Orthogonal Maximum Margin Projection 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*—Dimensionality reduction techniques that can introduce low-dimensional feature representation with enhanced discriminatory power are of paramount importance in face recognition. In this paper, a novel subspace learning algorithm called orthogonal maximum margin projection(OMMP) is proposed. The OMMP algorithm is based on the maximum margin projection (MMP), which aims at discovering both geometrical and discriminant structures of the face manifold. First, OMMP considers both the local manifold structure and class label information by using the within-class and between-class graphs, as well as characterizing the separability of different classes with the margin criterion, then OMMP orthogonalizes the basis vectors of the face subspace. Experimental results on three databases show the effectiveness of the proposed OMMP algorithm.**

***Index Terms*—dimensionality reduction, face recognition, maximum margin projection(MMP), orthogonal maximum margin projection (OMMP)**

I. INTRODUCTION

Face recognition has attracted growing attention within the last two decades because of its potential applications in many fields, such us identity authentication, video surveillance, image search, and human-computation communication. As a result, numerous face recognition (FR) techniques have been proposed. One of the most successfully and well-studied techniques to FR is the appearance-based method[1], which operates directly on images or appearances of face objects and process the images as two-dimensional holistic patterns. However, a major challenge of the appearance-based FR methods is that the image data often lie in a high-dimensional feature space. Considering the curse-of-dimensionality[2-3], it is desirable to first project the images into the lower-dimensional subspace with the dimensionality reduction methods, then the conventional classification algorithms can be applied.

Manuscript received March 7, 2011; revised April 2, 2011; accepted April 10, 2011.
Project number: 70701013.
Corresponding author: Ziqiang Wang.

Principle component analysis(PCA)[4] and linear discrimiant analysis (LDA)[5] are two classical dimensionality reduction algorithm. PCA, also known as Karhunen–Loéve transformation, aims to find a set of mutually orthogonal bases that capture the global information of the data points in terms of variance. In contrast to unsupervised method such as PCA, LDA is a supervised dimensionality reduction approach. LDA aims to find an optimal transformation that maps the data into a lower-dimensional space that minimizes the within-class distance and simultaneously maximizes the between-class distance, thus achieving maximum discrimination. It is generally believed that the LDA-based algorithms are superior to PCA-based ones, since the former focus on the most discriminant feature extraction while the latter achieves simply data reconstruction. Although PCA and LDA have been extensively applied to face recognition, they are designed for discovering the global Euclidean structure. They fail to uncover the manifold structure hidden in the image space. In many cases, local manifold structure is more important than global Euclidean structure[6-12], especially when nearest-neighbor like classifiers are used for classification.

Recently, Locality preserving projection(LPP)[6] has been proposed for dimensionality reduction that preserves the local geometrical relationship and discovers its intrinsic manifold structure. The LPP is the linear approximations to the eigenfunctions of the Laplace-Beltrami operator on the compact Riemannian manifold. However, the basis functions obtained by the LPP are nonorthogonal. This makes it difficult to reconstruct the data. To overcome this shortcoming, Cai et al. propose an algorithm called orthogonal locality preserving projection (OLPP)[13], which can produce orthogonal basis functions and can have more locality preserving power than LPP. However, LPP and OLPP are unsupervised dimensionality reduction methods and do not take the label information into account. If the samples needed to be classified take on multimanifolds (corresponding to multi-classes), then the locality-preserving manifold learning algorithms may result in overlapped embeddings belonging to different classes, which deteriorates the

doi:10.4304/jcp.7.2.377-383

discrimination performance. In order to reduce the dimensionality of multimanifolds data appropriately, it is necessarily to preserve both the local geometrical and discriminant structures of the data manifold simultaneously, since the local structure of the data itself is not sufficient.

In this paper, we propose a new algorithm, called orthogonal maximum margin projection(OMMP) which is fundamentally based on the maximum margin projection(MMP)[14]. It focuses on local manifold discriminant analysis for face recognition. Its goal is to discover both geometrical and discriminant structures of the data manifold. OMMP builds two adjacency graphs which can best reflect the local geometrical structure of the underlying manifold and the class relationship between the sample points. Inspired by the ideas of MMP, we proposed an objective function that maximizes the margin between embeddings belong to different classes, and at the same time requires the basis functions to be orthogonal. By jointly considering the local manifold structure and the label information with the within-class graph and between-class graphs, as well as preserving the metric structure of face space with orthogonal basis functions, OMMP is expected to deliver much better performance in face recognition applications.

The rest of the paper is organized as follows. In Section II, we provide a brief review of the MMP algorithm. The OMMP algorithm is developed in Section III. The experimental results are reported in Section IV. Finally, the conclusions along with some directions for further research are presented in Section V.

II. BRIEF REVIEW OF MMP

MPP is a recently proposed linear dimensionality reduction algorithm[14]. It builds two adjacency graphs which can best reflect the local geometrical structure of the underlying manifold and the class relationship between the sample points. It is a graph-based approach for learning a linear approximation to the intrinsic data manifold by making use of both labeled and unlabeled data. Its goal is to discover both geometrical and discriminant structures of the face manifold.

Given a set of face images $x_1, x_2, \ldots, x_n \in R^p$, Let $X = [x_1, x_2, \ldots, x_n]$. MMP aims to find a linear transformation $a \in R^{p\times q}$ that maps each vector $x_i (i = 1, \ldots, n)$ in the *p*-dimensional space to a vector y_i in the lower *q*-dimensional space by $y_i = a^T x_i$ such that y_i represent x_i according to some optimal criterion. MMP seeks a subspace that preserves the local geometrical and discriminant structures of the data manifold.

Let S_b and S_w be weight matrices of between-class graph G_b and within-class graph G_w, respectively. Their definitions are as follows:

$$S_{w,ij} = \begin{cases} \gamma, & \text{if } x_i \text{ and } x_j \text{ share the same label,} \\ 1, & \text{if } x_i \text{ or } x_j \text{ is unlabeled} \\ & \text{but } x_i \in N_w(x_j) \text{or } x_j \in N_w(x_i), \\ 0, & \text{otherwise.} \end{cases} \quad (1)$$

$$S_{b,ij} = \begin{cases} 1, & \text{if } x_i \in N_b(x_j) \text{or } x_j \in N_b(x_i), \\ 0, & \text{otherwise.} \end{cases} \quad (2)$$

Let $N(x_i) = \{x_i^1, \ldots, x_i^k\}$ denote the set of its *k* nearest neighbors, and $l(x_i)$ represents the label of x_i. Specially, $N_b(x_i) = \{x_i^j \,|\, l(x_i^j) \neq l(x_i), j = 1, \ldots, k\}$ contains the neighbors having different labels, and $N_w(x_i) = N(x_i) - N_b(x_i)$ contains the rest of the neighbors.

MMP aims to find the local geometrical and discriminant structures of the face manifold. Given two weight matrices S_b and S_w, MMP can be obtained by solving the following maximization problem:

$$a_{opt} = \arg\max_a a^T X (\beta L_b + (1-\beta) S_w) X^T a \quad (3)$$

with the constraint

$$a^T X D_w X^T a = 1 \quad (4)$$

Where $L_b = D_b - S_b$ is the Laplacian matrix of G_b, D_b is a diagonal matrix whose entries on diagonal are column sum of S_b, i.e., $D_{b,ii} = \sum_{j=1}^{n} S_{b,ij}$. Likewise, $L_w = D_w - S_w$ is the Laplacian matrix of G_w, D_w is a diagonal matrix whose entries on diagonal are column sum of S_w, i.e., $D_{w,ii} = \sum_{j=1}^{n} S_{w,ij}$.

The objective function in MMP incurs a heavy penalty if neighboring points x_i and x_j in the within-class graph are mapped far apart. Meanwhile, it incurs a heavy penalty if neighboring points x_i and x_j between-class graph are mapped close together. Therefore, maximizing it is an attempt to ensure that if x_i and x_j are "close" and share the same label, then $y_i (= w^T x_i)$ and $y_j (= w^T x_j)$ are close as well. Furthermore, maximizing it is an attempt to ensure that if x_i and x_j are "close" but have different labels, then $y_i (= w^T x_i)$ and $y_j (= w^T x_j)$ are far apart. Finally, the projection vector a that maximizes (3) is given by the maximum eigenvalue solution to the generalized eigenvalue problem:

$$X(\beta L_b + (1-\beta)S_w)X^T a = \lambda X D_w X^T a \quad (5)$$

where $\beta \in [0,1]$ is a suitable constant which controls the weight between a within-class graph and a between-class graph.

Note that if the number of samples is less than the number of features, it implies that both $X(\beta L_b + (1-\beta)S_w)X^T$ and XD_wX^T are singular. In such a case, one can first project the face vector into the PCA subspace by throwing away those zero singular values.

Once the eigenvector are obtained by solving (5). Let A be the transformation matrix. That is, $y_i = A^T x_i$ and $y_j = A^T x_j$. Therefore, the Euclidean distance between y_i and y_j in the lower dimensional subspace can be computed in the following:

$$\begin{aligned} d(y_i, y_j) &= \|y_i - y_j\| \\ &= \sqrt{(y_i - y_j)^T (y_i - y_j)} \\ &= \sqrt{(A^T x_i - A^T x_j)^T (A^T x_i - A^T x_j)} \\ &= \sqrt{(x_i - x_j)^T AA^T (x_i - x_j)} \end{aligned}$$

If A is orthogonal matrix, i.e., $AA^T = I$, it implies that the orthogonal basis functions of MMP can preserve the metric structure.

III. ORTHOGONAL MAXIMUM MARGIN PROJECTION ALGORITHM

In this section, we describe a novel manifold learning algorithm for face recognition, called orthogonal maximum margin projection(OMMP),which can be thought of as a combination of MMP and orthogonal transformation.

Note that the small sample size problem occurs frequently in appearance-based face recognition, since the number of face images (n) is much lower than the dimension of the face image vector (p). The fact that XD_wX^T is $p \times p$ matrices implies that it is singular. To overcome this problem, we first project the image set into a PCA subspace to make XD_wX^T nonsingular. Another consideration of using PCA as preprocessing is for noise reduction. The algorithmic procedure of OMMP is formally stated below:

1) PCA Projection. We project the image set $\{x_i\}$ into the PCA subspace by throwing away the components corresponding to zero eigenvalue. For the sake of simplicity, we still use x to represent the images in the PCA subspace in the following steps. Let W_{PCA} denote the transformation matrix of PCA.

2) Constructing the Within-Class and Between-Class Graphs. Let G denote a graph with n nodes. The ith node corresponds to the face image x_i. We construct two graphs, that is, within-class graph G_w and between-class graph G_b, to represent the intraclass compactness and interclass separability. In the within-class graph, for each node x_i, set the adjacency matrix $S_{w,ij} = S_{w,ji} = 1$ if x_i is among the k-nearest neighbor of x_j in the same class. Otherwise, set $S_{w,ij} = S_{w,ji} = 0$. In the between-class graph, set the similarity matrix $S_{b,ij} = 1$ if the pair (x_i, x_j) is among the shortest pair among the set $\{(x_i, x_j), l(x_i) \neq l(x_j)\}$, where $l(x)$ represents the class label of x. Otherwise, set $S_{b,ij} = 0$. The within-class and between-class graph model the local geometrical and discriminant structures of the face manifold.

3)Computing Orthogonal Eigenvector. Compute the orthogonal orthogonal basis vectors for the generalized eigenvector problem:

$$X(\beta L_b + (1-\beta)S_w)X^T a = \lambda X D_w X^T a \quad (6)$$

where $L_b = D_b - S_b$ is the Laplacian matrix of between-class graph G_b, D_b is a diagonal matrix whose entries on diagonal are column sum of S_b, i.e., $D_{b,ii} = \sum_{j=1}^{n} S_{b,ij}$, D_w is a diagonal matrix whose entries on diagonal are column sum of S_w, i.e., $D_{w,ii} = \sum_{j=1}^{n} S_{w,ij}$.

The generalized eigenvectors of (5) are non-orthogonal. The OMMP algorithm aims at finding a set of orthogonal basis vectors $a_1, a_2, \ldots, a_m$ which maximize the following objective function:

$$a_{opt} = \arg\max_a \frac{a^T X(\beta L_b + (1-\beta)S_w)X^T a}{a^T X D_w X^T a} \quad (7)$$

subject to the constraint

$$a_m^T a_1 = a_m^T a_2 = \cdots = a_m^T a_{m-1} = 0$$

Since the XD_wX^T is positive definite after PCA projection, for any a, we can always normalize it such that $a^T XD_wX^T a = 1$.Then the above maximization problem is equivalent to maximizing the following objective function:

$$a_{opt} = \arg\max_a a^T X(\beta L_b + (1-\beta)S_w)X^T a \quad (8)$$

subject to

$$a_m^T a_1 = a_m^T a_2 = \cdots = a_m^T a_{m-1} = 0, a^T X D_w X^T a = 1.$$

Where a_1 is the eigenvector of the matrix $\left(XD_wX^T\right)^{-1}X(\beta L_b+(1-\beta)S_w)X^T$ associated with the largest eigenvalue.

In order to compute the *m*th basis vector that maximizes the above objective function, the Lagrange multiplier is used to transform (8) to include all the constraints:

$$\begin{aligned}O_m &= a_m^T X\left(\beta L_b+(1-\beta)S_w\right)X^T a_m\\ &\quad -\lambda\left(a_m^T XD_wX^Ta_m-1\right)\\ &\quad -\mu_1 a_m^T a_1-\cdots-\mu_{m-1}a_m^T a_{m-1}\end{aligned}$$

The optimization is performed by $\frac{\partial O_m}{\partial a_m}=0$.

$$\begin{aligned}&2X(\beta L_b+(1-\beta)S_w)X^Ta_m-2\lambda XD_wX^Ta_m\\&-\mu_1a_1-\cdots-\mu_{m-1}a_{m-1}=0\end{aligned}\tag{9}$$

As the derivation process of the Ref.[15], multiplying the left side of (9) by a_m^T, we obtain

$$\begin{aligned}&2a_m^TX\left(\beta L_b+(1-\beta)S_w\right)X^Ta_m\\&-2\lambda a_m^TXD_wX^Ta_m=0\\&\Rightarrow\lambda=\frac{a_m^TX\left(\beta L_b+(1-\beta)S_w\right)X^Ta_m}{a_m^TXD_wX^Ta_m}\end{aligned}\tag{10}$$

Thus λ represents the expression to be maximized.

Multiplying the left side of (9) successively by $a_1^T\left(XD_wX^T\right)^{-1},\ldots,a_{m-1}^T\left(XD_wX^T\right)^{-1}$, we can obtain the following a set of $(m-1)$ equations:

$$\mu_1a_1^T\left(XD_wX^T\right)^{-1}a_1+\cdots+\mu_{m-1}a_1^T\left(XD_wX^T\right)^{-1}a_{m-1}$$
$$=2a_1^T\left(XD_wX^T\right)^{-1}X(\beta L_b+(1-\beta)S_w)X^Ta_m$$
$$\mu_1a_{21}^T\left(XD_wX^T\right)^{-1}a_1+\cdots+\mu_{m-1}a_2^T\left(XD_wX^T\right)^{-1}a_{m-1}$$
$$=2a_2^T\left(XD_wX^T\right)^{-1}X(\beta L_b+(1-\beta)S_w)X^Ta_m$$
$$\cdots\cdots$$
$$\mu_1a_{m-1}^T\left(XD_wX^T\right)^{-1}a_1+\cdots+\mu_{m-1}a_{m-1}^T\left(XD_wX^T\right)^{-1}a_{m-1}$$
$$=2a_{m-1}^T\left(XD_wX^T\right)^{-1}X(\beta L_b+(1-\beta)S_w)X^Ta_m$$

We use matrix notation define

$\mu^{(m-1)}=[\mu_1,\ldots,\mu_{m-1}]^T$, $A^{(m-1)}=[a_1,\ldots,a_{m-1}]^T$,

$V^{(m-1)}=\left[V_{ij}^{(m-1)}\right]^T=\left[A^{(m-1)}\right]^T\left(XD_wX^T\right)^{-1}A^{(m-1)}$,

$V_{ij}^{(m-1)}=a_i^T\left(XD_wX^T\right)^{-1}a_j$.

Using this simplified notation, the previous set of $(m-1)$ equations can be represented in a single matrix relationship

$$V^{(m-1)}\mu^{(m-1)}$$
$$=2\left[A^{(m-1)}\right]^T\left(XD_wX^T\right)^{-1}X(\beta L_b+(1-\beta)S_w)X^Ta_m$$

Thus

$$\begin{aligned}&\mu^{(m-1)}=\\&2\left[V^{(m-1)}\right]^{-1}\left[A^{(m-1)}\right]^T\left(XD_wX^T\right)^{-1}X(\beta L_b+(1-\beta)S_w)X^Ta_m\end{aligned}\tag{11}$$

Let us now multiply the left side of (9) by $\left(XD_wX^T\right)^{-1}$, we can obtain the following equation:

$$2\left(XD_wX^T\right)^{-1}X(\beta L_b+(1-\beta)S_w)X^Ta_m-2\lambda a_m$$
$$-\mu_1\left(XD_wX^T\right)^{-1}a_1-\cdots-\mu_{m-1}\left(XD_wX^T\right)^{-1}a_{m-1}=0$$

This can be expressed using matrix notation as

$$2\left(XD_wX^T\right)^{-1}X(\beta L_b+(1-\beta)S_w)X^Ta_m-2\lambda a_m$$
$$-\left(XD_wX^T\right)^{-1}A^{(m-1)}\mu^{(m-1)}=0$$

With (11) , we obtain

$$\left\{I-\left(XD_wX^T\right)^{-1}A^{(m-1)}\left[V^{(m-1)}\right]^{-1}\left[A^{(m-1)}\right]^T\right\}$$
$$\cdot\left(XD_wX^T\right)^{-1}X(\beta L_b+(1-\beta)S_w)X^Ta_m=\lambda a_m$$

As shown in (10), considering λ as the criterion to be maximized, thus a_m is the eigenvector of

$$U^{(m)}=\left\{I-\left(XD_wX^T\right)^{-1}A^{(m-1)}\left[V^{(m-1)}\right]^{-1}\left[A^{(m-1)}\right]^T\right\}$$
$$\cdot\left(XD_wX^T\right)^{-1}X(\beta L_b+(1-\beta)S_w)X^T$$

associated with the largest eigenvalue of $U^{(m)}$.

4)OMMP Embedding. Let $W_{OMMP}=[a_1,a_2,\ldots,a_m]$ be the solutions of (6), ordered according to their eigenvalues, $\lambda_1>\lambda_2>\ldots>\lambda_m$.Thus, the embedding is as follows:

$$x_i\rightarrow y_i=W^Tx_i\tag{12}$$

$$W=W_{PCA}W_{OMMP}\tag{13}$$

where y_i is a *q*-dimensional representation of the face image x_i,and W is the transformation matrix.

VI. EXPERIMENTAL RESULTS

In this section, several experiments are carried out to show the effectiveness of our proposed OMMP algorithm for face recognition. Three face databases are used in our experimental study, including Yale, ORL and CMU PIE databases. The proposed algorithm is compared with OLPP, MMP and two well-known linear methods, i.e., Eigenface method (PCA) [4], and Fisherface method (PCA+LDA)[5]. In our proposed OMMP algorithm, there are three parameter values to be set: k, β and γ. The parameter k denotes the number of nearest neighbors. The parameter β controls the weight between a within-class graph and a between-class graph. The parameter γ controls the weight between labeled and unlabeled images. As the MMP algorithm in [14], we empirically set them as $k = 5$, $\beta = 0.5$,and $\gamma = 50$ in the experiments.

In all the experiments, all the original images are realigned by fixing the locations of the two eyes, cropped, and resized to 32×32 pixels, with 256 gray levels per pixel. No further preprocessing is done. To perform the face recognition, we first compute the face subspace from the training samples; then the new face image to be identified is projected into the face subspace with our proposed OMMP algorithm; finally, the nearest neighbor classifier is adopted to identify new facial images, where the Euclidean metric is used as the distance measure.

A. Databases

The Yale face database (http://cvc.yale.edu/projects/ yalefaces/ yalefaces.html) contains 165 grayscale images of 15 individuals (each person has 11 different images) under various facial expressions and lighting conditions. For each individual, l (= 2, 3, 4, 5,6) images are randomly selected for training and the rest are used for testing. In general, the performance of all these algorithms mentioned previously varies with the number of dimensions. For each given l , we average the results over 20 random splits and report the maximal average recognition rate.

The ORL database (http://www.uk.research.att.com/ facedatabase.html) contains 400 images grouped into 40 distinct subjects with ten different images for each. The images were captured at different times, and for some objects, the images may vary in facial expression and facial detail. All the images were taken with a tolerance for some tilting and rotation of the face up to 20 degrees. The experimental protocol is the same as before.

The CMU PIE(Pose, Illumination and Expression) database (http://www.ews.uiuc.edu/ dengcai2/ Data/ data. html) contains more than 40,000 facial images of 68 people. The face images were captured by 13 synchronized cameras and 21 flashes, under varying pose, illumination, and expression. In this experiment, five near frontal poses (C05, C07, C09, C27, and C29) and the images under different illuminations, lighting and expressions are used such that each individual has 170 near frontal face images. For each individual, l (= 5, 10, 20, 30,40) images are randomly selected for training and the rest are used for testing. For each given l , we average the results over 20 random splits and report the best results.

B. Results

The best recognition rates as well as the optimal dimensionality obtained by each algorithm are reported on the Table （I-III）. These results are also showed in the Figure (1-3). From these experimental results, we can make the following observations:

1)Our proposed OMMP algorithm consistently outperforms Eigenface, Fisherface, OLPP and MMP algorithms.

2) The local-manifold-structure based algorithm(such as OLPP, MMP and OMMP) are consistently better than the global-Euclidean-structure based algorithm(such as Eigenface and Fisherface methods), which indicates that the OLPP,MMP and OMMP algorithms encode more discriminating information in the low-dimensional face subspace by preserving local structure which is more important than the global structure for classification.

3) The orthogonal basis function can improve face recognition accuracy for OMPP beyond the corresponding non-orthogonal algorithms. The results demonstrate that the OMMP can have more locality preserving power than OMMP. This result is consistent with the observation in [6] and [16] that the locality preserving power is directly related to the discriminating power. Thus the OMMP algorithm has more discriminating power than MPP.

4) Although both OLPP and OMMP are orthogonal manifold learning algorithms for face recognition, the average recognition accuracy of OMMP is always better than that of OLPP. The main reason could be attributed to the fact that OLPP may not well consider the between-locality property of the data points. For face recognition, it is hard to determine whether label information or local manifold structure is more important. By jointly considering the local manifold structure and the label information (utilized in both within-class graph and between-class graph), as well as characterizing the separability of different classes with the margin criterion, OMMP performs much better than OLPP in face recognition applications.

V. CONCLUSION

In this paper, we propose a subspace learning algorithm for face recognition, called orthogonal maximum margin projection (OMMP), to discover the underlying face manifold structure. OMMP considers both the local manifold structure and label information, as well as preserving the metric structure of face space with orthogonal basis functions. Experimental results on Yale, ORL, and PIE databases show the effectiveness of our proposed algorithm. However, OMMP is essentially linear. Our future work is to extend OMMP to nonlinear map with kernel trick.

APPENDIX A TABLE (I-III) AND FIGURE(1-3)

TABLE I.
RECOGNITION RATES COMPARISON ON THE YALE DATABASE

Algorithms	Recognition rate	Dimension
Eigenface	75.8%	33
Fisherface	80.5%	14
OLPP	82.6%	14
MMP	85.3%	15
OMMP	89.6%	14

TABLE II.
RECOGNITION RATES COMPARISON ON THE ORL DATABASE

Algorithms	Recognition rate	Dimension
Eigenface	86.2%	195
Fisherface	92.5%	39
OLPP	95.4%	59
MMP	96.8%	40
OMMP	98.6%	40

TABLE III.
RECOGNITION RATES COMPARISON ON THE CMU PIE DATABASE

Algorithms	Recognition rate	Dimension
Eigenface	80.6%	150
Fisherface	84.3%	67
OLPP	92.8%	421
MMP	94.5%	114
OMMP	97.6%	110

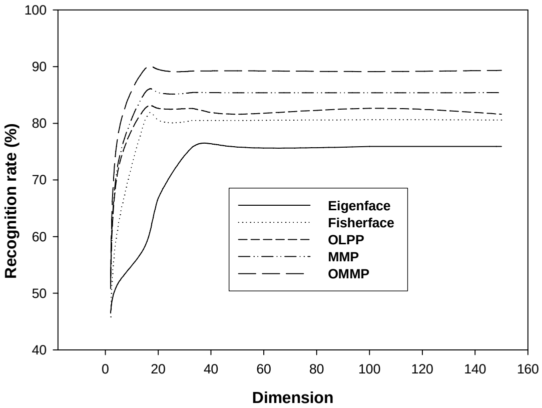

Figure 1. Recognition rate versus reduced dimensionality on the Yale database

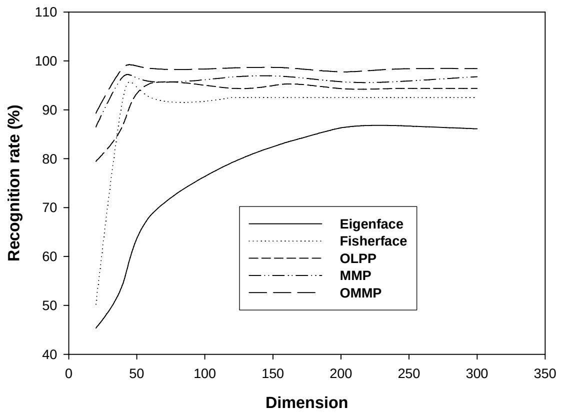

Figure 2. Recognition rate versus reduced dimensionality on the ORL database

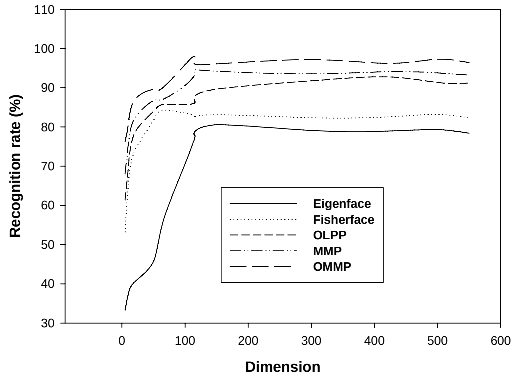

Figure 3. Recognition rate versus reduced dimensionality on the CMU PIE database

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] H. Murase, S.K.Nayar, “Visual learning and recognition of 3-d objects from appearance,” *International Journal of Computer Vision*, vol.14, pp.5-24,January 1995.

[2] R.O.Duda, P.E.Hart, and D. G. Stork, *Pattern Classification*, 2nd edition. Hoboken, NJ: Wiley-Interscience, 2000.

[3] K. Fukunaga, *Introduction to Statistical Pattern Recognition*, 2nd edition. New York: Academic Press,1990.

[4] M. Turk, A. Pentland, “Eigenfaces for recognition,” *Journal of Cognitive Neuroscience*, vol.3, pp.71-86, Winter 1991.

[5] P. N. Belhumeur, J. P. Hepanha, and D. J. Kriegman, “Eigenfaces vs. fisherfaces: recognition using class specific linear projection,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol.19, pp.711-720, July 1997.

[6] 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.

[7] 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.

[8] M.Sugiyama, “Dimensionality reduction of multimodal labeled data by local fisher discriminant analysis, ” *Journal of Machine Learning Research*, vol.8, pp.1027-1061,May 2007.

[9] 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.

[10] D.Xu, S.Yan, D.Tao, S.Lin, and H.-J.Zhang, “Marginal fisher analysis and its variants for human gait recognition

and content-based image retrieval," *IEEE Transactions on Image Processing*, vol.16,pp.2811-2821, November 2007.

[11] D.Cai, X.He, and J.Han, "SRDA: an efficient algorithm for large scale discriminant analysis," *IEEE Transactions on Knowledge and Data Engineering*, vol.20, pp.1-20, January 2008.

[12] D.Cai, X.He, and J.Han, "Document clustering using locality preserving indexing," *IEEE Transactions on Knowledge and Data Engineering*, vol.17, pp.1624-1637, December 2005.

[13] 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.

[14] D.Cai, X.He, and J.Han, "Learning a maximum margin subspace for image retrieval," *IEEE Transactions on Knowledge and Data Engineering*, vol.20, pp.189-201, February.2008.

[15] J. Duchene, S.Leclercq, "An optimal transformation for discriminant and principal component analysis," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol.10, pp. 978–983, November 1988.

[16] X. He,P. Niyogi, "Locality preserving projections," in *Advances in Neural Information Processing Systems*, Cambridge, MA:MIT Press,2003,pp.385-392.

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 a 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.