&T Database

The AT&T face database includes 40 distinct individuals and each individual has 10 different images. Each image is down-sampled to the size of 28×23. The training number t per class is 2, 4, 6, respectively. The best results of each algorithm are reported in Table I. For $t = 4$, the results of accuracy rate versus dimension are shown in Figure 1.

As can be seen in Table I and Figure 1, the performances of the proposed algorithms are much better than those of LDA and LPP, especially when the reduced dimension is very low. SI-LPP also gains an improvement over LPP.

B. UMIST Database

The UMIST repository is a multiview database, consisting of 575 images of 20 people, each covering a wide range of poses from profile to frontal views. We down-sample the size of each image to 28×23. The training number t per class is 4, 6, 8, respectively. The best results of each algorithm are reported in Table II. For $t = 6$, the results of accuracy rate versus dimension are shown in Figure 2.

Similarly to the above experiments, the performances of the proposed algorithms are significantly better than those of LDA and LPP, especially when the reduced dimension is very low. SI-LPP also gains an improvement over LPP.

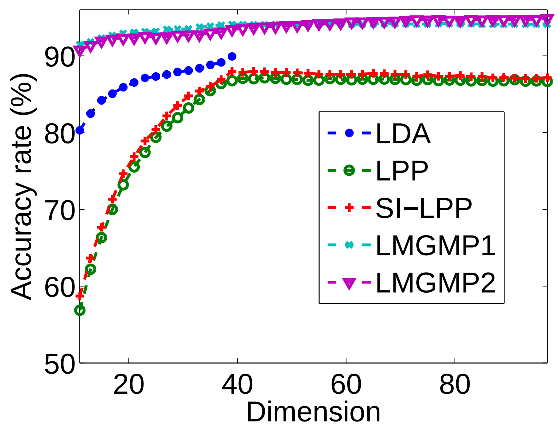

FIG. 1: Accuracy vs. dimensions on the AT&T database (4 train).

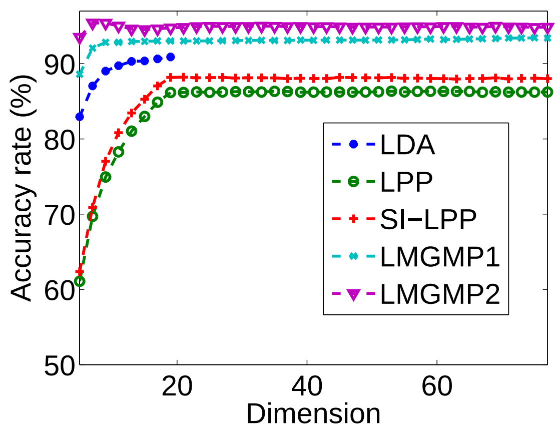

FIG. 2: Accuracy vs. dimensions on the UMIST database (6 train).

VII. CONCLUSION

In this paper, we point out that the Locality Preserving Projections(LPP) algorithm, which is developed recently and has been frequently used in many applications, does not satisfy the shift invariance property. The reason is analyzed and the shift invariant LPP algorithm is derived. We also reveal the geometrical meaning of the derived shift invariant LPP. Based on the analysis of the geometrical meaning, we propose two novel algorithms to minimize the locality and maximize the globality simultaneously under an orthogonal projection matrix. Several recently proposed works related to our algorithms are discussed. Experimental results on face recognition demonstrate the effectiveness and superiority of the proposed algorithms.

---

[1] R. O. Duda., P. E. Hart., and D. G. Stork, *Pattern Classification.* Hoboken, NJ: Wiley-Interscience, 2000.

[2] X. He and P. Niyogi, "Locality preserving projections," in *NIPS*, 2003.

[3] M. Belkin and P. Niyogi, "Laplacian eigenmaps for dimensionality reduction and data representation," *Neural Computation*, vol. 15, no. 6, pp. 1373–1396, 2003.

[4] J. Yang, D. Zhang, J. Y. Yang, and B. Niu, "Globally maximizing, locally minimizing: Unsupervised discriminant projection with applications to face and palm biometrics," *IEEE Trans. PAMI*, vol. 29, no. 4, pp. 650–664, 2007.

[5] J. Yan, N. Liu, B. Zhang, S. Yan, Z. Chen, Q. Cheng, W. Fan, and W.-Y. Ma, "OCFS: optimal orthogonal centroid feature selection for text categorization," in *SIGIR*, 2005, pp. 122–129.

[6] J. Ye, "Characterization of a family of algorithms for generalized discriminant analysis on undersampled problems," *Journal of Machine Learning Research*, vol. 6, pp. 483–502, 2005.

[7] C. H. Q. Ding, T. Li, W. Peng, and H. Park, "Orthogonal nonnegative matrix tri-factorizations for clustering," in *KDD*, 2006, pp. 126–135.

[8] X. Liu, J. Yin, Z. Feng, J. Dong, and L. Wang, "Orthogonal neighborhood preserving embedding for face recognition," in *ICIP*, 2007, pp. I–133–I–136.

[9] H. Wang, S. Chen, Z. Hu, and W. Zheng, "Locality-preserved maximum information projection," *IEEE Transactions on Neural Networks*, vol. 19, no. 4, pp. 571–585, 2008.

[10] E. Kokiopoulou and Y. Saad, "Orthogonal neighborhood preserving projections: A projection-based dimensionality reduction technique," *IEEE Trans. PAMI*, vol. 29, no. 12, pp. 2143–2156, 2007.

[11] D. Cai, X. He, J. Han, and H.-J. Zhang, "Orthogonal laplacianfaces for face recognition," *IEEE Trans. on Image Processing*, vol. 15, no. 11, pp. 3608–3614, 2006.

[12] F. R. K. Chung, *Spectral Graph Theory.* CBMS Regional Conference Series in Mathematics, No. 92, American Mathematical Society, February 1997.

[13] S. Yan, D. Xu, B. Zhang, H. Zhang, Q. Yang, and S. Lin, "Graph embedding and extensions:

A general framework for dimensionality reduction," *PAMI*, vol. 29, no. 1, pp. 40–51, 2007.

[14] F. Nie, S. Xiang, and C. Zhang, "Neighborhood minmax projections." in *IJCAI*, 2007, pp. 993–998.

[15] H. Wang, S. Yan, D. Xu, X. Tang, and T. S. Huang, "Trace ratio vs. ratio trace for dimensionality reduction," in *CVPR*, 2007.

Biographies:

Feiping Nie received his B.S. degree in Computer Science from North China University of Water Conservancy and Electric Power, China in 2000, and received his M.S. degree in Computer Science from Lanzhou University, China in 2003. He is currently a Ph.D. Candidate in the Department of Automation, Tsinghua University, China. His research interests include machine learning, pattern recognition, data mining and image processing.

Shiming Xiang received his B.S. degree from Department of Mathematics of Chongqing Normal University, China in 1993 and M.S. degress from Department of Mechanics & Mathematics of Chongqing University, China in 1996, and Ph.D. degree from Institute of Computing Technology, Chinese Academy of Sciences, China in 2004. He was a post Doctor in Department of Automation, Tsinghua University until 2006. He is currently an associate researcher in Institute of Automation Chinese Academy of Science. His interests include computer vision, pattern recognition, machine learning, etc.

Yangqiu Song received his B.S. degree from Department of Automation, Tsinghua University, China in 2003. He is currently a Ph.D candidate in Department of Automation, Tsinghua University. His research interests focus on machine learning and its applications.

Changshui Zhang received his B.S. degree in Mathematics from Peking University, China in 1986, and Ph.D. degree from Department of Automation, Tsinghua University in 1992. He is currently a professor of Department of Automation, Tsinghua University. He is an Associate Editor of the journal Pattern Recognition. His interests include artificial intelligence, image processing, pattern recognition, machine learning, evolutionary computation and complex system analysis, etc. Tel.: +86-10-672-82-447, +86-10-627-82-455; fax: +86-10-627-86-911; E-mails: zcs@mail.tsinghua.edu.cn

TABLE I: Accuracy ± dev. (%) on the AT&T Database.

	2 TRAIN	4 TRAIN	6 TRAIN
LDA	77.5 ± 2.9	89.9 ± 1.8	91.6 ± 2.0
LPP	77.4 ± 2.7	87.1 ± 1.9	88.6 ± 2.3
SI-LPP	79.1 ± 2.6	88.1 ± 1.9	89.1 ± 2.2
LMGMP1	82.9 ± 2.8	94.2 ± 1.6	96.9 ± 1.5
LMGMP2	**84.6** ± **2.6**	**94.9** ± **1.7**	**97.2** ± **1.3**

TABLE II: Accuracy ± dev. (%) on the UMIST database.

	4 TRAIN	6 TRAIN	8 TRAIN
LDA	85.0 ± 3.5	90.9 ± 2.4	93.6 ± 2.1
LPP	79.6 ± 3.7	86.3 ± 2.6	90.2 ± 2.4
SI-LPP	81.8 ± 3.8	88.2 ± 2.6	91.5 ± 2.4
LMGMP1	86.5 ± 3.7	93.5 ± 2.1	96.3 ± 1.4
LMGMP2	**90.6** ± **2.9**	**95.4** ± **1.3**	**97.3** ± **1.2**

Figure caption list:

Accuracy vs. dimensions on the AT&T database (4 train).

Accuracy vs. dimensions on the UMIST database (6 train).