Journal of Theoretical and Applied Information Technology
15th November 2012. Vol. 45 No.1

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

ORTHOGONAL APPROXIMATELY HARMONIC PROJECTION FOR FACE RECOGNITION

[1, 2]XIA SUN, [2]ZIQIANG WANG, [2]LIJUN SUN

[1] School of Computer Science and Technology, Wuhan University of Technology, Wuhan, China

[2] School of Information Science and Engineering, Henan University of Technology, Zhengzhou, China

ABSTRACT

Face recognition has attracted growing attention for applications such as identity authentication and human-computer interface. However, a major challenge of face recognition is that the captured face image often lies in a high-dimensional feature space. To overcome the curse of dimensionality problem and improve the performance of face recognition, a novel manifold learning algorithm called orthogonal approximately harmonic projection (OMMP) is proposed in this paper. The OAHP algorithm is based on the harmonic projection (AHP) and explicitly considers the local geometrical structure and cluster structure of the face space. Meanwhile, the OAHP method can produce orthogonal basis vectors to preserve the metric structure of face space, which greatly enhances the discriminating power of the reduced lower-dimensional feature space. Experimental results on three face databases show that the proposed OAHP performs much better than related algorithms in terms of recognition rate.

Keywords: *Face Recognition, Orthogonal Approximately Harmonic Projection, Manifold Learning*

1 INTRODUCTION

In the past decades, face recognition has received extensive attention due to its potential applications in many fields, such as information surveillance, identity authentication, and human-computer interface. As a result, many face recognition algorithms have been proposed, and surveys in this area can be found in [1]. In general, a face image of size is represented as a vector in the image space . However, the image space is always of very high dimensionality, ranging from several hundreds to thousands. Therefore, it is often necessary to conduct dimensionality reduction to acquire an efficient and discriminative representation before formally conducting classification. In fact, face images can be usually considered as samples drawn from a low-dimensional manifold and artificially embedded in a high-dimensional ambient space. To find a meaningful low-dimensional representation of high-dimensional data, the most representative dimensionality reduction algorithms are principal component analysis (PCA) and linear discriminant analysis (LDA)[2].

PCA aims to reduce data dimensionality by performing a covariance analysis between factors, it projects the data along the directions where the data vary the most. Applying PCA algorithm to face recognition, Turk and Pentland[3] developed the well-known Eigenfaces method. Since PCA is an unsupervised method, it is only optimal with respect to presentation and reconstruction while not for discriminating one face class from others. Unlike PCA, LDA is a supervised dimensionality reduction algorithm. It aims at finding an optimal transformation that maps the data into a lower-dimensional space that minimizes the within-class scatter and simultaneously maximizes the between-class scatter, thus achieving maximum discrimination. LDA has been extensively applied to face recognition, and the popular Fisherfaces[2] method is found on the LDA algorithm. The drawback of LDA is that it requires large training sample size for good generalization. For face recognition, it is generally believed that algorithms based on LDA are superior to those based on PCA when sufficient labeled face images are provided. However, both PCA and LDA are designed for discovering only the global Euclidean structure, whereas the local manifold structure is ignored. In fact, a number of research efforts have shown that the face images possibly reside on a nonlinear submanifold[4-6]. To analyze the high-dimensional data that lie on or near a submanifold of the ambient space, many manifold learning-based dimensionality reduction algorithms have been proposed, such as isometric feature mapping (ISOMAP)[7], locally linear embedding(LLE)[8], and Laplacian

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

eigenmap(LE)[9]. ISOMAP aims to find the low-dimensional representations for a data set by approximately preserving the geodestic distances of the data pairs. LLE maps the data into a low-dimensional space by preserving the relationship between the neighboring points. LE aims to preserve proximity relationships by manipulations on an undirected weighted graph, which indicates neighbor relations of pairwise data points. These nonlinear manifold learning algorithms have achieved impressive results on some benchmark artificial data sets. Nevertheless, the nonlinearity makes them computationally expensive. In addition, they are defined only on the training data set, and it is unclear how the mapping can be computed for new testing points. Therefore, they are not suitable for face recognition. The approximately harmonic projection (AHP)[10] method is recently proposed to model the local manifold structure. AHP is a linear manifold learning method based on the harmonic framework, and the optimal transformation can be obtained by approximating the Dirichlet integral. However, the basis vectors obtained by the AHP method are not orthogonal, which makes it difficult to reconstruct the data.

To cope with the above drawback of AHP, we propose a new manifold learning algorithm termed orthogonal AHP (OAHP) for face recognition. OAHP is fundamentally based on the AHP method. It uses the approximate affine hull of the nearest neighbor graph to model the local geometrical structure of face manifold. The projection vectors are then obtained by solving a generalized eigenvalue problem. Similar to the AHP, the OAHP algorithm can also preserve the local geometrical structure, but meanwhile it requires the basis vectors to be orthogonal, which makes it more effective for preserving the intrinsic geometrical structure and the metric structure of the face space. Furthermore, our experimental results show that OAHP have more locality preserving power than AHP. In fact, previous researches have shown that locality preserving power is directly related to the discriminating power. Therefore, OAHP might be optimal in discriminating face images with different classes which is the ultimate goal of face recognition.

The remainder of the paper is organized as follows. In section 2, we provide a brief review of the AHP algorithm. Section 3 introduces our proposed OAHP algorithm for face recognition. The experimental results on face recognition are reported in Section 4. Finally, we present the conclusions in Section 5.

2 BRIEF REVIEW OF AHP

AHP is a recently proposed linear manifold learning method for dimensionality reduction [10]. It is based on the approximate affine hull and explicitly utilizes the edge length to reflect the geometrical structure of the manifold structure of the data space.

Given a set of face images $\{x_1,\cdots,x_n\}\subset\mathbb{R}^m$, let $X=[x_1,\cdots,x_n]$. Let W^c and W^b be two weight matrices defined on the face images. The objective function of AHP is defined as follows:

$$\begin{aligned} a_{opt} &= \arg\min_a \frac{1}{2}\sum_{i\sim j}\frac{1}{d_{ij}}\left(a^T x_i - a^T x_j\right)^2 \\ &= \arg\min_a a^T X\left(D^c - W^c\right)X^T a \end{aligned} \tag{1}$$

with the constraint

$$\begin{aligned} &3\sum_{i\sim j}\int_0^{d_{ij}}\left(a^T x_i + \frac{t}{d_{ij}}\left(a^T x_i - a^T x_j\right)\right)^2 dt \\ &= a^T X\left(2D^b + W^b\right)X^T a = 1 \end{aligned} \tag{2}$$

where d_{ij} denotes the length of the edge between x_i and x_j, W^c and W^b are two matrices defined as follows: if x_i and x_j are connected, then $W_{ij}^c = 1/d_{ij}$ and $W_{ij}^b = d_{ij}$; otherwise, $W_{ij}^c = W_{ij}^b = 0$. D^c and D^b are two diagonal matrices defined as $D_{ii}^c = \sum_j W_{ij}^c$, $D_{ii}^b = \sum_j W_{ij}^b$.

The objective function in AHP aims to use the approximate affine hull of the graph to separate data points sampled from different components. Therefore, minimizing it is to ensure that if x_i and x_j lie the multiple connected components, then $y_i\left(=a^T x_i\right)$ and $y_j\left(=a^T x_j\right)$ are made close by the optimal projection. Finally, the projection vector a that minimizes (1) is given by the minimum eigenvalue solution to the following generalized eigenvalue problem:

$$X\left(D^c - W^c\right)X^T a = \lambda X\left(2D^b + W^b\right)X^T a \tag{3}$$

Note that, to avoid the singularity problem existed in AHP, one may first apply PCA to remove the components corresponding to zero eigenvalues. Thus, the projection vector of AHP can be considered as the eigenvectors of the matrix $\left(X\left(2D^b + W^b\right)X^T\right)^{-1} X\left(D^c - W^c\right)X^T$ associated

Journal of Theoretical and Applied Information Technology
15th November 2012. Vol. 45 No.1

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

with the smallest eigenvalues. In addition, since $\left(X\left(2D^b+W^b\right)X^T\right)^{-1}X\left(D^c-W^c\right)X^T$ is not usually symmetric, the AHP projection axes are not orthogonal, which makes it less effective for preserving the metric structure of the face space.

3 OAHP ALGORITHM FOR FACE RECOGNITION

In this section, we propose a novel manifold learning algorithm termed orthogonal AHP (OAHP) and apply it to the face recognition.

It is worth noting that the singular 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 space ($m\times m$). To cope with this issue, we first project each face image into the PCA subspace by removing the components corresponding to zero eigenvalues, so that the matrix $X\left(2D^b+W^b\right)X^T$ becomes nonsingular. Another consideration of using PCA as preprocessing is for noise reduction. The algorithmic procedure of OAHP is formally stated as follows.

1) PCA projection. We project each face image x_i into the PCA subspace by throwing away the components corresponding to zero eigenvalue. For simplicity, we still adopt x to denote the image in the PCA subspace and let A_{PCA} denote the transformation matrix of PCA in the following steps.

2) Constructing the nearest neighbor graph and computing the edge weight. Let graph G represent a graph with n nodes, where each node i denotes a face image x_i. We construct an edge between nodes i and j if x_i is among the k nearest neighbors of x_j or x_j is among the k nearest neighbors of x_i. Similar to the AHP method, OAHP preserves the geometrical structure of the graph by simultaneously using edge length and orientation. For each edge, let $e_{ij}=x_j-x_i$ denote the edge vector which has an orientation from x_i to x_j, and the edge gradient is calculated according to

$$\nabla f_{e_{ij}}=\frac{a^Tx_j-a^Tx_i}{d_{ij}} \tag{4}$$

where $d_{ij}=\left\|x_j-x_i\right\|$ is the edge length. Then, two edge weight matrices W^c and W^b of the graph G are defined as follows: if x_i and x_j are connected, then $W_{ij}^c=1/d_{ij}$ and $W_{ij}^b=d_{ij}$; otherwise, $W_{ij}^c=W_{ij}^b=0$. In addition, D^c and D^b are two diagonal matrices defined as $D_{ii}^c=\sum_j W_{ij}^c$, $D_{ii}^b=\sum_j W_{ij}^b$.

3) Computing the orthogonal basis vectors of OAHP. The projection vector a that minimizes (1) under the constraint (2) is given by the eigenvectors associated with the smallest eigenvalues of the following generalized eigen-problem:

$$X\left(D^c-W^c\right)X^Ta=\lambda X\left(2D^b+W^b\right)X^Ta \tag{5}$$

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

$$a_1=\arg\min_a\frac{a^TX\left(D^c-W^c\right)X^Ta}{a^TX\left(2D^b+W^b\right)X^Ta} \tag{6}$$

and

$$a_d=\arg\min_a\frac{a^TX\left(D^c-W^c\right)X^Ta}{a^TX\left(2D^b+W^b\right)X^Ta} \tag{7}$$

with the constraint

$$a_d^Ta_1=a_d^Ta_2=\cdots=a_d^Ta_{d-1}=0 \tag{8}$$

Since the $X\left(2D^b+W^b\right)X^T$ is positive definite after PCA projection, following the strategy suggested in [6] and [11], we can also normalize it such that $a^TX\left(2D^b+W^b\right)X^Ta=1$ for any a. Then the above minimization problem can be equivalently transformed into the following optimal objective function:

$$a_{opt}=\arg\min_a a^TX\left(D^c-W^c\right)X^Ta \tag{9}$$

with the constraint

$$a^TX\left(2D^b+W^b\right)X^Ta=1 \tag{10}$$

$$a_d^Ta_1=a_d^Ta_2=\cdots=a_d^Ta_{d-1}=0 \tag{11}$$

where a_1 is the eigenvector of the matrix $\left(X\left(2D^b+W^b\right)X^T\right)^{-1}X\left(D^c-W^c\right)X^T$ associated with the smallest eigenvalue.

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

In order to compute the d-th basis vector that minimizes the above optimal objective function, the Lagrange multiplier is used to transform (9) to include the constraints (10) and (11):

$$\begin{aligned} L_d &= a_d^T X\left(D^c - W^c\right)X^T a_d \\ &\quad -\lambda\left(a_d^T X\left(2D^b + W^b\right)X^T a_d - 1\right) \\ &\quad -\mu_1 a_d^T a_1 - \cdots - \mu_{d-1} a_d^T a_{d-1} \end{aligned} \tag{12}$$

The optimization can be performed by $\partial L_d / \partial a_d = 0$, then we have

$$\begin{aligned} &2X\left(D^c - W^c\right)X^T a_d - 2\lambda X\left(2D^b + W^b\right)X^T a_d \\ &-\mu_1 a_1 - \cdots - \mu_{d-1} a_{d-1} = 0 \end{aligned} \tag{13}$$

Multiplying the left side of (13) successively by $a_1^T\left(X\left(2D^b+W^b\right)X^T\right)^{-1}, \ldots, a_{d-1}^T\left(X\left(2D^b+W^b\right)X^T\right)^{-1}$, we can obtain the following set of $(d-1)$ equations:

$$\begin{aligned} &\mu_1 a_1^T\left(X\left(2D^b+W^b\right)X^T\right)^{-1} a_1 \\ &+\cdots-\mu_{d-1} a_1^T\left(X\left(2D^b+W^b\right)X^T\right)^{-1} a_{d-1} \\ &= 2a_1^T\left(X\left(2D^b+W^b\right)X^T\right)^{-1} X\left(D^c-W^c\right)X^T a_d \end{aligned}$$

$$\begin{aligned} &\mu_1 a_2^T\left(X\left(2D^b+W^b\right)X^T\right)^{-1} a_1 \\ &+\cdots-\mu_{d-1} a_2^T\left(X\left(2D^b+W^b\right)X^T\right)^{-1} a_{d-1} \\ &= 2a_2^T\left(X\left(2D^b+W^b\right)X^T\right)^{-1} X\left(D^c-W^c\right)X^T a_d \end{aligned}$$

$$\cdots\cdots$$

$$\begin{aligned} &\mu_1 a_{d-1}^T\left(X\left(2D^b+W^b\right)X^T\right)^{-1} a_1 \\ &+\cdots-\mu_{d-1} a_{d-1}^T\left(X\left(2D^b+W^b\right)X^T\right)^{-1} a_{d-1} \\ &= 2a_{d-1}^T\left(X\left(2D^b+W^b\right)X^T\right)^{-1} X\left(D^c-W^c\right)X^T a_d \end{aligned}$$

If we define the following notions,

$$\mu^{(d-1)} = \left[\mu_1, \ldots, \mu_{d-1}\right]^T \tag{14}$$

$$A^{(d-1)} = \left[a_1, \ldots, a_{d-1}\right] \tag{15}$$

$$\begin{aligned} B^{(d-1)} &= \left[B_{ij}^{(d-1)}\right] \\ &= \left[A^{(d-1)}\right]^T \left(X\left(2D^b+W^b\right)X^T\right)^{-1} A^{(d-1)} \end{aligned} \tag{16}$$

$$B_{ij}^{(d-1)} = a_i^T\left(X\left(2D^b+W^b\right)X^T\right)^{-1} a_j \tag{17}$$

Then by using the above notions (14)-(17), the previous set of $(d-1)$ equations can be represented in a single matrix relationship

$$\begin{aligned} &B^{(d-1)}\mu^{(d-1)} \\ &= 2\left[A^{(d-1)}\right]^T\left(X\left(2D^b+W^b\right)X^T\right)^{-1} X\left(D^c-W^c\right)X^T a_d \end{aligned} \tag{18}$$

Consequently, we have

$$\begin{aligned} &\mu^{(d-1)} = 2\left[B^{(d-1)}\right]^{-1}\left[A^{(d-1)}\right]^T \\ &\times\left(X\left(2D^b+W^b\right)X^T\right)^{-1} X\left(D^c-W^c\right)X^T a_d \end{aligned} \tag{19}$$

In addition, multiplying the left side of (13) by $\left(X\left(2D^b+W^b\right)X^T\right)^{-1}$, we can obtain

$$\begin{aligned} &2\left(X\left(2D^b+W^b\right)X^T\right)^{-1} X\left(D^c-W^c\right)X^T a_d \\ &-2\lambda a_d - \mu_1\left(X\left(2D^b+W^b\right)X^T\right)^{-1} a_1 \\ &-\cdots-\mu_{d-1}\left(X\left(2D^b+W^b\right)X^T\right)^{-1} a_{d-1} = 0 \end{aligned} \tag{20}$$

which can be simply represented as follows by using the above notions (14)-(17):

$$\begin{aligned} &2\left(X\left(2D^b+W^b\right)X^T\right)^{-1} X\left(D^c-W^c\right)X^T a_d \\ &-2\lambda a_d - \left(X\left(2D^b+W^b\right)X^T\right)^{-1} A^{(d-1)}\mu^{(d-1)} = 0 \end{aligned} \tag{21}$$

By combining (19) and (21), we can obtain

$$\begin{aligned} &\left\{I - \left(X\left(2D^b+W^b\right)X^T\right)^{-1} A^{(d-1)}\left[B^{(d-1)}\right]^{-1}\left[A^{(d-1)}\right]^T\right\} \\ &\cdot\left(X\left(2D^b+W^b\right)X^T\right)^{-1} X\left(D^c-W^c\right)X^T a_d = \lambda a_d \end{aligned}$$

Thus, the basis vector a_d can be regarded as the eigenvector of the matrix

$$\begin{aligned} &\left\{I - \left(X\left(2D^b+W^b\right)X^T\right)^{-1} A^{(d-1)}\left[B^{(d-1)}\right]^{-1}\left[A^{(d-1)}\right]^T\right\} \\ &\cdot\left(X\left(2D^b+W^b\right)X^T\right)^{-1} X\left(D^c-W^c\right)X^T \end{aligned}$$

associated with the smallest eigenvalue. Finally, we can obtain the optimal orthogonal basis vectors $\{a_1, a_2, \cdots, a_d\}$.

Journal of Theoretical and Applied Information Technology
15th November 2012. Vol. 45 No.1

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

4) OAHP projection. Let $A_{OAHP}=[a_1,a_2,\cdots,a_d]$, then the lower-dimensional feature presentation y of the face image x via OAHP can be represented as follows:

$$x \to y = \left(A_{PCA}A_{OAHP}\right)^T x \tag{22}$$

5) Face recognition in the lower-dimensional feature space. Once we obtain the lower-dimensional feature representations of face images by using (22). Face recognition becomes a pattern classification task. Thus, we can apply the nearest neighbor classifier to identify different facial images.

Note that the orthogonal basis vectors of OAHP can preserve the metric structure of the data space. We simply prove this argument in the following. Let $A_{OAHP}=[a_1,a_2,\cdots,a_d]$ be the transformation matrix, then the Euclidean distance between two data points in the reduced feature space can be calculated as follows:

$$\begin{aligned} D(y_i,y_j) &= \|y_i - y_j\| \\ &= \|A_{OAHP}^T x_i - A_{OAHP}^T x_j\| \\ &= \|A_{OAHP}^T (x_i - x_j)\| \\ &= \sqrt{(x_i-x_j)^T A_{OAHP}A_{OAHP}^T (x_i-x_j)} \end{aligned} \tag{23}$$

Since A_{OAHP} is an orthogonal matrix, i.e., $A_{OAHP}A_{OAHP}^T = I$, the metric structure of the data space is preserved.

The proposed OAHP algorithm tries to preserve the local manifold structure by minimizing the objective function (1) under the orthogonal constraint of the basis vectors. A face image is transformed into the local geometry preserving subspace for recognition. Previous researches have shown that the eigenvalues of the subspace can reflect the locality preserving power [4-6], and smaller eigenvalues has more local geometry preserving power. Figure 1 shows the eigenvalues of AHP and OAHP on the ORL face database. As can be seen, the eigenvalues of OAHP are consistently smaller than those of AHP, which indicates that OAHP has more local geometry preserving power than AHP.

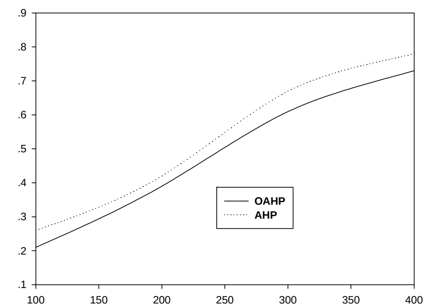

Figure 1: Eigenvalues Of AHP and OAHP

4 EXPERIMENTAL RESULTS

In this section, three experiments have been conducted to show the effectiveness of our proposed OAHP method on three face databases. The OAHP method is compared with the PCA, LDA, and AHP methods. For AHP and OAHP, we adopt the same nearest neighbor graph structure. Meanwhile, we first apply PCA to avoid the singular problem before using LDA, AHP, and OAHP methods.

In this study, three bench-mark face databases were used for testing: the Yale database (http: //cvc. yale.edu/projects/yalefaces/yalefaces.html), the Olivetti Research Laboratory (ORL) database (http://www.uk.research.att.com/facedatabase.html), and the CMU PIE (pose, illumination, and expression) database[12]. In all the experiments, the following preprocessing steps were applied: First, all of color images are converted into gray ones. Then, the centers of the eyes of an face image are manually detected, aligned, cropped, and re-sized to 32×32, which is further normalized to zero mean and unit variance. Some sample images after preprocessing of the three face databases are shown in Figure 2 to Figure 4, respectively.


Figure 2: Face Image Examples Of The Yale Database


Figure 3: Face Image Examples Of The ORL Database


Figure 4: Face Image Examples Of The CMU PIE Database

Journal of Theoretical and Applied Information Technology
15th November 2012. Vol. 45 No.1

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

In short, 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 Yale face database was constructed at the Yale Center for Computational Vision and Control. It contains 165 gray scale images of 15 individuals. The images demonstrate variations in lighting condition, facial expression (normal, happy, sad, sleepy, surprised, and wink). In this test, recognition rates were determined by the "leaving-one out" strategy: To classify an image of an individual, that image was first removed from the data set. Then, for each evaluation, 10 rounds of experiments are repeated with random selection of the training data, and the average result is recorded as final recognition rates. The recognition rates and the optimal dimensionality obtained by PCA, LDA, AHP, and OAHP are shown in Table 1. The best results occur when using 50, 14, 15, 15 dimensions for PCA, LDA, AHP, and OAHP, respectively. The recognition rates of PCA, LDA, AHP, and OAHP are 75.6%, 92.8%, 94.9%, and 96.8%, respectively. As can be seen, the OAHP method outperforms the original AHP method and performs the best among the compared algorithms. Figure 5 shows the plots of recognition rate versus different reduced dimensionality on the Yale face database.

Table I: Performance Comparisons On The Yale Database

Method	Recognition rate	Dimensionality
PCA	75.6%	50
LDA	92.8%	14
AHP	94.9%	15
OAHP	96.8%	15

Table II: Performance Comparisons On The ORL Database

Method	Recognition rate	Dimensionality
PCA	86.1%	190
LDA	92.3%	39
AHP	93.5%	40
OAHP	96.7%	50

Table III: Performance Comparisons On The CMU PIE Database

Method	Recognition rate	Dimensionality
PCA	82.4%	150
LDA	94.6%	67
AHP	95.2%	90
OAHP	97.4%	120

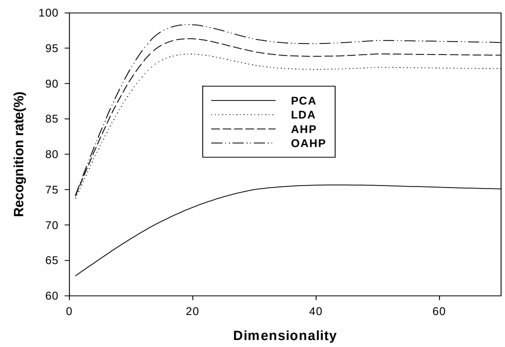

Figure 5: Recognition Rate Versus Reduced Dimensionality On The Yale Database

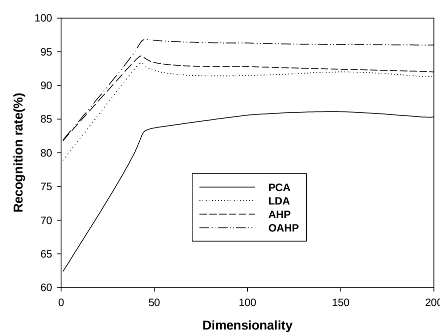

Figure 6: Recognition Rate Versus Reduced Dimensionality On The ORL Database

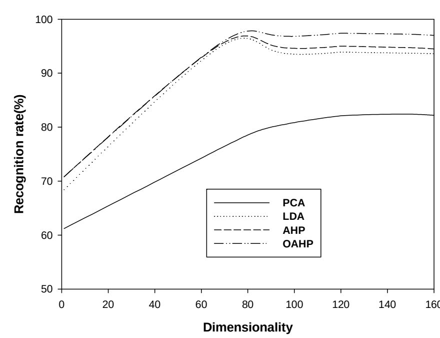

Figure 7: Recognition Rate Versus Reduced Dimensionality On The CMU PIE Database

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

The ORL face database contains a set of face images taken at the Olivetti Research Laboratory in Cambridge. It contains 400 images of 40 individuals. Some images were captured at different times and have different variations including expression, lighting and facial details (glasses/no glasses). The images were taken with a tolerance for some tilting and rotation of the face up to 20 degrees. In this test, a random subset with five images per individual was chosen to form training set, and the rest of the database was considered to be the testing set. In order to obtain steady results, we average the results over ten random splits. The recognition results are shown in Table 2. As can be seen, our proposed OAHP consistently outperforms the PCA, LDA, and AHP methods. OAHP has the best performance and the maximum achieved recognition rate is 96.7% when keeping 50 dimensions. For the PCA and LDA methods, the best recognition rates are 86.1% and 92.3%, respectively. While for AHP, it is 93.5%. Figure 6 depicts the plots of recognition rate versus different reduced dimensionality on the ORL face database.

The CMU-PIE face database contains 68 subjects with 41368 face images as a whole. The face images were captured by 13 synchronized cameras and 21 flashes under variations in pose, illumination, and expression. We choose the five near frontal poses (C05, C07, C09, C27, C29) and use all the images under different illuminations, lighting and expressions which leaves us 170 near frontal face images for each individual. Within the 170 face images for each individual in this test, a random set with 100 face images per individual are used for training and the other 70 for testing. We average the results over 10 random splits. Table 3 shows the recognition results. As can be seen, the best results occur when using 150, 67, 90, 120 dimensions for PCA, LDA, AHP, and OAHP, respectively. The recognition rates of PCA, LDA, AHP, and OAHP are 82.4%, 94.6%, 95.2%, and 97.4%, respectively. Therefore, the OAHP method achieves the best performance among the compared methods. Figure 7 shows the plots of recognition rate versus different reduced dimensionality on the CMU PIE face database.

In summary, from the above experimental results, we can make the following observations.

1) Our proposed OAHP algorithm consistently outperforms PCA, LDA, and AHP algorithms, which demonstrates that the orthogonal basis vector constraint can effectively enhance the performance of OAHP algorithm.

2) The PCA algorithm gives relatively poor performance since it is an unsupervised learning method.

3) The manifold learning-based algorithms, i.e., AHP and OAHP, perform much better than conventional algorithms, i.e., PCA and LDA. The possible explanations are as follows: Both PCA and LDA can only discover the global Euclidean structure, while both AHP and OAHP can encode more discriminating information by preserving local manifold structure which is more important than the global Euclidean structure for classification.

4) Although the AHP method performs much better than PCA and LDA methods by using local geometrical structure and cluster structure for face recognition, it still performs worse than our proposed OAHP method. This result demonstrates that OAHP can have more locality preserving power than AHP by enforcing the orthogonal basis vector constraint, which is consistent with the observation in [5] and [6] that the locality preserving power is directly related to the discriminating power. Thus the OAHP method has more discriminating power than AHP.

5 CONCLUSION AND FUTURE WORK

In this paper, we have proposed a novel manifold learning algorithm for face recognition, called orthogonal approximately harmonic projection (OAHP). It combines the locality preserving power of approximately harmonic projection (AHP) and orthogonal basis vectors constraint to provide an effective approach for dimensionality reduction. The experimental results on three face databases show that the proposed algorithm performs better than other related algorithms. However, OAHP is essentially linear. Our future work is to extend OAHP to nonlinear map with kernel trick.

ACKNOWLEDGMENTS

This work is supported by NSFC (Grant No. 70701013), the National Science Foundation for Post-doctoral Scientists of China (Grant No. 2011M500035), and the Specialized Research Fund for the Doctoral Program of Higher Education of China (Grant No.20110023110002).

REFERENCES:

[1] W.Zhao, R. Chellappa, P. Phillips, and A. Rosenfeld, "Face recognition: a literature survey", ACM Computing Surveys, Vol.35, No.4, 2003, pp.399-458.

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

[2] P.Belhumeur, J. Hespanha, and D. Kriegman, "Eigenfaces vs. fisherfaces: recognition using class specific linear projection," IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol.19, No.7, 1997, pp.711-720.

[3] M.A.Turk and A.P.Pentland, "Eigenfaces for recognition", Journal of Cognitive Neuroscience, Vol.3, No.1, 1991, pp.71-86.

[4] X.He, S.Yan, Y.Hu, P.Niyogi, and H.-J. Zhang, "Face recognition using Laplacianfaces", IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol.27, No.3, 2005, pp.328-340.

[5] Taiping Zhang, Bin Fang, Yuan Yan Tang, Guanghui He, and Jing Wen, "Topology preserving non-negative matrix factorization for face recognition", IEEE Transactions on Image Processing, Vol.17,No.4,2008,pp.574-584.

[6] Deng Cai, Xiaofei He, Jiawei Han, and HongJiang Zhang, "Orthogonal Laplacianfaces for face recognition", IEEE Transactions on Image Processing, Vol.15,No.11,2006,pp.3608-3614.

[7] J. Tenenbaum, V. Silva, and J. Langford, "A global geometric framework for nonlinear dimensionality reduction", Science, Vol.290, No.5500, 2000, pp.2319-2323.

[8] S.Roweis and L. Saul, "Nonlinear dimensionality reduction by locally linear embedding", Science, Vol.290, No.5500, 2000, pp.2323-2326.

[9] M.Belkin and P.Niyogi, "Laplacian eigenmaps for dimensionality reduction and data representation", Neural Computation, Vol.15, No.6, 2003, pp.1373-1396.

[10] Binbin Lin, Xiaofei He, Yuan Zhou, Ligang Liu, and Ke Lu, "Approximately harmonic projection: theoretical analysis and an algorithm", Pattern Recognition, Vol.43, No.10, 2010, pp.3307-3313.

[11] J. Duchene and S. Leclercq, "An optimal transformation for discriminant and principal component analysis", IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol.10, No.6, 1988,pp.978-983.

[12] T. Sim, S. Baker, and M. Bsat, "The CMU pose, illuminlation, and expression database", IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol.25, No.12,2003,pp.1615-1618.