Manifold Adaptive Kernel Local Fisher Discriminant Analysis for Face Recognition

Ziqiang Wang
Henan University of Technology, Zhengzhou, 450001, China
Email: wzqagent@126.com

Xia Sun
Henan University of Technology, Zhengzhou, 450001, China
Email: sunxiamail@126.com

***Abstract*—To efficiently cope with the high dimensionalities and complex nonlinear variations of face images in face recognition task, a novel manifold adaptive kernel local Fisher discriminant analysis algorithm is proposed in this paper. The core idea of this algorithm is as follows: First, the local manifold structure of the face image is modeled by a nearest neighbor graph. Then, an original input kernel function is deformed with respect to the local manifold structure. Finally, the resulting manifold adaptive kernel function is incorporated into the kernel local Fisher discriminant analysis(LFDA) method, which leads to the manifold adaptive kernel LFDA(MAKL) algorithm for face recognition. Experimental results on three popular face databases show that the proposed algorithm performs much better than other related algorithms.**

***Index Terms*—face recognition, local Fisher discriminant analysis, manifold learning, kernel function**

I. INTRODUCTION

During the past two decades, face recognition has received a lot of attention because of its wide application in many fields[1-4], such as identity authentication, information security, surveillance, human–computer interface, and so on. However, a major challenge of face recognition is that the captured face image data often lies in a high-dimensional space,ranging from several hundreds to thousands. Due to the consideration of the curse of dimensionality, a common way to cope with this problem is to use dimensionality reduction techniques. Dimensionality reduction could effectively avoid the “curse of dimensionality”, improve performance and computational efficiency of pattern classification, suppress noise, and alleviate storage requirement[5]. Therefore, the dimensionality reduction-based face recognition algorithm has attracted growing atttention in the computer vision and pattern recognition fields.

The most representative dimensionality reduction algorithms include principal component analysis (PCA) and linear discriminant analysis (LDA)[6]. PCA projects the data points into a lower dimensional subspace, in which the sample variance is maximized. It computes the eigenvectors of the sample covariance matrix and approximates the original data by a linear combination of the leading eigenvectors. PCA is optimal for reconstruction but is not optimal for discrimination. Unlike PCA is unsupervised, LDA is a supervised dimensionality reduction algorithm. LDA searches for the project axes on which the data points of different classes are far from each other while requiring data points of the same class to be close to each other. LDA encodes discriminatory information by finding directions that maximize the ratio of between-class scatter to within-class scatter. While these two algorithms have yielded impressive results on face recognition, they may fail to detect the underlying manifold structure as they are designed for discovering only the global Euclidean structure for face representation and recognition[7-9].

Recently, a number of research efforts have shown that the face images possibly lie on a nonlinear submanifold hidden in the image space, and face representation is fundamentally related to the problem of manifold learning[9-12]. Given a set of high-dimensional data points, manifold learning takes the local structure information into account, aiming to directly discover an intrinsically low-dimensional manifold space embedded in the ambient space. The most well-known manifold learning algorithms include Isomap[13], LLE[14], and Laplacian Eigenmap(LE)[15]. Isomap aims at finding a Euclidean embedding such that Euclidean distances in $\mathbb{R}^n$ can provide a good approximation to the geodesic distances on M . The basic idea of LLE is that the data points might reside on a nonlinear submanifold, but it might be reasonable to assume that each local neighborhood is linear. The Laplacian Eigenmap(LE) is based on spectral graph theory, it aims to preserve the similarities of the neighboring points. Despite the success of applying these manifold learning algorithms to many fields, they are defined only on the training data points but cannot be used for embedding new test data points. Therefore, they are not suitable for face recognition. To tackle the out-of-sample problem, locality preserving

Manuscript received March 7, 2012; revised April 2, 2012; accepted April 25, 2012.
Project number: 70701013.
Corresponding author: Ziqiang Wang.

doi:10.4304/jmm.7.6.387-393

projection (LPP)[16] obtains the face subspace by finding the optimal linear approximations to the eigenfunctions of the Laplace Betrami operator on the manifold. While LPP has attained reasonably good performance in face recognition, it is an unsupervised dimensionality reduction algorithm and does not take the valuable label information into account. Therefore, LPP might not be optimal in discriminating face images with different semantics which is the ultimate goal of face recognition. More recently, a new manifold learning algorithm called local Fisher discriminant analysis(LFDA)[17] is proposed for dimensionality reduction. By effectively combining the merits of LDA and LPP, LFDA maximizes between-class separability and preserves within-class local structure at the same time. Although LFDA has achieved reasonably good performance by integrating both local manifold structure and label information simultaneously, it is still a linear algorithm in nature. So LFDA may fail to discover the intrinsic manifold structure when the face image space is highly nonlinear.

In this paper, we proposed a novel manifold adaptive kernel dimensionality reduction algorithm for face recognition. By using a data-dependent norm on reproducing kernel Hilbert space (RKHS)[18], we can warp the structure of the RKHS to reflect the underlying geometry of the face image data. The traditional local Fisher discriminant analysis(LFDA) can then be performed in the manifold adaptive kernel space. We discuss how to kernelize the LFDA which gives rise to nonlinear manifold adaptive kernel LFDA algorithm for face recognition.

The remainder of the paper is organized as follows. Section II gives a brief review of the LFDA algorithm. Our manifold adaptive kernel LFDA algorithm for face recognition is introduced in Section III. The experimental results are presented in Section IV. Finally, we provide some concluding remarks and suggestions for future work in Section V.

II. Brief Review Of LFDA

LFDA is a recently proposed linear dimensionality reduction algorithm[17]. It is based on locality preserving projection and explicitly considers the class label information of the data space. LFDA aims to reduce the dimensionality of multimodal labeled data approximately by maximizing between-class separability and preserving the within-class local structure at the same time. LFDA encodes discriminatory information by finding the transformation matrix such that nearby data pairs in the same class are made close and the data pairs in different classes are separated from each other; far apart data pairs in the same class are not imposed to each other.

Given a set of face images $\mathrm{X}=\{x_1,x_2,\cdots,x_n\}\subset\mathbb{R}^p$ belonging to c classes, the number of face images in the ith class is n_i satisfying $\sum_{i=1}^{c}n_i=n$. Let A be an affinity matrix defined on the data points. The objective function of LFDA is as follows:

$$V_{opt}=\arg\max_V\frac{V^TS^{(b)}V}{V^TS^{(w)}V} \tag{1}$$

$$S^{(b)}=\frac{1}{2}\sum_{i,j=1}^{n}W_{ij}^{(b)}\left(x_i-x_j\right)\left(x_i-x_j\right)^T \tag{2}$$

$$S^{(w)}=\frac{1}{2}\sum_{i,j=1}^{n}W_{ij}^{(w)}\left(x_i-x_j\right)\left(x_i-x_j\right)^T \tag{3}$$

$$W_{ij}^{(b)}=\begin{cases}A_{ij}\left(\dfrac{1}{n}-\dfrac{1}{n_l}\right), & \text{if } c_i=c_j=l;\\ \dfrac{1}{n}, & \text{if } c_i\neq c_j.\end{cases} \tag{4}$$

$$W_{ij}^{(w)}=\begin{cases}\dfrac{A_{ij}}{n_l}, & \text{if } c_i=c_j=l;\\ 0, & \text{if } c_i\neq c_j.\end{cases} \tag{5}$$

where $S^{(b)}$ and $S^{(w)}$ denote the local between-class scatter matrix and local within-class scatter matrix, respectively, $W_{ij}^{(b)}$ and $W_{ij}^{(w)}$ denote the weight matrices of the local between-class adjacency graph and local within-class adjacency graph, respectively, c_i is the class label of the data point x_i, and $l\in\{1,2,\cdots,c\}$ is the class label. A_{ij} is the heat kernel weight.

$$A_{ij}=\begin{cases}\mathrm{e}^{-\left(\|x_i-x_j\|^2/t\right)}, & \text{if } x_i \text{ is among the } r \text{ nearest neighbor of } x_j \\ & \text{or } x_j \text{ is among the } r \text{ nerest neighbor of } x_i.\\ 0, & \text{otherwise.}\end{cases} \tag{6}$$

The justification for such choice and the setting of parameter can be referred to [15].

As can be seen from (1), by preserving the local geometric structure, LFDA aims to look for a transformation matrix V such that the data pairs in the same class are made close and the data pairs in different classes are separated from each other. Finally, the optimal V 's are the eigenvectors corresponding to the maximum eigenvalue of the generalized eigenvalue problem:

$$S^{(b)}V=\lambda S^{(w)}V \tag{7}$$

Therefore, the problem of LFDA is converted into the leading eigenvectors of $\left(S^{(w)}\right)^{-1}S^{(b)}$. For face recognition, a problem arises that the matrix $S^{(w)}$ cannot be guaranteed to be nonsingular since the number of available samples is smaller than the dimensionality of the samples. In this case, one can first apply PCA to remove the components corresponding to zero eigenvalues.

While LFDA has shown its promising results on many classification tasks, linear method fail to deliver good performance when face patterns are subject to large variations in illumination, facial expression and pose variations, which results in a highly complex nonlinear distribution. The limited success of LFDA algorithm should be attributed to its linear nature. As a result, it is reasonable to assume that a better solution to this inherent nonlinear problem could be achieved using kernel-based nonlinear methods.

III. Manifold Adaptive Kernel LFDA

In order to solve nonlinear problems, the conventional LFDA can be generalized to its nonlinear version with kernel trick[19], namely kernel LFDA. The kernel trick first maps the input data into an implicit feature space F with a nonlinear mapping, and then the data are analyzed in F. The kernel trick has been demonstrated to be able to effectively represent complicated nonlinear relations of the input data, and recently kernel-based nonlinear dimensionality reduction algorithms have been received more attention. Therefore, such a nonlinear generalization is meaningful in the sense that a kernelized LFDA would generally achieve better accuracy, and relax the restriction of LFDA being only a linear dimensionality reduction scheme. In this section, we will describe our manifold adaptive kernel LFDA design approach which is fundamentally based on LFDA and manifold adaptive kernel. We begin with a description of kernel LFDA.

A. Kernel LFDA

Let $\varphi : x \in \mathbb{R}^p \rightarrow \varphi(x) \in F$ be a nonlinear mapping from the input space $\mathbb{R}^p$ to a feature space F, the idea behind kernel LFDA is to perform a traditional LFDA in the feature space F instead of the input space $\mathbb{R}^p$. Then the face image data matrix in F can be denoted as $\varphi(x) = [\varphi(x_1), \varphi(x_2), \cdots, \varphi(x_n)]$. In implementation, the feature vector does not need to be computed explicitly, while it is just done by computing the inner product of two vectors in F with a kernel function $k(\cdot,\cdot)$:

$$k(x_1, x_2) = \langle \varphi(x_1), \varphi(x_2) \rangle \tag{8}$$

Let $\overline{S}^{(b)}$ and $\overline{S}^{(w)}$ be the local between-class and within-class scatter matrices in the feature space F, respectively. We have

$$\overline{S}^{(b)} = \frac{1}{2}\sum_{i,j=1}^{n} W_{ij}^{(b)} \left(\varphi(x_i) - \varphi(x_j)\right)\left(\varphi(x_i) - \varphi(x_j)\right)^T \tag{9}$$

$$\overline{S}^{(w)} = \frac{1}{2}\sum_{i,j=1}^{n} W_{ij}^{(w)} \left(\varphi(x_i) - \varphi(x_j)\right)\left(\varphi(x_i) - \varphi(x_j)\right)^T \tag{10}$$

Let U be the projective function in the feature space F, performing LFDA in F means maximizing the local between-class scatter $\overline{S}^{(b)}$ and minimizing the local within-class scatter $\overline{S}^{(w)}$. So the corresponding objective function (1) in the feature space is

$$U_{opt} = \arg\max_{U} \frac{U^T \overline{S}^{(b)} U}{U^T \overline{S}^{(w)} U} \tag{11}$$

which can be solved by the generalized eigenvalue problem:

$$\overline{S}^{(b)} U = \lambda \overline{S}^{(w)} U \tag{12}$$

Because the eigenvectors are linear combinations of $\varphi(x_i)$, there exist coefficient α_i such that

$$U = \sum_{i=1}^{n} \alpha_i \varphi(x_i) \tag{13}$$

Let $\alpha = [\alpha_1, \alpha_2, \cdots, \alpha_n]^T$, it can be proved that equation (11) is equivalent to

$$\alpha_{opt} = \arg\max_{\alpha} \frac{\alpha^T K L^{(b)} K \alpha}{\alpha^T K L^{(w)} K \alpha} \tag{14}$$

and the corresponding generalized eigenvalue problem is:

$$K L^{(b)} K \alpha = \lambda K L^{(w)} K \alpha \tag{15}$$

where K is the kernel matrix $K_{ij} = k(x_i, x_j)$, $L^{(b)}$ and $L^{(w)}$ are defined as:

$$L^{(b)} = D^{(b)} - W^{(b)} \tag{16}$$

$$L^{(w)} = D^{(w)} - W^{(w)} \tag{17}$$

where $D^{(b)}$ and $D^{(w)}$ are both diagonal matrices, and their entries are column sums of $W^{(b)}$ and $W^{(w)}$, respectively, $D_{ii}^{(b)} = \sum_{j=1}^{n} W_{ij}^{(b)}$ and $D_{ii}^{(w)} = \sum_{j=1}^{n} W_{ij}^{(w)}$.

So the problem of kernel LFDA is converted into finding the leading eigenvectors of $\left(K L^{(w)} K\right)^{-1} K L^{(b)} K$, each eigenvector α gives a projective function U in the feature space. For a new data point x, its projection onto U in the feature space F can be calculated by

$$\begin{aligned}\langle U, \varphi(x) \rangle &= \sum_{i=1}^{n} \alpha_i \langle \varphi(x_i), \varphi(x) \rangle \\ &= \sum_{i=1}^{n} \alpha_i k(x_i, x)\end{aligned} \tag{18}$$

For face recognition, a problem arises that the matrix $K L^{(w)} K$ and $K L^{(b)} K$ cannot be guaranteed to be nonsingular. In this case, we may first apply PCA to remove the components corresponding to zero eigenvalues.

B. Data-Dependent Manifold Adaptive Kernel

Since different kernel functions will produce different constructions of implicit feature space, identifying the appropriate kernel function for a given dataset is essential to all kernel-based learning techniques. For face recognition, it has shown that the face images possibly reside on a nonlinear submanifold. However, the nonlinear structure captured by the data-independent kernels such as the Gaussian kernel, polynomial kernel and Sigmoid kernel, may not be consistent with the intrinsic manifold structure which has been shown very useful for improving the learning performance by many previous researches[20]. To further improve the learning performance of kernel LFDA, in the following we discuss how to incorporate the manifold structure into the construction process of kernel function which leads to a data-dependent manifold adaptive kernel function.

Let L be a linear space with a positive semi-definite inner product(quadratic form) and let $S: H \to L$ be a bounded linear operator. We define $\tilde{H}$ to be the space of functions from H with the modified product

$$\langle f, g \rangle_{\tilde{H}} = \langle f, g \rangle_H + \langle Sf, Sg \rangle_L \quad (19)$$

It has been proved that $\tilde{H}$ is still a RKHS[18].

Given the data points $x_1, \cdots, x_n$, let $S: H \to \mathbb{R}^n$ be the evaluation map

$$S(f) = \left(f(x_1), \cdots, f(x_n)\right)^T \quad (20)$$

Denote $f = \left(f(x_1), \cdots, f(x_n)\right)^T$, note that $f \in L$, thus we have

$$\langle Sf, Sf \rangle_L = \langle f, f \rangle = f^T M f \quad (21)$$

where M is a positive semi-definite matrix.

Given an input kernel function k, and denote $k_x = \left(k(x, x_1), \cdots, k(x, x_n)\right)$. It can be shown that the reproducing kernel $\tilde{k}$ in $\tilde{H}$ is

$$\tilde{k}(x, z) = k(x, z) - k_x^T (I + MK)^{-1} M k_z \quad (22)$$

where I is an identity matrix, and K is the kernel matrix $K_{ij} = k(x_i, x_j)$ in H. The key issue now is the choice of M, so that the deformation of the kernel induced by the data-dependent norm, is motivated with respect to the underlying geometry of the data.

As suggested in [15], the manifold structure can be modeled by a nearest-neighbor graph which preserves the local structure of the face image space. Let G denote a nearest-neighbor graph with n nodes, the ith node corresponds to the face image x_i. The graph Laplacian matrix L is defined as $L = D - W$, where

$$W_{ij} = \begin{cases} 1, & \text{if } x_i \text{ and } x_j \text{ are adjacent} \\ 0, & \text{otherwise} \end{cases} \quad (23)$$

and D is a diagonal matrix given by $D_{ii} = \sum_j W_{ij}$. The graph Laplacian provides the following smoothness penalty on the graph:

$$f^T L f = \frac{1}{2} \sum_{i,j=1}^{n} \left(f(x_i) - f(x_j)\right)^2 W_{ij} \quad (24)$$

Finally, by setting $M = L$, we can obtain the following manifold adaptive kernel:

$$\tilde{k}(x, z) = k(x, z) - k_x^T (I + LK)^{-1} L k_z \quad (25)$$

In short, the key idea of manifold adaptive kernel is to construct a data-dependent norm on Reproducing Kernel Hilbert Spaces (RKHS), which can warp the structure of the RKHS to reflect the underlying geometry of the data. When an input kernel is deformed according to the manifold structure, the resulting kernel may be able to achieve better performance than the original input kernel.

C. The Manifold Adaptive Kernel LFDA Algorithm

After calculating the manifold adaptive kernel matrix $\tilde{K}$: $\tilde{K}_{ij} = \tilde{k}(x_i, x_j)$ in terms of (25), we can employ the manifold adaptive kernel to find the optimal projection of kernel LFDA in the new RKHS $\tilde{H}$. Let us replace kernel matrix K in (15) with the deformed kernel matrix $\tilde{K}$, we obtain

$$\tilde{K} L^{(b)} \tilde{K} \alpha = \lambda \tilde{K} L^{(w)} \tilde{K} \alpha \quad (26)$$

Therefore, for a new testing face image x, the feature vector extracted from the new RKHS is derived as

$$x \to y = \sum_{i=1}^{n} \alpha_i \tilde{k}(x_i, x) \quad (27)$$

where y is the lower-dimensional representation of the face image x.

We summarize our manifold adaptive kernel LFDA algorithm (MAKL) as follows:

1) PCA processing: We project the face images x_i into the PCA subspace by throwing away the components corresponding to zero eigenvalue.
2) Calculate the original input kernel matrix K: $K_{ij} = k(x_i, x_j)$.
3) Construct a r nearest-neighbor G with weight matrix W defined in (23), and compute the graph Laplacian $L = D - W$.
4) Calculate the manifold adaptive kernel matrix $\tilde{K}$: $\tilde{K}_{ij} = \tilde{k}(x_i, x_j)$ according to (25).
5) Find α by solving the generalized eigenvalue problem in (26), and obtain the low-dimensional feature vector of high-dimensional face image data via (27).

Once we get lower-dimensional feature representations of the original face images with the manifold adaptive kernel LFDA algorithm, face recognition becomes a pattern recognition task. The traditional classifier algorithms can be applied to identify different face images in the reduced semantic space. In this paper, we apply the nearest-neighbor classifier for its simplicity, and the Euclidean metric is used as our distance measure.

IV. Experimental Results

In this section, to investigate the performance of our proposed manifold adaptive kernel LFDA algorithm (MAKL) for face recognition, we compare it with the kernel PCA algorithm (KPCA)[21], the kernel LDA algorithm(KLDA)[22], the kernel LPP algorithm (KLPP) [16], and the local Fisher discriminant analysis algorithm (LFDA)[17], four of the most representative dimensionality reduction algorithms for face recognition.

Note that the three algorithms KLPP, LFDA and MAKL need to construct an adjacent graph on the face images. In the following experiments, we use the same adjacent graph for these three algorithms and the nearest neighbor number is set to be 7. For the kernel function selection of the KPCA, KLDA and KLPP algorithms, we adopted the popular Gaussian kernel function.

$$k(x,y)=\exp\left(-\frac{\|x-y\|^2}{2\sigma^2}\right) \tag{28}$$

where the value for σ was set to be 0.5 multiplied by the average of pairwise distances in the training data. For fair comparison, we also set the original input kernel function of MAKL algorithm to the above Gaussian kernel function. In addition, KLDA, KLPP, LFDA and MAKL algorithms involve a preceding PCA stage to void the singularity problem, we keep 98% data energy in the PCA stage.

Our empirical study on face recognition was conducted based on three benchmark databases: the Olivetti Research Laboratory (ORL), Yale, and CMU PIE face database. In all the experiments, preprocessing to locate the faces was applied. All the original face images are manually aligned by fixing the locations of two eyes, cropped and then resized to 32×32 pixels, with 256 gray levels per pixel. Each face image is represented by a 1024-dimensional vector in the image space. Some sample face images after preprocessing of the three databases are shown in Fig.1-Fig.3. To perform face recognition, we first obtain the face subspace with dimensionality reduction algorithms. Then, the new face image to be identified is projected into the face subspaces. Finally, the nearest neighbor classifier is adopted to identify the new face image, where the Euclidean metric is used as the distance measure.

The ORL face database (http://www.uk.research. att.com /facedatabase.html) contains 400 images of 40 individuals, some images were captured at different times and have different variations including expression (open or closed eyes, smiling or nonsmiling) and facial details (glasses or no glasses). A random subset with five images per individual was taken to constitute the training set, and the rest images of the database make up the testing set. The Yale database (http://cvc.yale.edu/projects/yalefaces/ yalefaces.html) contains 165 front view face images of 15 individuals, the images demonstrate variations in lighting condition, facial expression (normal, happy, sad, sleepy, surprised, and wink). A random subset with six images per individual was taken to form the training set, and the rest images of the database was considered as testing set. The CMU PIE face database contains 68 subjects 41368 face images as a whole[23], the face images were captured by 13 synchronized cameras and 21 flashes, under varying pose, illumination, and expression. We use a subset of 5 near front poses (C05, C07, C09, C27, and C29) and illuminations indexed as 08 and 11 is used, thus each person has ten images, within the ten face images for each individual in our experiment, a random subset with five images per individual was used to form the training set, and the rest of the database was considered to be the testing set. In order to obtain steady results, these trials were independently conducted ten times and the average accuracy was reported.

The recognition results are shown in Table I- Table III. The recognition accuracy rates versus the reduced dimensions on the tree face databases are shown in Fig.4-Fig.6. As can be seen, the main observations from the performance comparisons include:

1) Our proposed MAKL algorithm consistently outperforms the KPCA, KLDA, KLPP and LFDA algorithms, which demonstrate that MAKL can effectively utilize the data-dependent manifold adaptive kernel function for face recognition.
2) The KPCA algorithm performs poorly. This is probably because KPCA is unsupervised learning method and does not encode valuable discrimination information.
3) The KLDA algorithm performs comparatively to the KLPP algorithm. This demonstrates that it is hard to evaluate whether local manifold structure or class label information is more important.
4) Although the LFDA algorithm outperforms KPCA, KLDA and KLPP algorithms by using both local manifold structure and class label information, it is still a linear algorithm and is inadequate to describe the nonlinear face image space due to the high variability of the image content and style. Therefore, it performs worse than the kernel-based MAKL algorithm.

V. CONCLUSIONS

We have introduced a novel dimensionality reduction algorithm for face recognition called Manifold Adaptive Kernel LFDA(MAKL). Unlike most of traditional dimensionality reduction algorithms which seek the data-independent nonlinear structure of the face image space, our proposed algorithm explicitly considers both the intrinsic manifold structure and discriminative information. Experiments on three face databases demonstrate the effectiveness of the proposed algorithm. Since the proposed MAKL algorithm is a general nonlinear dimensionality reduction algorithm for high-dimensional data, we plan to apply the algorithm to video and audio classification in the future.

APPENDIX A FIGURE(1-6) AND TABLE (I-III)

Figure 1. Face image examples of the ORL database.

Figure 2. Face image examples of the Yale database.

Figure 3. Face image examples of the PIE database.

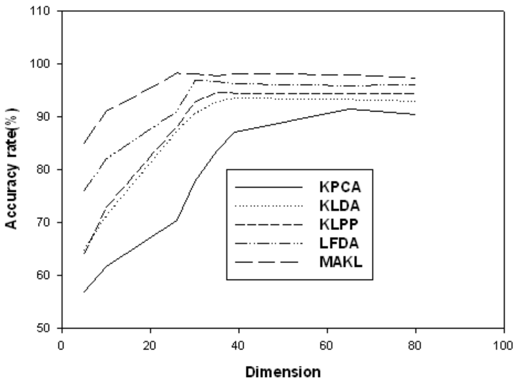

Figure 4. Accuracy rate versus reduced dimensionality on the ORL database.

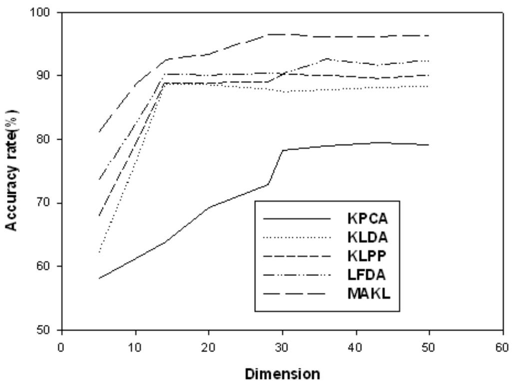

Figure 5. Accuracy rate versus reduced dimensionality on the Yale database.

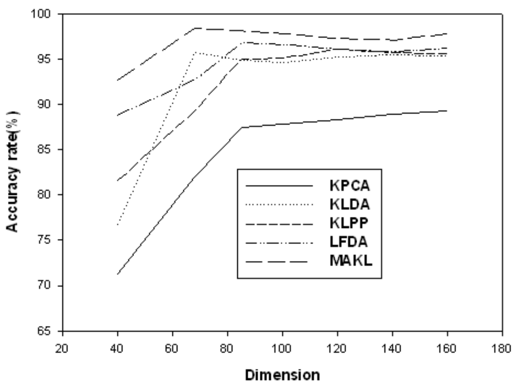

Figure 6. Accuracy rate versus reduced dimensionality on the PIE database.

TABLE I.
PERFORMANCE COMPARISONS ON THE ORL DATABASE

Algorithms	Dimension	Recognition rate
KPCA	65	91.5%
KLDA	39	93.3%
KLPP	35	94.6%
LFDA	30	96.8%
MAKL	26	98.2%

TABLE II.
PERFORMANCE COMPARISONS ON THE YALE DATABASE

Algorithms	Dimension	Recognition rate
KPCA	43	79.4%
KLDA	14	88.7%
KLPP	30	90.2%
LFDA	36	92.6%
MAKL	28	96.5%

TABLE III.
PERFORMANCE COMPARISONS ON THE PIE DATABASE

Algorithms	Dimension	Recognition rate
KPCA	160	89.3%
KLDA	67	95.7%
KLPP	120	96.1%
LFDA	85	96.8%
MAKL	68	98.4%

ACKNOWLEDGMENT

This work is supported by the National Natural Science Foundation of China under Grant No.70701013, the Natural Science Foundation of Henan Province under Grant No. 102300410020,0611030100, and the Natural Science Foundation of Henan University of Technology under Grant No. 08XJC013 and 09XJC016.

REFERENCES

[1] S.Zafeiriou, G.Tzimiropoulos, M.Petrou, and T.Stathaki, “Regularized kernel discriminant analysis with a robust kernel for face recognition and verification,” *IEEE Transactions on Neural Networks and Learning Systems*, vol.23, pp.526-534,March 2012.

[2] A.K.Jain, B.Klare, and U.Park, “Face matching and retrieval in forensics applications,” *IEEE Multimedia*, vol.19, pp.20-20,January 2012.

[3] X.-H. Han, Y.-W. Chen, and X.Ruan, “Multilinear supervised neighborhood embedding of a local descriptor tensor for scene/object recognition,” *IEEE Transactions on Image Processing*,vol.21,pp.1314-1326, March 2012.

[4] A.Li, S.Shan,and W.Gao, "Coupled bias-variance tradeoff for cross-pose face recognition," *IEEE Transactions on Image Processing*, vol.21,pp.305-315,January 2012.
[5] H.Wang,S.Chen,Z.Hu,and W.Zheng, "Locality-preserved maximum information projection," *IEEE Transactions on Neural Networks*,vol.19,pp.571-585, April 2008.
[6] R.O.Duda, P.E.Hart, and D. G. Stork, *Pattern Classification*, 2nd edition. Hoboken, NJ: Wiley-Interscience, 2000.
[7] D.Cai, X.He, J.Han, and H.-J.Zhang, "Orthogonal Laplacianfaces for face recognition," *IEEE Transactions on Image Processing*, vol.15, pp.3608-3614, November 2006.
[8] S.Yan, D.Xu, B.Zhang, H.-J.Zhang, Q.Yang, and S.Lin, "Graph embedding and extensions: a general framework for dimensionality reduction," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol.29,pp.40-51, January 2007.
[9] 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,pp.328-340, March 2005.
[10] T.-K. Kim, J.Kittler,and R.Cipolla, "On-line learning of mutually orthogonal subspaces for face recognition by image sets," *IEEE Transactions on Image Processing*, vol.19, pp.1067-1074,April 2010.
[11] C.Chen,J.Zhang,and R.Fleischer, "Distance approximating dimension reduction of Riemannian manifolds," *IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics*,vol.40,pp.208-217, February 2010.
[12] T.Zhang,B.Fang,Y.Y.Tang,G.He,and J.Wen, "Topology preserving non-negative matrix factorization for face recognition," *IEEE Transactions on Image Processing*, vol.17,pp.574-584,April 2008.
[13] J. Tenenbaum, V. Silva, and J. Langford, "A global geometric framework for nonlinear dimensionality reduction," *Science*, vol. 290,pp.2319-2323, December 2000.
[14] S. Roweis and L. Saul, "Nonlinear dimensionality reduction by locally linear embedding," *Science*, vol. 290, pp. 2323-2326, December 2000.
[15] M.Belkin and P.Niyogi, "Laplacian eigenmaps for dimensionality reduction and data representation," *Neural Computation*,vol.15,pp.1373-1396,June 2003.
[16] X. He,P. Niyogi, "Locality preserving projections," in *Advances in Neural Information Processing Systems*, Cambridge, MA:MIT Press,2003,pp.385-392.
[17] M.Sugiyama,"Dimensionality reduction of multimodal labeled data by local Fisher discriminant analysis," *Journal of Machine Learning Research*, vol.8,pp.1027-1061,July 2007.
[18] V. Sindhwani, P. Niyogi, and M. Belkin, "Beyond the point cloud: from transductive to semi-supervised learning," in *Proceedings of 2005 International Conference of Machine learning*, New York:ACM Press, 2005, pp.824-831.
[19] V. N. Vapnik, *The Nature of Statistical Learning Theory*. New York: Springer-Verlag, 1995.
[20] D.Cai and X.He,"Manifold adaptive experimental design for text categorization," *IEEE Transactions on Knowledge and Data Engineering*,vol.24,pp.707-719,April 2012.
[21] K.I.Kin, K.Jung, and H.J.Kim, "Face recognition using kernel principal component analysis," *IEEE Signal Processing Letters*, vol.9,pp.40-42, February 2002.
[22] Q.Liu, H.Lu, and S.Ma, "Improving kernel Fisher discriminant analysis for face recognition," *IEEE Transactions on Circuits and Systems for Video Technology*,vol.14,pp.42-49,January 2004.
[23] T. Sim, S. Baker, and M. Bsat, "The CMU pose, illuminlation, and expression database," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol.25, pp.1615-1618, December 2003.

Ziqiang Wang was born Zhoukou, Henan, China in 1973. He received the PhD degree from China University of Mining &Technology (Beijing) and the Master of Science degree from Xi'an Petroleum University, both in computer application, in 2011 and 2002, respectively. His research interests include data mining and pattern recognition.

He is currently an associate professor in the Henan University of Technology.

Xia Sun was born Xi'an, Shanxi, China in 1978. She received the Master of Science degree in computer application from Huazhong University of Science and Technology in 2008. Her research interests include data mining and pattern recognition.

She is currently a lecturer in the Henan University of Technology.