Pattern Recognition 43 (2010) 4165–4176

Contents lists available at ScienceDirect

Pattern Recognition

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

# LPP solution schemes for use with face recognition

Yong Xu [a,*], Aini Zhong [a], Jian Yang [b], David Zhang [c]

[a] Bio-Computing Research Center, Shenzhen Graduate School, Harbin Institute of Technology, China
[b] School of Computer Science & Technology, Nanjing University of Science & Technology, Nanjing, Jiangsu 210094, China
[c] The Biometrics Research Center, The Hong Kong Polytechnic University, Kowloon, Hong Kong

ARTICLE INFO

*Article history:*
Received 18 February 2010
Received in revised form
19 May 2010
Accepted 18 June 2010

*Keywords:*
Face recognition
Feature extraction
Locality preserving projection
Small sample size problems

ABSTRACT

Locality preserving projection (LPP) is a manifold learning method widely used in pattern recognition and computer vision. The face recognition application of LPP is known to suffer from a number of problems including the small sample size (SSS) problem, the fact that it might produce statistically identical transform results for neighboring samples, and that its classification performance seems to be heavily influenced by its parameters. In this paper, we propose three novel solution schemes for LPP. Experimental results also show that the proposed LPP solution scheme is able to classify much more accurately than conventional LPP and to obtain a classification performance that is only little influenced by the definition of neighbor samples.



## 1. Introduction

Locality preserving projection (LPP) is a manifold learning method [1–7] widely used in pattern recognition and computer vision. LPP is also well-known as a linear graph embedding method [8,9]. When LPP transforms different samples into new representations using the same linear transform, it tries to preserve the local structure of the samples, i.e., the neighbor relationship between samples [10–15] so that samples that were originally in close proximity in the original space remain so in the new space. We note that the original LPP method was unsupervised and was proposed for only vector samples, not being able to be directly applied to image samples. Here 'unsupervised' means that when producing the transforming axis the original LPP method does not exploit the class-label information. Hereafter this method is referred to as conventional LPP.

There have been several types of improvements to conventional LPP. The first type of improvement is supervised LPP [16–19], which seeks to improve the performance of LPP in recognition problems by exploiting the class-label information of samples in the training phase. The main difference between unsupervised LPP and supervised LPP is that unsupervised LPP uses only the distance metric between samples to determines 'neighbor samples' whereas supervised LPP uses both the distance metric and the class label of samples to determine 'neighbor samples'. Supervised LPP does not regard two samples from two different classes are 'neighbors' even if they are in close proximity to each other. Since the weight matrix is determined on the basis of neighbor relationship between samples, having different weight matrices is also one of the main differences between supervised LPP and unsupervised LPP. It is usually thought that supervised LPP can outperform unsupervised LPP in classification applications owing to the use of the class-label information. Local discriminant embedding (LDE) [20] and marginal Fisher analysis (MFA) [21] can also be viewed as supervised LPP methods. This is because their training phases both exploit the class-label information of samples. They are derived by using a motivation partially similar to LPP and each of them is based on an eigen-equation formally similar to the eigen-equation of LPP. On the other hand, since LDE and MFA partially borrow the idea of discriminant analysis and try to produce satisfactory linear separability, their ideas are also somewhat different from the idea of preserving the local structure of LPP. LDE and MFA can be viewed as two combinations of the locality preserving technique and the linear discriminant analysis. The two methods probably perform worse than the conventional supervised LPP in preserving the local structure.

The second type of improvement changes conventional LPP to a nonlinear transform method by using the kernel trick [19–24]. This type of improvement transforms a sample into a linear combination of a number of kernel functions each being determined by this sample and one training sample. The method uses the same linear combination coefficients to transform each sample into the new representation. Because the kernel function is nonlinearly related to the sample, the transform mapping is nonlinear. The third type of improvement to conventional LPP mainly focuses on directly implementing LPP for two-dimensional rather than one-dimensional vectors. This allows LPP to have a

* Corresponding author. Tel.: +86 752 26032458; fax: +86 752 26032461.
*E-mail address:* laterfall286@yahoo.com (Y. Xu).

0031-3203/$ - see front matter © 2010 Elsevier Ltd. All rights reserved.
doi:10.1016/j.patcog.2010.06.016

higher computational efficiency. This type of improvement has been referred to as two-dimensional locality preserving projection (2DLPP) [25,26]. The fourth type of improvement to conventional LPP seeks to obtain LPP solutions with different solution properties. Examples of this type of improvement include orthogonal locality preserving method [27], uncorrelated LPP feature extraction method [28], the fast implementation algorithm for unsupervised orthogonal LPP [29], and the LPP algorithm for the SSS problem [30].

Different improvements to conventional LPP can also be regarded as implementations of the idea of locality preserving projection under different constraint conditions or in different cases. For example, unsupervised LPP requires all the samples to preserve their local neighbor relationship, whereas supervised LPP requires only samples from the same class to preserve their neighbor relationship. As conventional LPP was devised for vector data, its implementation on an image-based application requires that the image be converted into a vector in advance. However, 2DLPP is devised for image matrices, which means that 2DLPP directly applies the idea of locality preserving projection to image matrices rather than the vector corresponding to the image. Recently it has been demonstrated that LPP is theoretically related and formally similar to other linear dimensionality reduction methods and the main difference between LPP and them is in the weight matrix. Indeed, many popular linear dimensionality reduction methods including unsupervised LPP, supervised LPP, linear discriminant analysis (LDA), MFA, LDE and neighborhood preserving embedding (NPE) can be described as the implementations of the linear graph embedding framework with different weight matrices [31]. Conventional LPP and its improvements have been used in face recognition, image retrieval, document analysis, data clustering, etc. [11,16–18,32,33].

As in image-based applications, conventional LPP should first convert the image into the vector and as conventional LPP obtains the transforming axes by solving the minimum or maximum eigenvalue solution of a generalized eigen-equation, conventional LPP usually suffers from several problems. The first problem is that the dimensionality of the sample is usually larger than the number of the samples and the generalized eigen-equation cannot be directly solved due to the matrix singularity problem. This problem is also referred to as the small sample size (SSS) problem. An image-based recognition problem such as face recognition is usually a SSS problem. On the other hand, image-based recognition covers a wide range of pattern recognition problems. Thus, the study of how to properly apply LPP to the SSS problem is crucial. To the best of our knowledge, no satisfactory approach to this study has been proposed. Most of previous LPP-based image recognition applications avoid the SSS problem. For example, a number of face recognition applications of conventional LPP first reduce the size of the face image and then implement the conventional LPP algorithm for the resized images. In order to make the conventional LPP algorithm workable, the dimensionality of the vector of the resized image should be smaller than the number of the training samples. Consequently, in order to avoid the SSS problem, the original image usually should be resized into a very small size. This will cause a large quantity of image information loss. Another example of avoiding the SSS problem is to first reduce the dimensionality of the sample by performing principal component analysis and then to carry out the conventional LPP algorithm [23]. But there are no guidelines for how to use principal component analysis to transform the sample into a proper dimensionality. If the extent of reduction is too great, there will be considerable information loss. On the other hand, if the dimension reduction extent is small, the corresponding eigen-equation is still singular and cannot be solved directly.

A further drawback of conventional LPP is that if it is implemented by solving the minimum eigenvalue problem, the minimum eigenvalue solution is not always optimal for preserving the local structure. There are two reasons for this. The first is that if there are zero eigenvalues, conventional LPP will take as transforming axes the eigenvectors corresponding to the zero-eigenvalues of the generalized eigen-equation. As a result, after conventional LPP transforms samples into a new space using these transforming axes, a sample statistically will have the same representation as its neighbors, which will be formally demonstrated in Section 2. This is not how locality preserving projection works. The goal of LPP is not to make samples have the same representation but is to preserve the neighbor relationships between samples. The second reason is that the classifier cannot correctly classify samples when conventional LPP is implemented in the unsupervised case, since two neighbor samples from two different classes might have the same representation in the new space.

We also note that when implementing a LPP solution scheme, we should define a specific number of neighbor samples for each particular sample. In practice, it is not known how different values of this number influence the classification performance.

In this paper, we propose three new solution schemes for LPP. These new schemes have three advantages. The first is that they can be directly implemented no matter whether there exists the SSS problem or not. The second is that they are consistent with the goal of LPP and have a clear justification. The third advantage is that experimental results show that these schemes are more accurate than conventional LPP. This paper also conducts experiments to show the effect on classification performance of the number of neighbor samples and the value of the parameter $k$ of $k$-nearest-neighbor classifiers (KNNC). The experimental results show that the improved LPP solution scheme 3 is not only computationally efficient, but also classifies much more accurately than conventional LPP. It is also a well-behaved LPP solution scheme whose classification accuracy is little influenced by the definition of neighbor samples.

The remainder of the paper is organized as follows. In Section 2 we introduce the algorithm of conventional LPP and analyze its characteristics. In Section 3 we present our LPP solution schemes and show their characteristics. In Section 4 we describe the experimental results. Section 5 offers our Conclusion.

## 2. Description of LPP

LPP was proposed as a way to transform samples into a new space and to ensure that samples that were in close proximity in the original space remain so in the new space. The goal of LPP is to minimize the following function:

$$\frac{1}{2}\sum_{ij}(y_i-y_j)^2 w_{ij}, \tag{1}$$

where $y_i$, $y_j$ are transform results of vector samples $\mathbf{x}_i$, $\mathbf{x}_j$, and $w_{ij}$ is the weight coefficient. $y_i$ is obtained by using a transforming axis $\mathbf{z}$. That is, we have $y_i=\mathbf{x}_i^T\mathbf{z}$ and $y_j=\mathbf{x}_j^T\mathbf{z}$. The function defined in Eq. (1) can be rewritten as

$$\frac{1}{2}\sum_{ij}(\mathbf{z}^T\mathbf{x}_i-\mathbf{z}^T\mathbf{x}_j)^2 w_{ij}=\sum_{ij}\mathbf{z}^T\mathbf{x}_i w_{ij}\mathbf{x}_i^T\mathbf{z}-\sum_{ij}\mathbf{z}^T\mathbf{x}_i w_{ij}\mathbf{x}_j^T\mathbf{z}. \tag{2}$$

By defining a matrix $\mathbf{W}$and a dialog matrix $\mathbf{D}$ as $(\mathbf{W})_{ij}=w_{ij}$, $(\mathbf{D})_{ii}=\sum_j w_{ij}$ we can transform (2) into

$$\mathbf{z}^T\mathbf{X}(\mathbf{D}-\mathbf{W})\mathbf{X}^T\mathbf{z}=\mathbf{z}^T\mathbf{X}\mathbf{L}\mathbf{X}^T\mathbf{z}, \tag{3}$$

where $\mathbf{L}=\mathbf{D}-\mathbf{W}$, $n$ is the number of the training samples $\mathbf{x}_1,\mathbf{x}_2,\ldots,\mathbf{x}_n$ and the matrix $\mathbf{X}$ is defined as $\mathbf{X}=[\mathbf{x}_1\ \mathbf{x}_2\ \ldots\ \mathbf{x}_n]$. $\mathbf{x}_1,\mathbf{x}_2,\ldots,\mathbf{x}_n$ are supposed to be column vectors. Note that both $\mathbf{L}$ and $\mathbf{D}$ are positive semi-definitive matrices. $\mathbf{W}$ is a symmetric matrix and the element $w_{ij}$ (also referred to as the weight coefficient) is defined as follows:

$$w_{ij}=\begin{cases}\exp(-\|\mathbf{x}_i-\mathbf{x}_j\|^2/t) & \text{if } \mathbf{x}_j \text{ (or } \mathbf{x}_i\text{) is one of } Wk \text{ neighbors of } \mathbf{x}_i \text{ (or } \mathbf{x}_j\text{)}\\ 0 & \text{otherwise}\end{cases}$$

where the parameter $Wk$ is set to a positive constant. This definition shows that if $\mathbf{x}_j$ (or $\mathbf{x}_i$) is one of the $Wk$ neighbors of $\mathbf{x}_i$ (or $\mathbf{x}_j$), then the weight coefficient is set to $w_{ij}=\exp(-||\mathbf{x}_i-\mathbf{x}_j||^2/t)$; otherwise the weight coefficient is set to zero. It is clear that the value of the non-zero weight coefficient is determined by the distance between samples. The smaller the distance between two neighbor samples $\mathbf{x}_i$ and $\mathbf{x}_j$, the larger the weight coefficient $w_{ij}$. As a result, the minimization of the function in Eq. (1) implies that if the weight coefficient of two samples is large, their transform results should be in close proximity in the new space. Therefore, through the LPP transform with the objection function (1), close samples in the original space will be still close in the new space.

### 2.1. Conventional LPP solution scheme 1

When LPP takes the following function as its goal

$$\underset{\mathbf{z}^T\mathbf{X}\mathbf{D}\mathbf{X}^T\mathbf{z}=1}{\operatorname{argmin}}\ \mathbf{z}^T\mathbf{X}\mathbf{L}\mathbf{X}^T\mathbf{z}, \tag{4}$$

we refer to the corresponding solution scheme as the conventional LPP solution scheme 1. The goal of the conventional LPP solution scheme 1 is to obtain the $\mathbf{z}$ that satisfies $\arg\min(\mathbf{z}^T\mathbf{X}\mathbf{L}\mathbf{X}^T\mathbf{z}/\mathbf{z}^T\mathbf{X}\mathbf{D}\mathbf{X}^T\mathbf{z})$. It can be easily proven that the optimal solution $\mathbf{z}$ of Eq. (4) is identical to the eigenvector corresponding to the minimum eigenvalue of the following generalized eigen-equation:

$$\mathbf{X}\mathbf{L}\mathbf{X}^T\mathbf{z}=\lambda\mathbf{X}\mathbf{D}\mathbf{X}^T\mathbf{z}. \tag{5}$$

It is clear that Eq. (5) has the same minimum eigenvalue solution as the following eigen-equation:

$$(\mathbf{X}\mathbf{D}\mathbf{X}^T)^{-1}\mathbf{X}\mathbf{L}\mathbf{X}^T\mathbf{z}=\lambda\mathbf{z}. \tag{6}$$

In other words, the conventional LPP solution scheme 1 takes as the optimal transforming axis the eigenvector corresponding to the minimum eigenvalue of (6). In real-world applications, the conventional LPP solution scheme 1 usually first sorts the eigenvectors in increasing order of the corresponding eigenvalues and then takes the eigenvectors corresponding to the smallest eigenvalues as transforming axes to implement the LPP transform. The dimension of the transform result of the sample is identical to the number of the transforming axes that are used.

The conventional LPP solution scheme 1 may suffer from the following drawback: in the case of the so-called SSS problem where since the rank of $\mathbf{X}\mathbf{D}\mathbf{X}^T$ is smaller than its order and $\mathbf{X}\mathbf{D}\mathbf{X}^T$ is a singular matrix, the generalized eigen-equation (6) cannot be directly solved. Indeed, the rank of $\mathbf{X}\mathbf{D}\mathbf{X}^T$ must be no larger than the number of the training samples. On the other hand, the order of $\mathbf{X}\mathbf{D}\mathbf{X}^T$ is equal to the dimension of the sample vector. Consequently, if the dimension of the sample vector is larger than the number of the samples, the matrix $\mathbf{X}\mathbf{D}\mathbf{X}^T$must be singular and both (5) and (6) cannot be solved directly. This is demonstrated in detail by the following theorems and inference.

**Theorem 1.** *Let $N$ and $n$ be the dimension of the sample vector and the number of the training samples, respectively. If $N>n$, then $\mathbf{X}\mathbf{D}\mathbf{X}^T$ and $\mathbf{X}\mathbf{L}\mathbf{X}^T$ must be singular.*

**Proof.** The maximum possible rank of both $\mathbf{X}$ and $\mathbf{X}^T$ is $n$. This can be described by $rank(\mathbf{X})=rank(\mathbf{X}^T)\le n$. It is known that the maximum possible rank of the product of two matrices is smaller than or equal to the smaller of the ranks of the two matrices. That is, if $rank(\mathbf{A})=r_A$ and $rank(\mathbf{B})=r_B$, then $rank(\mathbf{AB})\le\min(r_A,r_B)$, where $\mathbf{A}$ and $\mathbf{B}$ are two matrices. As a result, we know that $rank(\mathbf{X}\mathbf{D}\mathbf{X}^T)\le n$ and $rank(\mathbf{X}\mathbf{L}\mathbf{X}^T)\le n$.

Both $\mathbf{X}\mathbf{D}\mathbf{X}^T$ and $\mathbf{X}\mathbf{L}\mathbf{X}^T$ are $N\times N$ matrices. Since $N>n$, it is certain that $rank(\mathbf{X}\mathbf{D}\mathbf{X}^T)<N$ and $rank(\mathbf{X}\mathbf{L}\mathbf{X}^T)<N$, which means that $\mathbf{X}\mathbf{D}\mathbf{X}^T$ and $\mathbf{X}\mathbf{L}\mathbf{X}^T$ are both singular matrices.

We can conclude that in the case of the SSS problem, since $\mathbf{X}\mathbf{D}\mathbf{X}^T$ is singular the eigen-equations (5) and (6) cannot be directly solved. This is the first drawback of the conventional LPP solution scheme 1.

If one converts Eqs. (5) and (6), respectively, into (7) and (8), the conventional LPP solution can be worked out directly:

$$\mathbf{X}\mathbf{L}\mathbf{X}^T\mathbf{z}=\lambda(\mathbf{X}\mathbf{D}\mathbf{X}^T+\mu\mathbf{I})\mathbf{z}, \tag{7}$$

$$(\mathbf{X}\mathbf{D}\mathbf{X}^T+\mu\mathbf{I})^{-1}\mathbf{X}\mathbf{L}\mathbf{X}^T\mathbf{z}=\lambda\mathbf{z}, \tag{8}$$

where $\mu$ is a small positive constant and $\mathbf{I}$ is the identity matrix. In practice, a similar procedure for solving singular eigen-equations has been widely used in numerical computation [34,35]. However, even if we do so, we will find that when the eigenvector of the minimum eigenvalue of (7) or (8) serves as the LPP transforming axis of the SSS problem, the transform result will represent data poorly. This is because in the case above the minimum eigenvalue is zero and the conventional LPP solution scheme will take as the transforming axis the eigenvector of the zero-eigenvalue of (7) or (8). Consequently, as shown in the following theorem, neighbor samples will statistically produce the same transform result.

**Theorem 2.** *After the conventional LPP solution scheme 1 transforms samples into new representations by using the eigenvectors corresponding to the minimum eigenvalues as transforming axes and where the minimum eigenvalue of the eigen-equation is zero, a sample will statistically have the same representation as its neighbors.*

**Proof.** If both sides of Eq. (7) are left multiplied by $\mathbf{z}^T$, we have

$$\mathbf{z}^T\mathbf{X}\mathbf{L}\mathbf{X}^T\mathbf{z}=\lambda\mathbf{z}^T(\mathbf{X}\mathbf{D}\mathbf{X}^T+\mu\mathbf{I})\mathbf{z}. \tag{9}$$

If Eq. (7) has zero eigenvalues and $\mathbf{z}_0$ is the eigenvector corresponding to a zero eigenvalue, then we obtain

$$\mathbf{z}_0^T\mathbf{X}\mathbf{L}\mathbf{X}^T\mathbf{z}_0=0. \tag{10}$$

As mentioned in Section 2.1, there is $1/2\sum_{ij}(y_i-y_j)^2w_{ij}=\mathbf{z}^T\mathbf{X}\mathbf{L}\mathbf{X}^T\mathbf{z}$. Consequently, if the eigenvector corresponding to a zero eigenvalue is taken as the transforming axis, the transform result will statistically satisfy the following condition:

$$\frac{1}{2}\sum_{ij}(y_i-y_j)^2w_{ij}=0. \tag{11}$$

Note that $w_{ij}\ge0$ and $(y_i-y_j)^2\ge0$ are certain for arbitrary $i$ and $j$. Hence, Eq. (11) means that for arbitrary $i$ and $j$, $(y_i-y_j)^2w_{ij}=0$ should be satisfied. In particular, for two neighbor samples $w_{ij}>0$

is satisfied. As a result, $(y_i-y_j)^2w_{ij}=0$ implies that in the transform space the neighbor samples must have the same representation. This does not preserve the local structure of data as in LPP, which requires that the transform results of neighbor samples be in close proximity rather than the same. This is one drawback of the conventional LPP solution scheme 1. According to the matrix eigenvalue theory, we know that the number of zero-eigenvalues of either of (7) or (8) is $N-r$. $N$ and $r$ are the dimensionality of the sample vector and the rank of $\mathbf{XLX}^T$, respectively.

A further potential drawback of the conventional LPP solution scheme 1 is that even if the neighbor samples are from different classes, in the transform space obtained using the conventional LPP solution scheme 1 they might also statistically have the same representation, which is disadvantageous for pattern classification problems. In other words, it is possible for the conventional LPP solution scheme 1 to produce the same representation for samples from different classes, especially for the samples located on the border of two classes. All unsupervised LPP methods might suffer from this same drawback.

### 2.2. Conventional LPP solution scheme 2

The following describes another conventional LPP solution scheme. According to the definition of $\mathbf{L}$, we know that

$$\mathbf{z}^T\mathbf{XLX}^T\mathbf{z}=\mathbf{z}^T\mathbf{XDX}^T\mathbf{z}-\mathbf{z}^T\mathbf{XWX}^T\mathbf{z}. \tag{12}$$

As a result, the objective function as shown in Eq. (4) is equivalent to the following one:

$$\arg\max_{\mathbf{z}^T\mathbf{XDX}^T\mathbf{z}=1}\mathbf{z}^T\mathbf{XWX}^T\mathbf{z}. \tag{13}$$

It is easy to prove that (13) is identical to $\arg\max(\mathbf{z}^T\mathbf{XWX}^T\mathbf{z}/\mathbf{z}^T\mathbf{XDX}^T\mathbf{z})$ and the optimal solution of Eq. (13) is the eigenvector corresponding to the maximum eigenvalue of either of the following generalized eigen-equations:

$$\mathbf{XWX}^T\mathbf{z}=\lambda\mathbf{XDX}^T\mathbf{z}, \tag{14}$$

i.e.,

$$(\mathbf{XDX}^T)^{-1}\mathbf{XWX}^T\mathbf{z}=\lambda\mathbf{z}. \tag{15}$$

Clearly, given the objective function (13), we should exploit the maximum eigenvalue solution of Eqs. (15) or (14) to implement the LPP transform. In other words, after we work out the eigenvalues and eigenvectors of Eq. (15), we should select as the LPP transforming axes the eigenvectors corresponding to maximum eigenvalues of Eq. (15). We refer to this solution scheme as the conventional LPP solution scheme 2. When the conventional LPP solution scheme 2 is applied to the SSS problem, it also suffers from the matrix singularity problem and the generalized eigen-equation is also not directly solvable. On the other hand, if $\mathbf{XDX}^T$ is replaced by $(\mathbf{XDX}^T+\mu\mathbf{I})$, the equation becomes directly solvable and the solution does not suffer from the same zero-eigenvalue problem as the conventional LPP solution scheme 1 encounters. In other words, since the conventional LPP solution scheme 2 selects the eigenvectors corresponding to maximum eigenvalues of (15) as the transforming axes, the conventional LPP solution scheme 2 will not encounter the drawback that in the transform space neighbor samples statistically have the same representation.

## 3. New LPP solution schemes

This section describes three new LPP solution schemes.

### 3.1. Improved LPP solution scheme 1

#### 3.1.1. Formal description

In this subsection, we describe our improvement to the conventional LPP solution scheme 1, i.e., the improved solution scheme 1. This solution scheme does not have the same drawbacks as the conventional LPP solution scheme 1.

First, we demonstrate that effective solutions of the conventional LPP solution scheme 1 should be from a subspace of $\mathbf{XDX}^T$. For simplicity, we define that $\mathbf{D}_1=\mathbf{XDX}^T$, $\mathbf{L}_1=\mathbf{XLX}^T$ and $\mathbf{W}_1=\mathbf{XWX}^T$. Suppose that $\vec{\boldsymbol{\alpha}}_1,\vec{\boldsymbol{\alpha}}_2,\ldots,\vec{\boldsymbol{\alpha}}_r$ are the eigenvectors corresponding to the positive eigenvalues of $\mathbf{D}_1$ and $\vec{\boldsymbol{\alpha}}_{r+}1,\vec{\boldsymbol{\alpha}}_{r+}2,\ldots,\vec{\boldsymbol{\alpha}}_N$ are the unit eigenvectors corresponding to the zero eigenvalues. Here, we regard eigenvalues that are less than $0.2\times10^{-10}$ as zero eigenvalues. According to the nature of LPP, the ability of preserving the neighbor relationship can be measured by $\mathbf{z}^T\mathbf{L}_1\mathbf{z}/\mathbf{z}^T\mathbf{D}_1\mathbf{z}$. That is, the smaller $\mathbf{z}^T\mathbf{L}_1\mathbf{z}/\mathbf{z}^T\mathbf{D}_1\mathbf{z}$, the better the local structure of samples is preserved. We can rewrite Eq. (5) as $\mathbf{L}_1\mathbf{z}=\lambda\mathbf{D}_1\mathbf{z}$ or

$$\mathbf{z}^T\mathbf{L}_1\mathbf{z}=\lambda\mathbf{z}^T\mathbf{D}_1\mathbf{z}. \tag{16}$$

We denote the range space of $\mathbf{D}_1$ by $\mathbf{D}_1^p=span\{\vec{\boldsymbol{\alpha}}_1,\vec{\boldsymbol{\alpha}}_2,\ldots,\vec{\boldsymbol{\alpha}}_r\}$ and represent the null space of $\mathbf{D}_1$ by $\mathbf{D}_1^q=span\{\vec{\boldsymbol{\alpha}}_{r+}1,\vec{\boldsymbol{\alpha}}_{r+}2,\ldots,\vec{\boldsymbol{\alpha}}_N\}$. If $\mathbf{z}$ is from $\mathbf{D}_1^q$, i.e., $\mathbf{z}\in\mathbf{D}_1^q$, we obtain $\mathbf{z}^T\mathbf{D}_1\mathbf{z}=0$. This is because if $\mathbf{z}\in\mathbf{D}_1^q$, then $\mathbf{z}$ can be expressed as a linear combination of the basis of $\mathbf{D}_1^q$, $\vec{\boldsymbol{\alpha}}_{r+}1,\vec{\boldsymbol{\alpha}}_{r+}2,\ldots,\vec{\boldsymbol{\alpha}}_N$. That is, we have $\mathbf{z}=\sum_{i=1}^{N-r}\gamma_i\vec{\boldsymbol{\alpha}}_{i+}r$ and $\mathbf{z}^T\mathbf{D}_1\mathbf{z}=\sum_{i=1}^{N-r}\sum_{j=1}^{N-r}\gamma_i\gamma_j\vec{\boldsymbol{\alpha}}_{i+}r^T\mathbf{D}_1\vec{\boldsymbol{\alpha}}_{j+}r$. Since $\vec{\boldsymbol{\alpha}}_{j+}r$ is the eigenvector corresponding to a zero eigenvalue of $\mathbf{D}_1$, $\mathbf{D}_1\vec{\boldsymbol{\alpha}}_{j+}r=0$ is satisfied for each $\vec{\boldsymbol{\alpha}}_{j+}r$ and consequently $\mathbf{z}^T\mathbf{D}_1\mathbf{z}=0$. Therefore, in this case, it is not guaranteed that the $\mathbf{z}^T\mathbf{L}_1\mathbf{z}/\mathbf{z}^T\mathbf{D}_1\mathbf{z}$ has the minimum value. On the other hand, if $\mathbf{z}\in\mathbf{D}_1^p$, we can conclude that $\mathbf{z}^T\mathbf{D}_1\mathbf{z}>0$. This is because $\mathbf{z}\in\mathbf{D}_1^p$ means $\mathbf{z}=\sum_{i=1}^{r}\eta_i\vec{\boldsymbol{\alpha}}_i$ and

$$\mathbf{z}^T\mathbf{D}_1\mathbf{z}=\sum_{i=1}^{r}\sum_{j=1}^{r}\eta_i\eta_j\vec{\boldsymbol{\alpha}}_i^T\mathbf{D}_1\vec{\boldsymbol{\alpha}}_j=\sum_{i=1}^{r}\sum_{j=1}^{r}\lambda_j\eta_i\eta_j\vec{\boldsymbol{\alpha}}_i^T\vec{\boldsymbol{\alpha}}_j=\sum_{i=1}^{r}\lambda_i\eta_i^2$$

where $\lambda_i$ is the eigenvalue corresponding to the eigenvector $\vec{\boldsymbol{\alpha}}_i$ of $\mathbf{D}_1$, $i=1,\ldots,r$. We note that $\lambda_i>0$ and there is at least one non-zero $\eta_i$; therefore, $\mathbf{z}^T\mathbf{D}_1\mathbf{z}=\sum_{i=1}^{r}\lambda_i\eta_i^2>0$. Thus, in the case where $\mathbf{z}$ is from $\mathbf{D}_1^p$, we are able to obtain small $\mathbf{z}^T\mathbf{L}_1\mathbf{z}/\mathbf{z}^T\mathbf{D}_1\mathbf{z}$. As a result, we can conclude that effective solutions of the conventional LPP solution scheme 1 should be from the range space of $\mathbf{D}_1$ rather than from its null space.

We design the improved solution scheme 1 as follows. We first define that $\mathbf{R}=[\vec{\boldsymbol{\alpha}}_1\ \vec{\boldsymbol{\alpha}}_2\ \ldots\ \vec{\boldsymbol{\alpha}}_r]$ and $\mathbf{S}=[\vec{\boldsymbol{\alpha}}_{r+}1\ \vec{\boldsymbol{\alpha}}_{r+}2\ \ldots\ \vec{\boldsymbol{\alpha}}_N]$. Using $\mathbf{R}$, we respectively transform $\mathbf{D}_1$, $\mathbf{L}_1$ and $\mathbf{W}_1$ into the following matrices: $\overline{\mathbf{D}}=\mathbf{R}^T\mathbf{D}_1\mathbf{R}$, $\overline{\mathbf{L}}=\mathbf{R}^T\mathbf{L}_1\mathbf{R}$ and $\overline{\mathbf{W}}=\mathbf{R}^T\mathbf{W}_1\mathbf{R}$. We then construct the following eigen-equation:

$$\overline{\mathbf{L}}\overline{\mathbf{z}}=\lambda\overline{\mathbf{D}}\overline{\mathbf{z}}. \tag{17}$$

We can directly solve this equation since $\overline{\mathbf{D}}$ is of full rank. Let $\rightarrow\vec{\boldsymbol{\beta}}_1,\vec{\boldsymbol{\beta}}_2,\ldots,\vec{\boldsymbol{\beta}}_r$ denote the eigenvectors corresponding to eigenvalues $\lambda_1,\lambda_2,\ldots,\lambda_r$ in the increasing order of Eq. (17). If the sample is required to be transformed into an $m$-dimensional space, we should select $\vec{\boldsymbol{\beta}}_1,\vec{\boldsymbol{\beta}}_2,\ldots,\vec{\boldsymbol{\beta}}_m$ as the $m$ transforming axes of LPP. We refer to this solution scheme as the improved LPP solution scheme 1.

*3.1.2. Transform procedure of the improved LPP solution scheme 1*

The transform procedure of the improved LPP solution scheme 1 is as follows. Using the matrix $\mathbf{R}$, we first produce $\mathbf{c}=\mathbf{x}^T\mathbf{R}$ from a sample $\mathbf{x}$. Then we transform $\mathbf{c}$ into $\mathbf{y}$ by carrying out $\mathbf{y}=\mathbf{cG}$, where $\mathbf{G}=[\vec{\beta}_1\ \vec{\beta}_2\ \dots\ \vec{\beta}_m]$. For simplicity of expression, the whole transform procedure can also be implemented and denoted by using $\mathbf{y}=\mathbf{x}^T\mathbf{RG}$.

### 3.2. Improved LPP solution scheme 2

In this subsection, we describe the improved LPP solution scheme 2, an improvement to the conventional LPP solution scheme 2. The improved LPP solution scheme 2 uses two separate procedures to perform two transforms, respectively. The first procedure is based on the following eigen-equation:

$$\overline{\mathbf{W}}\,\overline{\mathbf{z}}=\lambda\overline{\mathbf{D}}\overline{\mathbf{z}}, \tag{18}$$

where $\overline{\mathbf{W}}$ and $\overline{\mathbf{D}}$ are defined in Section 3.1. We can directly solve Eq. (18) since $\overline{\mathbf{D}}$ is of full rank. Let $\rightarrow\varsigma_1,\vec{\varsigma}_2,\dots,\vec{\varsigma}_r$ denote the eigenvectors corresponding to non-increasing eigenvalues $\lambda_1$, $\lambda_2$, …, $\lambda_r$ of Eq. (18). According to the goal of LPP, the eigenvectors corresponding to maximum eigenvalues of (18) should be used as transforming axes. As a result, if we select $\vec{\varsigma}_1,\vec{\varsigma}_2,\dots,\vec{\varsigma}_d$ as the $d$ transforming axes, we can produce $d$-dimensional features. The feature extraction procedure is presented as follows: let $\mathbf{U}=[\vec{\varsigma}_1\ \vec{\varsigma}_2\ \dots\ \vec{\varsigma}_d]$. Then the $d$-dimensional features of a sample $\mathbf{x}$ are obtained by the transform $\overline{\mathbf{y}}=\mathbf{x}^T\mathbf{RU}$.

The second procedure is as follows: Using $\mathbf{S}$, we transform $\mathbf{W}_1$ and $\mathbf{D}_1$ into $\tilde{\mathbf{W}}=\mathbf{S}^T\mathbf{W}_1\mathbf{S}$ and $\tilde{\mathbf{D}}=\mathbf{S}^T\mathbf{D}_1\mathbf{S}$, respectively, where $\mathbf{S}$ is also defined as $\mathbf{S}=[\vec{\alpha}_{r+}1\ \vec{\alpha}_{r+}2\ \dots\ \vec{\alpha}_N]$. In the transform space obtained using $\mathbf{S}$, $\tilde{\mathbf{z}}^T\tilde{\mathbf{D}}\ \tilde{\mathbf{z}}=0$ is satisfied for arbitrary $\tilde{\mathbf{z}}$. Therefore, in this space, the ratio $\tilde{\mathbf{z}}^T\tilde{\mathbf{W}}\ \tilde{\mathbf{z}}$ of $\tilde{\mathbf{z}}^T\tilde{\mathbf{D}}\ \tilde{\mathbf{z}}$ will be the infinity. According to the nature of the conventional LPP solution scheme 2, the $\tilde{\mathbf{z}}$ corresponding to large $\tilde{\mathbf{z}}^T\tilde{\mathbf{W}}\ \tilde{\mathbf{z}}$ is helpful for preserving the local structure of samples. Therefore, we can take as transforming axes of LPP the eigenvectors corresponding to maximum eigenvalues of the following eigen-equation:

$$\tilde{\mathbf{W}}\,\tilde{\mathbf{z}}=\lambda\tilde{\mathbf{z}}. \tag{19}$$

Suppose that $e$ eigenvectors $\vec{\xi}_1,\vec{\xi}_2,\dots,\vec{\xi}_e$ of Eq. (19) are used as transforming axes, and then the feature extraction result is $\tilde{\mathbf{y}}=\mathbf{x}^T\mathbf{SV}$, where $\mathbf{V}=[\vec{\xi}_1\ \vec{\xi}_2\ \dots\ \vec{\xi}_e]$. Hence, the final feature extraction result of a sample $\mathbf{x}$ obtained using the improved LPP solution scheme 2 is the following $(d+e)$-dimensional row vector:

$$\mathbf{y}=[\overline{\mathbf{y}}\tilde{\mathbf{y}}]. \tag{20}$$

### 3.3. Improved LPP solution scheme 3

In this subsection, we will describe the improved LPP solution scheme 3. We can also improve the conventional LPP solution scheme 2 in the following way: above of all, the objective function (13) is partially equivalent to the following one:

$$\arg\max(\mathbf{z}^T\mathbf{XWX}^T\mathbf{z}-u\mathbf{z}^T\mathbf{XDX}^T\mathbf{z}), \tag{21}$$

i.e.,

$$\arg\max\mathbf{z}^T\mathbf{X}(\mathbf{W}-u\mathbf{D})\mathbf{X}^T\mathbf{z}, \tag{22}$$

where $u$ is a positive constant. As a consequence, we can define the following eigen-equation:

$$\mathbf{X}(\mathbf{W}-u\mathbf{D})\mathbf{X}^T\mathbf{z}=\lambda\mathbf{z}. \tag{23}$$

We can implement the LPP transform by taking a number of eigenvectors corresponding to maximum eigenvalues of Eq. (23) as transforming axes. We refer to this scheme of solving the LPP transforming axes as the improved LPP solution scheme 3. In this scheme, the strategy of improving conventional LPP is to convert the objective function $\arg\max(\mathbf{z}^T\mathbf{XWX}^T\mathbf{z}/\mathbf{z}^T\mathbf{XDX}^T\mathbf{z})$ to (21). In other words, the improved LPP solution scheme 3 improves the conventional LPP solution scheme 2 by replacing the objective function in the division form with a function in the subtraction form. It is clear that (23) can be directly solved in all cases including the SSS case and the improved LPP solution scheme 3 has a lower computational cost than the improved LPP solution schemes 1 and 2. It should be pointed out that a similar strategy has been used to improve the Fisher discriminant analysis [36].

## 4. Experimental design and results

We used the ORL, AR, and FERET face image databases to test the proposed new LPP solution schemes and the conventional LPP solution schemes 1 and 2. As conventional LPP is an unsupervised method, all these solution schemes are implemented as unsupervised schemes to ensure a fair comparison. Since the $\mathbf{XDX}^T$ in all of the experiments on the conventional LPP solution schemes 1 and 2 is singular, we replaced $(\mathbf{XDX}^T)^{-1}$ with the pseudo-inverse of $\mathbf{XDX}^T$. We also slightly modified the conventional LPP solution scheme 1 as follows: when the eigen-equation has zero eigenvalues, we discarded zero eigenvalues and the corresponding eigenvectors and took only the eigenvectors corresponding to the non-zero smallest eigenvalues as transforming axes. We refer to this modified scheme as modified conventional solution scheme 1. When implementing conventional solution scheme 1 and modified conventional scheme 1, we regard eigenvalues that are less than $0.2\times10^{-11}$ as zero eigenvalues. We used the $k$-nearest-neighbor classifier (KNNC) to classify the transform results of the samples. In the experiment on the improved solution scheme 3 on the ORL face database, the parameter $u$ was set to 0.8. In the experiment on the improved solution scheme 3 on the AR and FERET face databases, the parameter $u$ was set to 0.02. In all experiments, the parameter $t$ in the definition of $w_{ij}$ is defined as the maximum value of the distances between arbitrary two training samples.

### 4.1. Experiment on the ORL face database

In the experiment on the ORL database (http://www.cam-orl.co.uk), all the face images in the ORL face database were captured against a dark homogeneous background. These images contain various facial expressions (smiling/not smiling, open/closed eyes) and facial details. The subjects were in an upright, frontal position but there was a tolerance for some tilting and rotation of up to about 20° [37]. Ten different images were obtained for each of the 40 subjects. The first five face images of each subject were used as training samples, and the remaining images were used as testing samples. In order to implement a LPP algorithm at a lower computational cost, we resized each original $112\times92$ image to $56\times46$ using the down-sampling method proposed in [38].

Fig. 1 shows the classification accuracies of six LPP solution schemes on the ORL face database. It describes the variation of the classification accuracy of each solution scheme with the number of the used transforming axes. Fig. 1(a), (b), (c), (d), (e) and (f), respectively, show the classification right rates of conventional solution scheme 1 (i.e., conventional 1), conventional solution scheme 2 (i.e., conventional 2), modified conventional solution

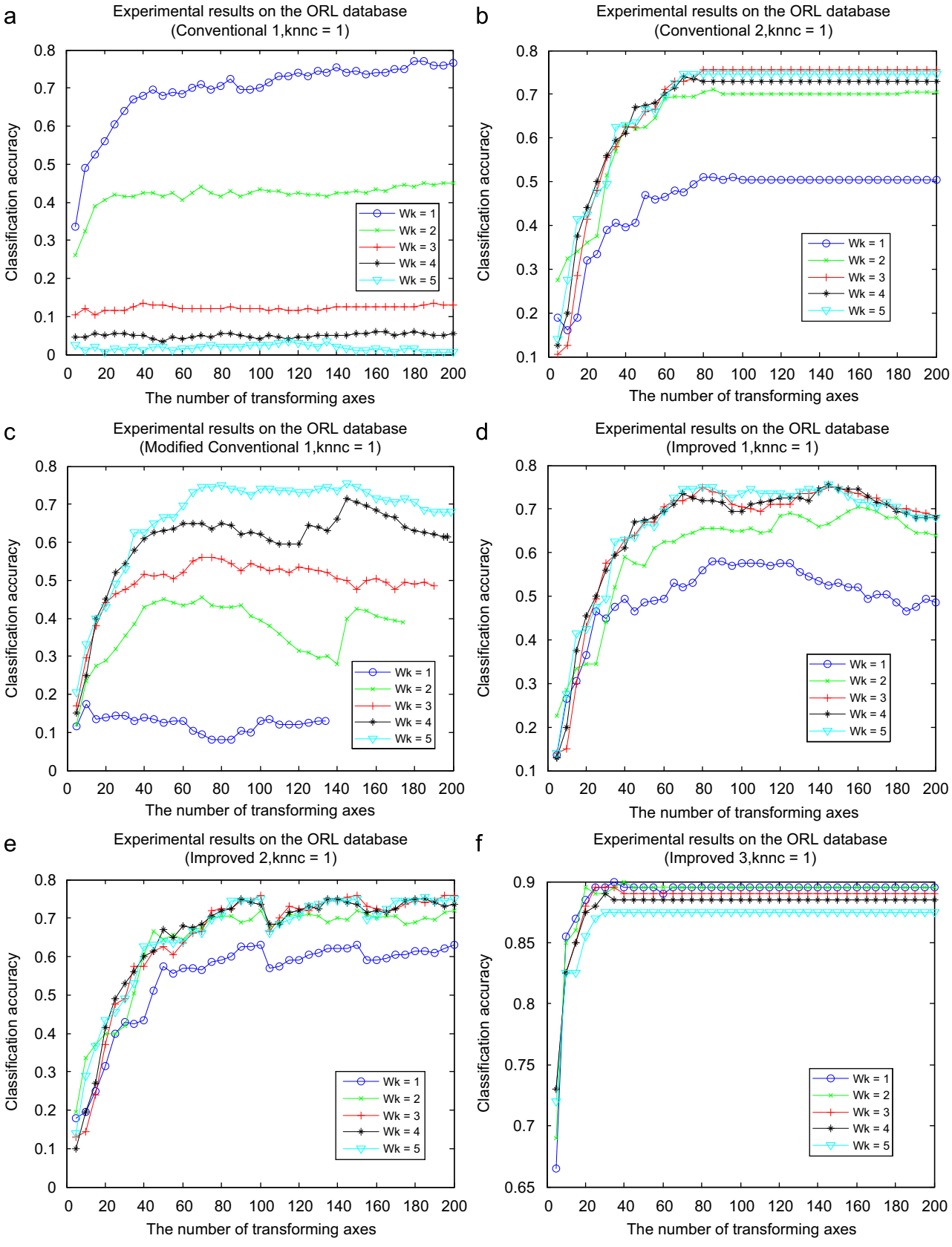


**Fig. 1.** The classification right rates (classification accuracies) of six LPP solution schemes on the ORL face database.

scheme 1 (i.e., modified conventional 1), the improved solution scheme 1 (i.e., improved 1), the improved solution scheme 2 (i.e., improved 2) and the improved solution scheme 3 (i.e., improved 3). *knnc*=1 means that the classifier is the 1-nearest-neighbor classifier. We used *Wk* to indicate how many neighbors were taken into account when we calculated the weight matrix. *Wk*=1, 2, 3, 4 and 5 mean that we define one, two, three, four and five neighbor samples for a sample when calculating the weight matrix, respectively. For example, 1 means that when calculating $w_{ij}$, we set the value of the parameter *Wk* in $w_{ij}$ to 1.

We can see that improved 3 is the most accurate of the solution schemes and that when more than 80 transforming axes are used, it almost always achieves close to or even the highest accuracy. However, the accuracies of the other schemes vary with the number of transforming axes. The value of *Wk* affects the classification accuracies of the different LPP solution schemes in various ways. For example, the classification accuracy of conventional 1 rapidly decreases as the value of *Wk* increases. However, the classification accuracy of modified conventional 1 increases as the value of *Wk* increases.

Table 1 shows the classification accuracies of different solution schemes with optimal settings of *Wk* and *knnc* on the ORL database. Fig. 2 shows the performance comparison of different solution schemes with optimal settings of *Wk* and *knnc* on the ORL database. The optimal settings of *Wk* and *knnc* mean that under these settings of *Wk* and *knnc*, the solution scheme achieves the best classification performance. We also see that improved 3 performs best.

**Table 1**
The classification accuracies of different solution schemes with optimal settings of *Wk* and *knnc* on the ORL database.

| | 5 | 20 | 40 | 60 | 80 | 100 | *Wk* | *knnc* |
|---|---|---|---|---|---|---|---|---|
| Conventional 1 | 0.31 | 0.59 | 0.70 | 0.71 | 0.74 | 0.74 | 1 | 3 |
| Conventional 2 | 0.11 | 0.42 | 0.63 | 0.71 | 0.76 | 0.76 | 3 | 1 |
| Modified conventional 1 | 0.21 | 0.43 | 0.63 | 0.70 | 0.75 | 0.74 | 5 | 1 |
| Improved 1 | 0.14 | 0.43 | 0.63 | 0.70 | 0.75 | 0.74 | 5 | 1 |
| Improved 2 | 0.14 | 0.44 | 0.63 | 0.64 | 0.71 | 0.75 | 5 | 1 |
| Improved 3 | 0.69 | 0.90 | 0.90 | 0.90 | 0.90 | 0.90 | 2 | 1 |

In the first row, 5, 20, 40, 60, 80 and 100 denote how many transforming axes were used. The shown values of *Wk* and *knnc* are the ones that enable a solution scheme to produce the best performance.

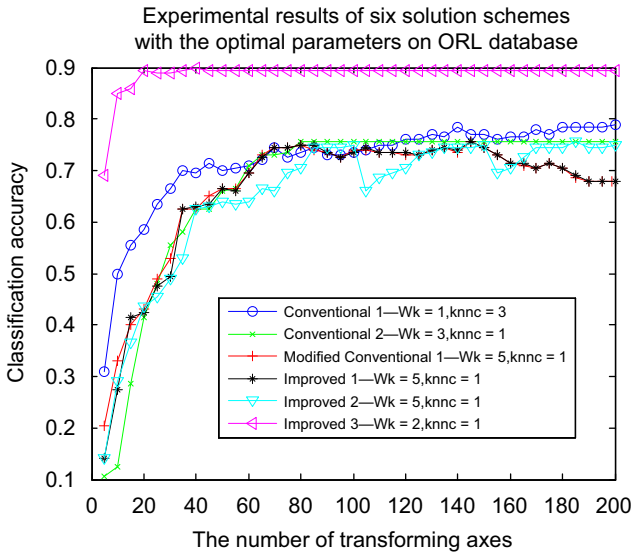


**Fig. 2.** Comparison of the performance of different solution schemes with optimal settings of *Wk* and *knnc* on the ORL database. The label of the solution scheme shows both the optimal values of *Wk* and *knnc* and the names of the solution schemes.

### 4.2. Experiment on the AR face database

The AR face database [39,40] contains over 4000 gray face images of 126 people. The images of each subject have different facial expressions, and were acquired under lighting conditions and with and without occlusions. Each subject provided 26 face images. We note that 12 face images of each subject are occluded with sunglasses or a scarf. The face images of 120 subjects were taken in two sessions. We used only the images of these 120 subjects in our experiment. We manually cropped the face portion of every image and then normalized them to $50 \times 40$ pixels [41]. Fig. 3 shows the normalized images of one subject. We used only the 14 non-occluded face images of each subject to test different solution schemes. The first and eighth images were used as training samples and the remaining images were used as testing samples.


**Fig. 3.** Normalized non-occluded face images with $50 \times 40$ pixels of one subject. (a) neutral expression, (b) smile, (c) anger, (d) scream, (e) left light on, (f) right light on, (g) all sides light on. (h), (i), (j), (k), (l), (m) and (n) were taken in the second session under the same conditions as (a), (b), (c), (d), (e), (f) and (g), respectively.

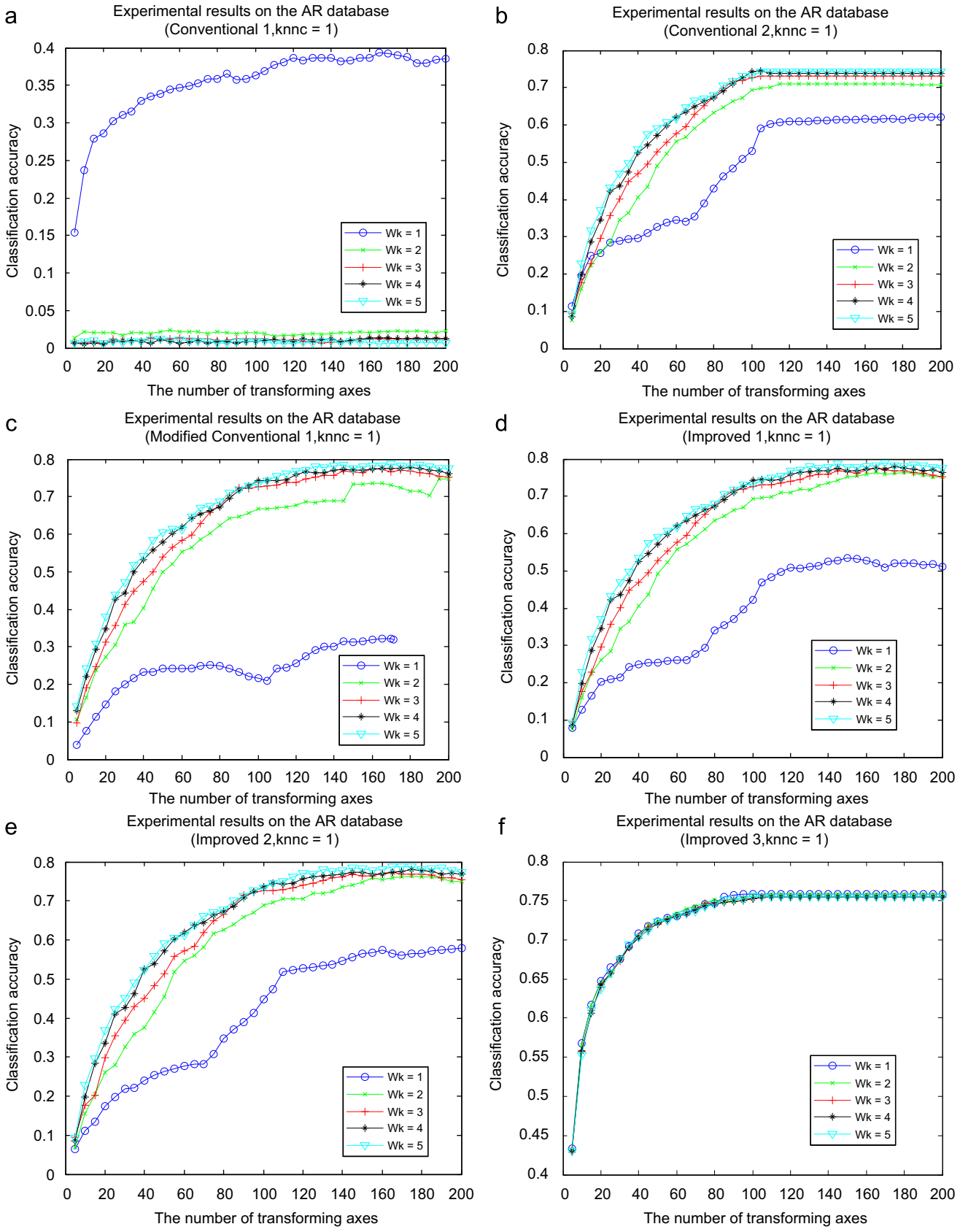


**Fig. 4.** The classification accuracies of six LPP solution schemes on the AR face database.

**Table 2**
The classification accuracies of different solution schemes with optimal settings of *Wk* and *knnc* on the AR database.

| | 5 | 20 | 40 | 60 | 80 | 100 | *Wk* | *knnc* |
|---|---|---|---|---|---|---|---|---|
| Conventional 1 | 0.16 | 0.31 | 0.34 | 0.37 | 0.38 | 0.39 | 1 | 2 |
| Conventional 2 | 0.10 | 0.37 | 0.54 | 0.62 | 0.68 | 0.73 | 5 | 1 |
| Modified conventional 1 | 0.14 | 0.38 | 0.54 | 0.61 | 0.69 | 0.73 | 5 | 1 |
| Improved 1 | 0.10 | 0.37 | 0.54 | 0.62 | 0.68 | 0.73 | 5 | 1 |
| Improved 2 | 0.10 | 0.37 | 0.52 | 0.61 | 0.67 | 0.73 | 5 | 1 |
| Improved 3 | 0.43 | 0.65 | 0.71 | 0.73 | 0.75 | 0.76 | 1 | 1 |

In the first row, 5, 20, 40, 60, 80 and 100 denote the number of the used transforming axes. The shown values of *Wk* and *knnc* are the ones that enable a solution scheme to produce the best performance.

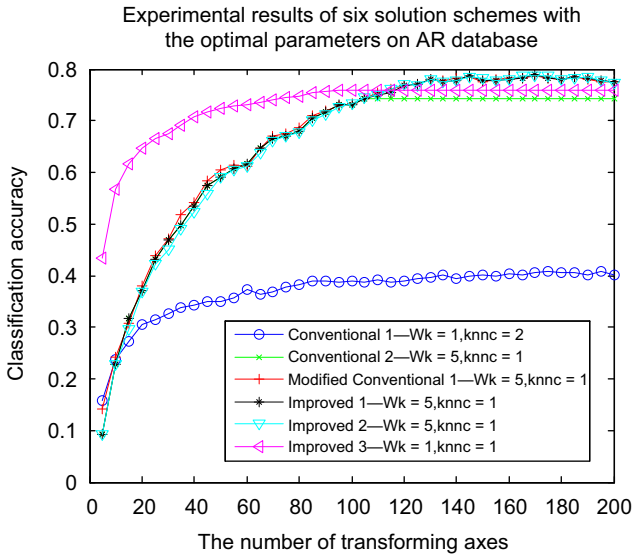


**Fig. 5.** The classification performance comparison of different solution schemes with respective optimal settings of *Wk* and *knnc* on the AR database. The optimal values of *Wk* and *knnc* are shown together with the solution scheme names in the label of the solution scheme.

Fig. 4 shows the classification accuracies of six LPP solution schemes on the AR face database. We can see that different values of *Wk* have obvious effects on the classification accuracies of all the LPP solution schemes except for improved 3. When $Wk=1$ conventional 1 obtains its highest classification accuracy. However, when $Wk=1$, conventional 2, modified conventional 1, improved 1 and improved 2 all obtain their lowest classification accuracies. In contrast, improved 3 with different values of *Wk* produces almost the same accuracies.

Table 2 shows the classification accuracies of different solution schemes with the respective optimal settings of *Wk* and *knnc* on the AR database. Fig. 5 compares the classification performance of different solution schemes with optimal settings of *Wk* and *knnc* on the AR database. Conventional 1 obtains the worst classification performance. If the solution schemes all use less than 100 transforming axes, improved 3 will produce much higher classification accuracy than the other solution schemes. When the number of the transforming axes used is larger than 100, the accuracy of improved 3 is similar to the highest accuracy of the other solution schemes, except conventional 1. Improved 1, improved 2, modified conventional 1 and conventional 2 also produce a similar optimal classification performance.

### 4.3. Experiment on the FERET face database

The FERET face image database was sponsored by the US Department of Defense through the DARPA Program [42,43]. This database has been widely used for testing and evaluating state-of-the-art face recognition algorithms. We conducted experiment on a subset of the FERET database. This subset includes 1,400 images of 200 individuals each having seven face images. It is composed of images named with two character strings: "ba," "bj," "bk," "be", "bf," "bd," and "bg" [44]. The images in this subset involve variations in facial expression, illumination, and pose. The facial portion of each original image was automatically cropped based on the location of eyes and the cropped image was resized to $80\times80$ pixels [41]. We used the down-sampling method in [37] to produce $40\times40$ face images. We took "ba" and "be" as training samples and used the remainder as testing samples.

Fig. 6 shows the classification accuracies of six LPP solution schemes with different values of *Wk* on the FERET face database. We see again that different values of *Wk* have little effect on the classification accuracy of improved 3. In addition, conventional 1 and improved 2 both obtain their highest classification accuracies when $Wk=1$. In contrast, the other solution schemes except for improved 3 obtain their lowest classification accuracies when $Wk=1$.

We use Fig. 7 and Table 3 to show the classification performance comparison of different solution schemes with optimal settings of *Wk* and *knnc* on the FERET database. Improved 3 is definitely more accurate than the other solution schemes. In contrast, conventional 1 obtains the lowest accuracy among all the solution schemes. The classification accuracy of improved 2 is slightly lower than improved 3 but higher than the other solution schemes.

## 5. Conclusion

When conventional LPP is applied to image data, it will suffer from several problems including the SSS problem and the problem that it does not preserve the locality structure. The three solution schemes proposed in this paper have clear advantages. First, as the proposed solution schemes do not suffer from the matrix singularity problem, they are all directly applicable to the SSS problem. Second, experimental results show that at least one of the proposed solution schemes can outperform conventional LPP in classification accuracy. It is especially noticeable that improved 3 is not only computationally efficient, but also classifies quite accurately. Moreover, it usually produces a stable high classification accuracy, little influenced by different values of the parameter *Wk*. In addition, the proposed LPP solution schemes have solid theoretical foundations.

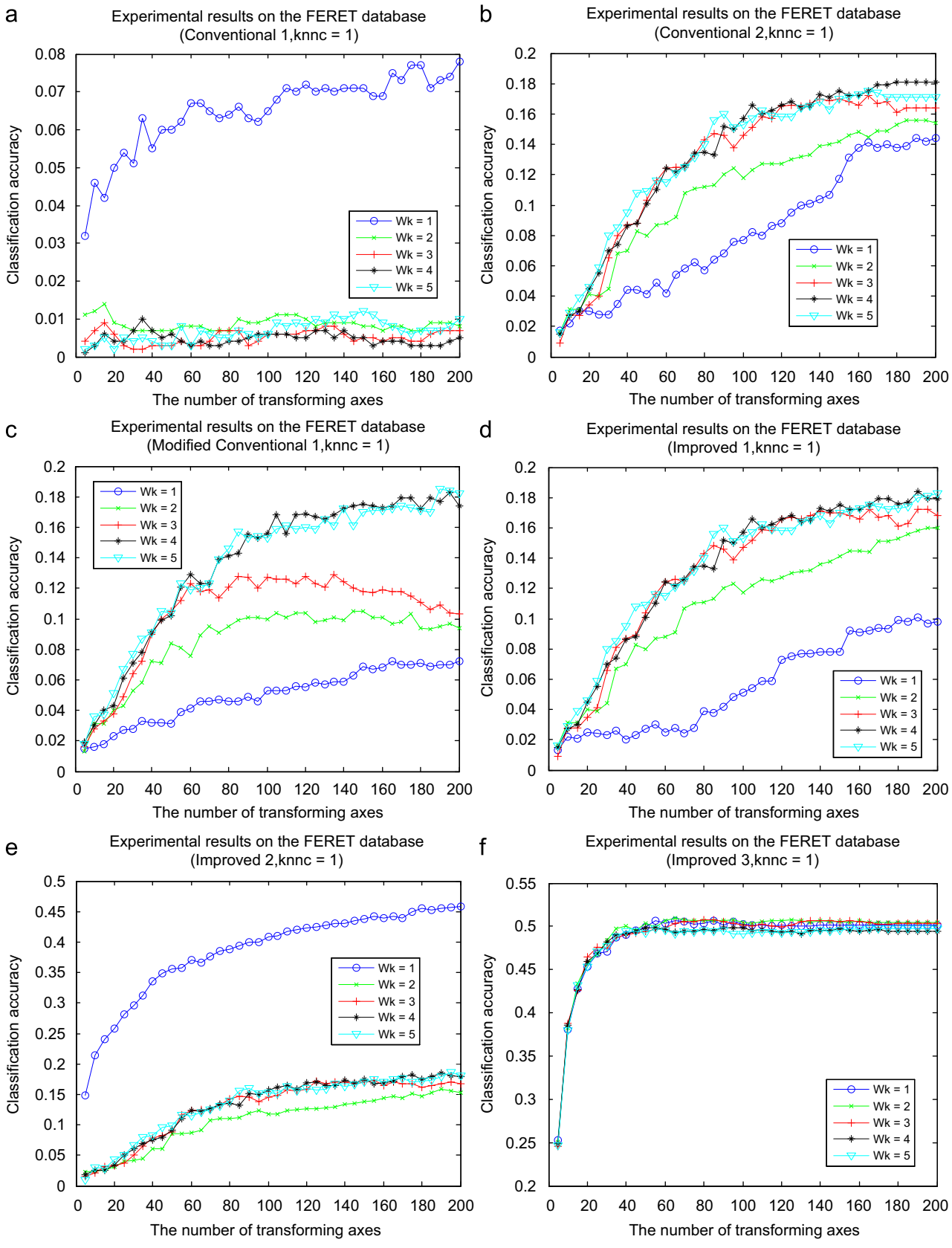


**Fig. 6.** The classification accuracies of six LPP solution schemes on the FERET face database.

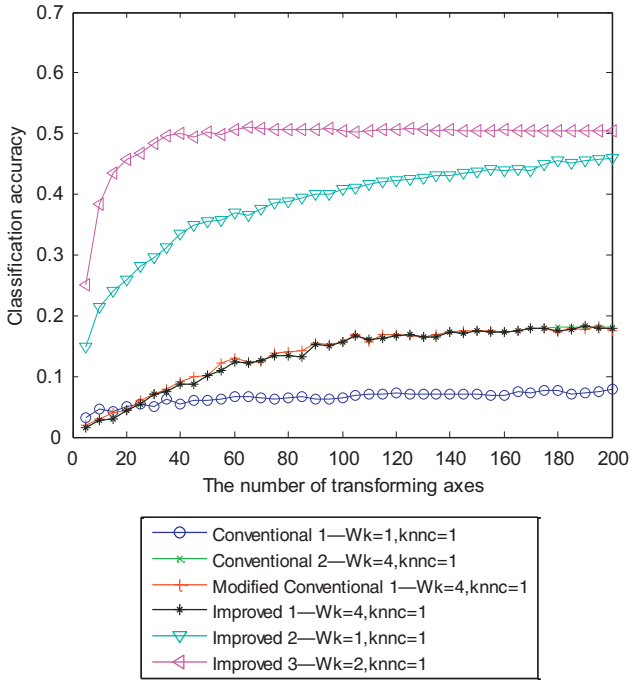


**Fig. 7.** Comparison of the classification performance of different solution schemes with optimal settings of *Wk* and *knnc* on the FERET database. The optimal values of *Wk* and *knnc* are shown together with the solution scheme names in the label of the solution scheme.

**Table 3**
The classification accuracies of different solution schemes with optimal settings of *Wk* and *knnc* on the FERET database.

| | 5 | 20 | 40 | 60 | 80 | 100 | *Wk* | *knnc* |
|---|---|---|---|---|---|---|---|---|
| Conventional 1 | 0.032 | 0.05 | 0.06 | 0.07 | 0.06 | 0.07 | 1 | 1 |
| Conventional 2 | 0.02 | 0.05 | 0.09 | 0.14 | 0.14 | 0.16 | 4 | 1 |
| Modified conventional 1 | 0.02 | 0.04 | 0.09 | 0.13 | 0.14 | 0.16 | 4 | 1 |
| Improved 1 | 0.02 | 0.05 | 0.09 | 0.12 | 0.14 | 0.16 | 4 | 1 |
| Improved 2 | 0.15 | 0.26 | 0.34 | 0.37 | 0.39 | 0.41 | 1 | 1 |
| Improved 3 | 0.25 | 0.46 | 0.50 | 0.51 | 0.51 | 0.50 | 2 | 1 |

In the first row, 5, 20, 40, 60, 80 and 100 denote the number of the transforming axes used. The values of *Wk* and *knnc* are those that enable a solution scheme to produce the best performance.

## Acknowledgments

This work was partially supported by the Program for New Century Excellent Talents in University (NCET-08-0156), the Fundamental Research Funds for the Central Universities (grant no. HIT.NSRIF.2009130), the NSFC under grant nos. 60973098 and 60632050, the National High-Tech Research and Development Plan of China (863) under contract no. 2007AA01Z195 as well as by the CERG fund from the HKSAR Government and the central fund from Hong Kong Polytechnic University.

## References

[1] H.S. Seung, D.D. Lee, The manifold ways of perception, Science 290 (2000).
[2] J.B. Tenenbaum, V. de Silva, J.C. Langford, A global geometric framework for nonlinear dimensionality reduction, Science 290 (2000).
[3] A. Shashua, A. Levin, S. Avidan, Manifold pursuit: a new approach to appearance based recognition, in: Proceedings of the International Conference on Pattern Recognition, August 2002.
[4] Y. Chang, C. Hu, M. Turk, Manifold of facial expression, in: Proceedings of the IEEE International Workshop Analysis and Modeling of Faces and Gestures, October 2003.
[5] K.-C. Lee, J. Ho, M.-H. Yang, D. Kriegman, Video-based face recognition using probabilistic appearance manifolds, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, vol. 1, 2003, pp. 313–320.
[6] S.T. Roweis, L.K. Saul, Nonlinear dimensionality reduction by locally linear embedding, Science 290 (2000).
[7] S. Roweis, L. Saul, and G. Hinton, Global coordination of local linear models, in: Proceedings of the Conference on Advances in Neural Information Processing System, vol. 14, 2001.
[8] Fan R.K. Chung , Spectral Graph Theory, Regional Conference Series in Mathematics, number 92, 1997.
[9] S. Yan, D. Xu, B. Zhang, H. Zhang, Graph embedding: a general framework for dimensionality reduction, CVPR 2 (2005) 830–837.
[10] M. Belkin, P. Niyogi, Laplacian eigenmaps and spectral techniques for embedding and clustering, Proceedings of the Conference on Advances in Neural Information Processing Systems (2001).
[11] J. Yang, D. Zhang, J.-Y. Yang, B. Niu, Globally maximizing, locally minimizing: unsupervised discriminant projection with applications to face and palm biometrics, IEEE Trans. Pattern Anal. Mach. Intell. 29 (4) (2007) 650–664.
[12] X. He, P. Niyogi, Locality preserving projections, Proceedings of the Conference on Advances in Neural Information Processing Systems (2003).
[13] X. He, S. Yan, et al., Face recognition using Laplacianfaces, IEEE Trans. Pattern Anal. Mach. Intell. 27 (3) (2005) 328–340.
[14] W. Min, K. Lu, X. He, Locality preserving projection, Pattern Recognition J 37 (4) (2004) 781–788.
[15] X. He, D. Cai, W. Min, Statistical and computational analysis of locality preserving projection, in: Proceedings of the 22nd International Conference on Machine Learning, 2005, pp. 281–288.
[16] X. He, S. Yan, Y. Hu, H. Zhang, Learning a locality preserving subspace for visual recognition, ICCV (2003) 385–393.
[17] Z. Zheng, F. Yang, W. Tan, J. Jia, J. Yang, Gabor feature-based face recognition using supervised locality preserving projection, Signal Processing 87 (2007) 2473–2483.
[18] X. He, Incremental semi-supervised subspace learning for image retrieval, ACM Multimedia (2004) 2–8.
[19] J. Cheng, Q. Liu, H. Lu, Y.-W. Chen, Supervised kernel locality preserving projections for face recognition, Neurocomputing 67 (2005) 443–449.
[20] H.-T. Chen, H.-W. Chang, T.-L. Liu., Local discriminant embedding and its variants, Proc. CVPR (2005).
[21] S. Yan, D. Xu, B. Zhang, H.-J. Zhang, Q. Yang, S. Lin, Graph embedding and extension: a general framework for dimensionality reduction, IEEE Trans. Pattern Anal. Mach. Intell. 29 (1) (2007).
[22] G. Feng, D. Hu, D. Zhang, Z. Zhou, An alternative formulation of kernel LPP with application to image recognition, Neurocomputing 69 (13–15) (2006) 1733–1738.
[23] J.-B. Li, J.-S. Pan, Shu-Chuan Chu, Kernel class-wise locality preserving projection, Inform. Sci. 178 (2008) 1825–1835.
[24] W. Zheng, J.-H. Lai, Regularized locality preserving learning of pre-image problem in kernel principal component analysis, ICPR 2 (2006) 456–459.
[25] D. Hu, G. Feng, Z. Zhou, Two-dimensional locality preserving projections (2DLPP) with its application to palmprint recognition, Pattern Recognition 40 (1) (2007) 339–342.
[26] R. Zhi, Q. Ruan, Facial expression recognition based on two-dimensional discriminant locality preserving projections, Neurocomputing 71 (2008) 1730–1734.
[27] D. Cai, X. He, Orthogonal locality preserving indexing, SIGIR (2005) 3–10.
[28] Haitao Zhao, Shaoyuan Sun, Zhongliang Jing, Local-information-based uncorrelated feature extraction, Opt. Eng. 45 (2) (2006).
[29] Y. Xu, F. Song, G. Feng, Y. Zhao, A novel local preserving projection scheme for use with face recognition, Expert Syst Appl 37 (9) (2010) 6718–6721.
[30] Y. Xu, D. Zhang, Use of unsupervised locality preserving projection method in the SSS problem and a new solution scheme, in: Proceedings of 2008 International Conference on Pattern Recognition.
[31] D. Cai, X. He, Y. Hu, J. Han, Thomas S. Huang: learning a spatially smooth subspace for face recognition. CVPR 2007.
[32] X. He, Deng Cai, Haifeng Liu, Wei-Ying Ma, Locality preserving indexing for document representation, SIGIR (2004) 96–103.
[33] D. Cai, X. He, J. Han, Document clustering using locality preserving indexing, IEEE Trans. Knowl. Data Eng. 17 (12) (2005) 1624–1637.
[34] Y. Xu, D. Zhang, Z. Jin, M. Li, J.-Y. Yang, A fast kernel-based nonlinear discriminant analysis for multi-class problems, Pattern Recognition 39 (6) (2006) 1026–1033.
[35] Y. Xu, J.-Y. Yang, J. Yang, A reformative kernel Fisher discriminant analysis, Pattern Recognition 37 (6) (2004) 1299–1302.
[36] F. Song, D. Zhang, Dg Mei, Z. Guo, A multiple maximum scatter difference discriminant criterion for facial feature extraction, IEEE Trans. Systems Man Cybernet. Pt. B 37 (6) (2007) 1599–1606.
[37] Y. Xu, D. Zhang, J.-Y. Yang, A feature extraction method for use with bimodal biometrics, Pattern Recognition 43 (2010) 1106–1115.
[38] Y. Xu, Z. Jin, Down-sampling face images and low-resolution face recognition, The third international conference on innovative computing, information and control, June 18–20, 2008, Dalian, China, pp. 392–395.

[39] A.M. Martinez and R. Benavente, The AR face database CVC Technical Report, no. 24, June 1998.
[40] A.M. Martinez and R. Benavente, The AR Face Database, http://rvl1.ecn.purdue.edu/~aleix/aleix_face_DB.html, 2003.
[41] J. Yang, D. Zhang, A.F. Frangi, J.Y. Yang, Two dimensional PCA: a new approach to appearance-based face representation and recognition, IEEE Trans. Pattern Anal. Machine Intell. 24 (1) (2004) 131–137.
[42] P.J. Phillips, H. Moon, S.A. Rizvi, P.J. Rauss, The FERET evaluation methodology for face-recognition algorithms, IEEE Trans. Pattern Anal. Machine Intell. 22 (10) (2000) 1090–1104.
[43] P.J. Phillips, The Facial Recognition Technology (FERET) Database, 〈http://www.itl.nist.gov/iad/humanid/feret/feret_master.html〉, 2004.
[44] J. Yang, J.-y Yang, A.F. Frangi, Combined Fisherfaces framework, Image Vision Comput. 21 (12) (2003) 1037–1044.

**Yong Xu** received his B.S. and M.S. degrees at Air Force Institute of Meteorology (China) in 1994 and 1997, respectively. He then received his Ph.D. degree in pattern recognition and intelligence system at the Nanjing University of Science and Technology (NUST) in 2005. From May 2005 to April 2007, he worked at Shenzhen graduate school, Harbin Institute of Technology (HIT) as a postdoctoral research fellow. Now he is an associate professor at Shenzhen graduate school, HIT. He also acts as a research assistant researcher at the Hong Kong Polytechnic University from August 2007 to June 2008. His current interests include pattern recognition, biometrics, and machine learning. He has published more than 40 scientific papers.

**Aini Zhong** obtained her B.S. degree at in 2008. Now she is working for her M.S. degree in Computer Science & Technology at Shenzhen graduate school, Harbin Institute of Technology. She is interested in pattern recognition and biometrics.

**Jian Yang** received the BS degree in mathematics from the Xuzhou Normal University in 1995. He received the MS degree in applied mathematics from the Changsha Railway University in 1998 and the Ph.D. degree from the Nanjing University of Science and Technology(NUST), on the subject of pattern recognition and intelligence systems in 2002. In 2003, he was a postdoctoral researcher at the University of Zaragoza, and in the same year, he was awarded the RyC program Research Fellowship sponsored by the Spanish Ministry of Science and Technology. From 2004 to 2006, he was a postdoctoral fellow at Biometrics Centre of Hong Kong Polytechnic University. From 2006 to 2007, he was a postdoctoral fellow at the Department of Computer Science of New Jersey Institute of Technology. Now, he is a professor in the School of Computer Science and Technology of NUST. He is the author of more than 50 scientific papers in pattern recognition and computer vision. His research interests include pattern recognition, computer vision and machine learning. Currently, he is an associate editor of Pattern Recognition Letters and Neurocomputing.

**David Zhang** graduated in Computer Science from Peking University. He received his M.Sc. in Computer Science in 1982 and his Ph.D. in 1985 from the Harbin Institute of Technology (HIT). From 1986 to 1988 he was a Postdoctoral Fellow at Tsinghua University and then an Associate Professor at the Academia Sinica, Beijing. In 1994 he received his second Ph.D. in Electrical and Computer Engineering from the University of Waterloo, Ontario, Canada. Currently, he is a Head, Department of Computing, and a Chair Professor at the Hong Kong Polytechnic University where he is the Founding Director of the Biometrics Technology Centre (UGC/CRC) supported by the Hong Kong SAR Government in 1998. He also serves as Visiting Chair Professor in Tsinghua University, and Adjunct Professor in Peking University, Shanghai Jiao Tong University, HIT, and the University of Waterloo. He is the Founder and Editor-in-Chief, International Journal of Image and Graphics (IJIG); Book Editor, Springer International Series on Biometrics (KISB); Organizer, the International Conference on Biometrics Authentication (ICBA); Associate Editor of more than ten international journals including IEEE Transactions and Pattern Recognition; and the author of more than 10 books and 200 journal papers. Professor Zhang is a Croucher Senior Research Fellow, Distinguished Speaker of the IEEE Computer Society, and a Fellow of both IEEE and IAPR.