Orthogonal Locality Sensitive Discriminant Analysis for Face Recognition*

YI JIN AND QIU-QI RUAN
Institute of Information Science
Beijing Jiaotong University
Beijing, 100044 P.R. China

An innovative appearance-based method that called Orthogonal Locality Sensitive Discriminant Analysis is presented for face recognition in this paper. Our algorithm is based on the Locality Sensitive Discriminant Analysis (LSDA) algorithm, which aims at finding a projection by maximizing the margin between data points from different classes at each local area. However, a major disadvantage of LSDA is that LSDA is non- orthogonal, and this makes it difficult to estimate the intrinsic dimensionality and to reconstruct the face data. Non-orthogonality distorts the local geometrical structure of the data submanifold. In this paper, an Orthogonal LSDA algorithm is proposed to preserve the local geometrical structure by computing the mutually orthogonal basis functions iteratively. Since it produces orthogonal basis functions and can have more local structure preserving power, it is expected to have more discriminating power than LSDA. Experiments based on the ORL and Yale face database show the impressive performance of the proposed method. Results show that our new algorithm outperforms the other popular approaches reported in the literature and achieves a much higher recognition rate.

Keywords: face recognition, locality preserving projections, locality sensitive discriminant analysis (LSDA), orthogonal basis functions, orthogonal locality sensitive discriminant analysis (OLSDA)

1. INTRODUCTION

The use of automatic face recognition technologies has increased substantially in the past several decades due to new areas of application such as human-computer intelligent interaction, access control, telecommunication and digital libraries, and smart environments. Feature extraction is a crucial step in face recognition and one of the successful classes of algorithms is appearance-based statistical algorithms. A number of linear subspace algorithms have been proposed for feature extraction in face recognition, for example, principal component analysis (PCA) [1], linear discriminant analysis (LDA) [2], independent component analysis (ICA) [3] and so on.

However, these methods effectively see only a linear manifold that based on the Euclidean structure. They fail to capture the structure which lies on a nonlinear submanifold hidden in the image space. Recently, more and more studies have shown that the face images likely exist in a nonlinear sub-manifold [4-6] which is an emerging research area. Therefore, some manifold learning algorithms which aim at preserving the local nonlinear structure of the face data have attracted more attention, examples of these

---

Received August 31, 2007; revised January 7 & April 23, 2008; accepted May 2, 2008.
Communicated by H. Y. Mark Liao.
* This paper was supported partly by the National Natural Science Foundation of China under grants No. 60472033 and No. 60672062.

algorithms being Isometric feature mapping (Isomap) [7], Locally Linear Embedding (LLE) [8], and Locality Preserving Projections (LPP) [9-11]. The first two algorithms are nonlinear but the LPP is a linear dimensionality reduction algorithm. For face image processing, we are especially interested in linear techniques attending considerations of computational load and recognition speed.

The LPP method aims to preserve the local structure at each neighborhood. Experiments show that LPP is able to extract nonlinear features in the local and nonlinear manifold and can thus perform better in face recognition [11]. Although the LPP method has many outstanding characteristics, it has inevitable disadvantages. In the first place, it is unsupervised in nature and fails to discover the discriminant structure in the data points. Secondly, the basis functions of LPP obtained by Laplacianface method are non-orthogonal [12]. This makes it difficult to reconstruct the face data. Many algorithms which aim to overcome the aforementioned disadvantages of LPP have been proposed such as Orthogonal Laplacianfaces [12], Discriminant Locality Preserving Projections (DLPP) [15], Locality Sensitive Discriminant Analysis (LSDA) [16], Local Discriminant Embedding (LDE) [17] and Marginal Fisher Analysis (MFA) [18]. Orthogonal Laplacianfaces uses orthogonal basis functions [12-14] to preserve the metric structure of the nonlinear submanifold space, but it ignores the crucial information between classes. LDE and MFA are essentially the same for both of them consider the between-class relations and the within-class relations are equally important. DLPP changes the objective function by introducing between-class scatter constraint into the objective function, while, by using the class labels, LSDA constructs a nearest graph which splits to a within-class graph and a between-class graph to model the local geometrical structure [16].

In this paper, an innovative manifold learning-based supervised algorithm for feature extraction and recognition named Orthogonal Locality Sensitive Discriminant Analysis (OLSDA) is presented. To take advantage of the strengths of Orthogonal Laplacianfaces and Locality Sensitive Discriminant Analysis and, at the same time, to avoid their disadvantages, we first map the data points into a subspace which optimally preserves not only the local neighborhood information but discriminant information as well. In particular, the margin between data points from different classes is maximized at each local neighborhood. Then an orthogonal basis function based constraint is added into the objective function of LSDA to emphasize the discriminant information. OLSDA shares the same locality preserving character as LSDA, but at the same time it requires the basis functions to be orthogonal and preserves the metric structure of the face space [12]. Experiments based on the ORL and Yale face database demonstrate that the new algorithm has more local structure preserving power than LSDA, Laplacianfaces (LPP), and Orthogonal Laplacianfaces (OLPP). Since it has been shown that the locality preserving power is directly related to the discriminating power [12], the OLSDA is expected to have more discriminating power than the other three algorithms.

The paper is organized as follows. In section 2, a brief review of the algorithm LSDA is given. Section 3 introduces the novel face recognition method-Orthogonal Locality Sensitive Discriminant Analysis (OLSDA). Experimental results for recognition using the ORL and Yale face database are given in section 4, and some concluding remarks are provided in section 5.

2. OVERVIEW OF THE LOCALITY SENSITIVE DISCRIMINANT ANALYSIS

In this section, a brief overview of LSDA [16] is given. LSDA shares some resemblances with LPP algorithm which uses graph to model the geometrical structure of the data points. The graph constructed by LSDA is split into two, within-class graph and between-class graph, by means of which, the discriminant structure is preserved.

Given m data points $\{x_1, x_2, \ldots, x_m\} \subset R^n$ sampled from the underlying submanifold M, two graphs, *i.e*. the within-class graph S_w and the between-class graph S_b can be constructed to discover both geometrical and discriminant structure of the data manifold. $l(x_i)$ is the class label of x_i. Therefore there are two subsets of k nearest neighbors, $N_w(x_i)$ and $N_b(x_i)$. $N_w(x_i)$ contains the neighbors sharing the same label with x_i, while $N_b(x_i)$ contains the neighbors having the different labels. Therefore,

$$\begin{aligned} N_w(x_i) &= \{x_i^j \mid l(x_i^j) = l(x_i), 1 \le j \le k\}, \\ N_b(x_i) &= \{x_i^j \mid l(x_i^j) \ne l(x_i), 1 \le j \le k\}. \end{aligned} \tag{1}$$

And the weight matrices of S_w and S_b can be defined as:

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

First by PCA projection, the extracted features are statistically uncorrelated by throwing away the components corresponding to zero eigenvalues. And the data extracted by the PCA lies in the new subspace for further feature extraction. Then a linear transformation $y_i = W^T x_i$ is found from the d-dimensional space to a line. The criterion of LSDA is to optimize the following two objections, where y_i is the one-dimensional representation of x_i:

$$\min \sum_{ij} (y_i - y_j)^2 S_{w,ij} \tag{3}$$

$$\max \sum_{ij} (y_i - y_j)^2 S_{b,ij} \tag{4}$$

under appropriate constrains. By algebra formulation, the objective Eqs. (3) and (4) can be reduced to:

$$\begin{aligned} \frac{1}{2}\sum_{ij} (y_i - y_j)^2 S_{w,ij} &= \sum_i w^T x_i D_{w,ii} x_i^T w - \sum_i w^T x_i S_{w,ij} x_j^T w \\ &= w^T X D_w X^T w - w^T X S_w X^T w \\ &= w^T X L_w X^T w \end{aligned} \tag{5}$$

and

$$\frac{1}{2}\sum_{ij}(y_i - y_j)^2 S_{b,ij} = \sum_i w^T x_i D_{b,ii} x_i^T w - \sum_i w^T x_i S_{b,ij} x_j^T w$$
$$= w^T X D_b X^T w - w^T X S_b X^T w \quad (6)$$
$$= w^T X L_b X^T w$$

where D_w and D_b are diagonal matrices; their values are the column (or row, since S_w and S_b are symmetric) sum of S_w and S_b, respectively, $D_{w,ii} = \sum_j S_{w,ij}$ and $D_{b,ii} = \sum_j S_{b,ij}$. $L_b = D_b - S_b$ is the Laplacian matrix of the between class graph. Considering that the value of D_w is not as high as possible, a constraint is imposed as follows:

$$y^T D_w y = 1 \Rightarrow w^T X D_w X^T w = 1. \quad (7)$$

Eq. (5) changes into the following:

$$w = \arg\min_w (1 - w^T X S_w X^T w) = \arg\max_w w^T X S_w X^T w. \quad (8)$$

Eq. (6) can then be written as follows:

$$w = \arg\max_w w^T X L_b X^T w. \quad (9)$$

Finally, a combined optimization problem is reduced to finding:

$$w = \arg\max_w w^T X H X^T w, \text{ s.t. } w^T X D_w X^T w = 1 \quad (10)$$

where $H = \alpha L_b + (1 - \alpha)S_w$ and α is a suitable constant verifying $0 \le \alpha \le 1$.

Thus, the transformation vector w that maximizes the objective function is obtained by maximizing the generalized eigenvalues problem:

$$XHX^T w = \lambda X D_w X^T w. \quad (11)$$

After some pre-processing steps on X, the basis functions of LSDA can also be regarded as the eigenvectors of the matrix $(XD_wX^T)^{-1}XHX^T$ associated with the largest eigenvalues, though XD_wX^T is non-singular in LSDA. Since $(XD_wX^T)^{-1}XHX^T$ is not symmetric in general, the basis functions of LSDA are non-orthogonal [12]. Let $W = [w_1, w_2, \ldots, w_k]$ be the transformation matrix by calculating the eigenvector $w_1, w_2, \ldots, w_k$ according to its k largest eigenvalues of Eq. (11), the linear mapping can be obtained as:

$$x_i \to y_i = W^T x_i, W = [w_1, w_2, \ldots, w_k]. \quad (12)$$

And the Euclidean distance between two data points in the reduced space can be computed as follows: $dist(y_i, y_j) = ||y_i - y_j|| = ||W^T(x_i - x_j)|| = \sqrt{(x_i - x_j)^T WW^T (x_i - x_j)}$. If W is an orthogonal matrix and $WW^T = I$, the metric structure is preserved. Actually, it is XD_wX^T-orthogonal for the constraint $w^TXD_wX^Tw = I$. Since the eigenvectors are non-orthogonal with each other, the Euclidean submanifold of the original images cannot be preserved completely.

3. THE ORTHOGONAL LOCALITY SENSITIVE DISCRIMINANT ANALYSIS

As mentioned before, one of the problems of LSDA is that it is essentially non-orthogonal, and therefore fails to preserve the metric structure of the face space. In this section, the new appearance-based Orthogonal Locality Sensitive Discriminant Analysis is induced in section 3.1 and the theoretical justifications of our algorithm will be presented in section 3.2.

3.1 The Algorithm of OLSDA

One problem which often arises in face recognition is that sometimes the matrix is singular. To overcome the problem, a pre-processing step using PCA is applied in the OLSDA algorithm.

1. PCA projection as a pre-processing step:

The face data is projected into the PCA subspace by the transformation matrix W_{PCA}. By PCA projection, the extracted features are statistically uncorrelated by throwing away the components corresponding to zero eigenvalues. And the data extracted by the PCA lies in the new subspace for further feature extraction and dimension reduction using the OLSAD algorithm.

2. Constructing the adjacency graph:

Let G_w and G_b denote the within-class graph and between-class graph. Correspondingly, there are two subsets containing the n nearest neighbors, within-class neighbors and between-class neighbors, $N_w(x_i)$ and $N_b(x_i)$. The ith node responds to the face data x_i. An edge is put between node i and node j if x_i is among the n nearest neighbors of x_j or x_j is among the n nearest neighbors of x_i.

3. Choosing the weight:

If node i and node j are connected, t is a variable parameter, and then weight can be calculated as follows:

$$S_{ij} = \exp(-||x_i - x_j||^2/t). \tag{13}$$

Otherwise, put $S_{ij} = 0$. Thus,

$$S_{w,ij} = \begin{cases} \exp(-||x_i - x_j||^2/t), & \text{if } x_i \in N_w(x_j) \text{ or } x_j \in N_w(x_i) \\ 0, & \text{otherwise} \end{cases},$$

$$S_{b,ij} = \begin{cases} \exp(-\|x_i - x_j\|^2/t), & \text{if } x_i \in N_b(x_j) \text{ or } x_j \in N_b(x_i) \\ 0, & \text{otherwise} \end{cases}. \tag{14}$$

The weight matrices $S_{w,ij}$ of graph G_w and $S_{b,ij}$ of graph G_b model the local and discriminant structure of the face manifold.

4. Computing the orthogonal basis functions of the OLSDA algorithm:

Let R^k be the k orthogonal basis vectors, we define:

$$P^{(k-1)} = [w_1, w_2, \ldots, w_{k-1}],$$
$$Q^{(k-1)} = [P^{(k-1)}]^T(XD_wX)^{-1}P^{(k-1)}. \tag{15}$$

The orthogonal basis vectors can then be determined as:

(1) The eigenvector w_1 of $(XD_wX^T)^{-1}XHX^T$ according to its k largest eigevalue.
(2) The eigenvector w_k of

$$R^k = \{I - (XDX^T)^{-1}P^{(k-1)}[Q^{(k-1)}]^{-1}[P^{(k-1)}]^T\} \cdot (XD_wX^T)^{-1}XHX^T \tag{16}$$

according to its largest eigenvalues.

Here $H = \alpha L_b + (1 - \alpha)S_w$ where α is a suitable constant verifying $0 \le \alpha \le 1$.

5. OLSDA mapping:

Considering $W_{OLSDA} = [w_1, w_2, \ldots, w_l]$, the linear mapping of the OLSDA is defined as follows:

$$x_i \to y_i = W^T x_i,\ W = W_{PCA}W_{OLSDA}. \tag{17}$$

3.2 The Justification of OLSDA

In this section, a theoretical foundation of the OLSDA algorithm is provided. Since the basic functions of LSDA are non-orthogonal, our main work is to find a set of orthogonal basis vectors that maximizes the locality sensitive discriminant function while at the same time preserving the locality and discriminant power of the projection map of w Eq. (17). Thus, the objective function of OLSDA is presented as follows:

$$w_1 = \arg\max_w \frac{w^T XHX^T w}{w^T XD_w X^T w}$$

and,

$$w_k = \arg\max_w \frac{w^T XHX^T w}{w^T XD_w X^T w}$$

with the constraint:

$$w_k^T w_1 = w_k^T w_2 = \cdots = w_k^T w_{k-1} = 0.$$

Note that, because XD_wX^T is non-singular after the PCA projection, a constraint $w^TXD_wX^Tw = 1$ is added to keep the ratio of w^TXHX^T and $w^TXD_wX^Tw$ unchanged. Obviously, w_1 can be computed as the eigenvector of the generalized eigenproblem $XHX^Tw = \lambda XD_wX^Tw$ associated with its largest eigenvalues, and in order to get the kth basis vector, we maximize the following objective function:

$$w = \arg\max_w \frac{w^T XHX^T w}{w^T XD_w X^T w}, \quad \text{s.t. } w_k^T w_1 = w_k^T w_2 = \cdots = w_k^T w_{k-1} = 0,\ w^T XD_w X^T w = 1. \tag{18}$$

The Lagrange theorem is used to solve the above objective function, and the objective function is transformed into:

$$T^{(k)} = w_k^T XHX^T w_k - \lambda(w_k^T XD_w X^T w_k - 1) - \mu w_k^T w_1 - \cdots - \mu_{k-1} w_k^T w_{k-1}. \tag{19}$$

The optimization is performed by setting the partial derivative of $T^{(k)}$ with respect to w_k to zero:

$$\frac{\partial T^{(k)}}{\partial w^k} = 0 \Rightarrow 2XHX^T w_k - 2\lambda XD_w X^T w_k - \mu_1 w_1 - \cdots - \mu_{k-1} w_{k-1} = 0. \tag{20}$$

For $w_k^T w_1 = w_k^T w_2 = \cdots = w_k^T w_{k-1} = 0,$ we obtain the following by multiplying the left side of Eq. (20) by w_k^T:

$$2w_k^T XHX^T w_k - 2\lambda XD_w X^T w_k = 0 \Rightarrow \lambda = \frac{w_k^T XHX^T w_k}{XD_w X^T w_k}. \tag{21}$$

Multiplying the left side of Eq. (20) by $w_k^T (XD_w X^T)^{-1}, \cdots, w_{k-1}^T (XD_w X^T)^{-1},$ a set of $(k-1)$ equations is obtained:

$$\mu_1 w_1^T (XD_w X^T)^{-1} w_1 + \cdots + \mu_{k-1} w_1^T (XD_w X^T)^{-1} w_{k-1} = 2w_1^T (XD_w X^T)^{-1} XHX^T w_k,$$
$$\mu_1 w_2^T (XD_w X^T)^{-1} w_1 + \cdots + \mu_{k-1} w_2^T (XD_w X^T)^{-1} w_{k-1} = 2w_2^T (XD_w X^T)^{-1} XHX^T w_k,$$
$$\cdots\cdots$$
$$\mu_1 w_{k-1}^T (XD_w X^T)^{-1} w_1 + \cdots + \mu_{k-1} w_{k-1}^T (XD_w X^T)^{-1} w_{k-1} = 2w_{k-1}^T (XD_w X^T)^{-1} XHX^T w_k.$$

The following can be defined:

$\mu^{(k-1)} = [\mu_1, \ldots, \mu_{k-1}]^T$, $P^{(k-1)} = [w_1, w_2, \ldots, w_{k-1}]$,

$$Q^{(k\text{-}1)} = [Q_{ij}^{(k\text{-}1)}] = [P^{(k\text{-}1)}]^T(XD_wX)^{-1}P^{(k\text{-}1)},\ Q_{ij}^{(k\text{-}1)} = w_i^T(XD_wX^T)^{-1}w_j.$$

Which, expressed as a matrix gives:

$$Q^{(k\text{-}1)}\mu^{(k\text{-}1)} = 2[P^{(k\text{-}1)}]^T(XD_wX^T)^{-1}XHX^Tw_k.$$

Thus,

$$\mu^{(k\text{-}1)} = 2[Q^{(k\text{-}1)}]^{-1}[P^{(k\text{-}1)}]^T(XD_wX^T)^{-1}XHX^Tw_k. \tag{22}$$

Then multiplying the left side of Eq. (20) by $(XD_wX^T)^{-1}$

$$\Rightarrow 2(XD_wX^T)^{-1}XHX^Tw_k - 2\lambda w_k - \mu_1(XD_wX^T)^{-1}w_1 - \ldots - \mu_{k\text{-}1}(XD_wX^T)^{-1}w_{k\text{-}1} = 0.$$

The matrix notation expression is:

$$2(XD_wX^T)^{-1}XHX^Tw_k - 2\lambda w_k - (XD_wX^T)^{-1}P^{(k\text{-}1)}\mu^{(k\text{-}1)} = 0. \tag{23}$$

With Eq. (22), we get the following function:

$$\{I - (XDX^T)^{-1}P^{(k\text{-}1)}[Q^{(k\text{-}1)}]^{-1}[P^{(k\text{-}1)}]^T\} \cdot (XD_wX^T)^{-1}XHX^Tw_k = \lambda w_k. \tag{24}$$

λ is just the criterion to be maximized, thus w_k is the eigenvector of

$$R^k = \{I - (XDX^T)^{-1}P^{(k\text{-}1)}[Q^{(k\text{-}1)}]^{-1}[P^{(k\text{-}1)}]^T\} \cdot (XD_wX^T)^{-1}XHX^T$$

according to its largest eigenvalues, where $H = \alpha L_b + (1 - \alpha)S_w$ and α is a suitable constant verifying $0 \leq \alpha \leq 1$.

Therefore, the orthogonal basis can be obtained by computing the eigenvector of $R^kw_k = \lambda w_k$ associated with its largest eigenvalues. This is the Orthogonal Locality Sensitive Discriminant Analysis (OLSDA) proposed in this paper.

4. EXPERIMENTAL RESULTS

In this section, two popular face databases, the ORL and the YALE, are used to test the performance of the proposed Locality Sensitive Discriminant Analysis (OLSDA) algorithm. For each database, we randomly select $n_p = 3, 5$ samples of each individual for training, and use the remaining facial images for testing. The images in all the experiments were cropped into a size of 32×32 pixels, with 256 gray levels per pixel. Thus, each image can be represented by a 1024-dimensional vector in image space. No further preprocessing has been done in our experiments. To evaluate the efficiency of the OLSDA algorithm, we compare the accurate recognition rate of the proposed method with that of other appearance-based methods such as Eigenfaces (PCA), Fisherfaces (LDA), Laplacianfaces (LPP) and Locality Sensitive Discriminant Analysis (LSDA) and so on.

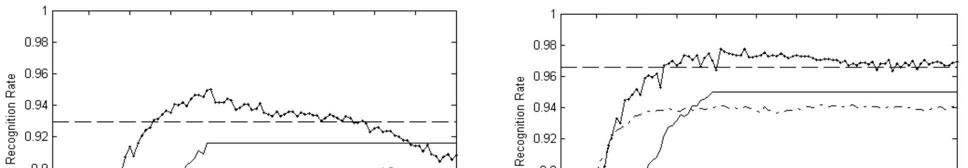
4.1 Experiments on the ORL Database

In this section, we use the ORL face database to evaluate the performance of the OLSDA algorithm in comparison with other traditional face recognition algorithms. ORL database consists of a total of 400 face images from 40 individuals, each providing 10 different images. The images were captured at different times and have different variations including expressions and decorations. 10 sample images of one individual are displayed in Fig. 1.

Fig. 1. Ten sample images from ORL database.

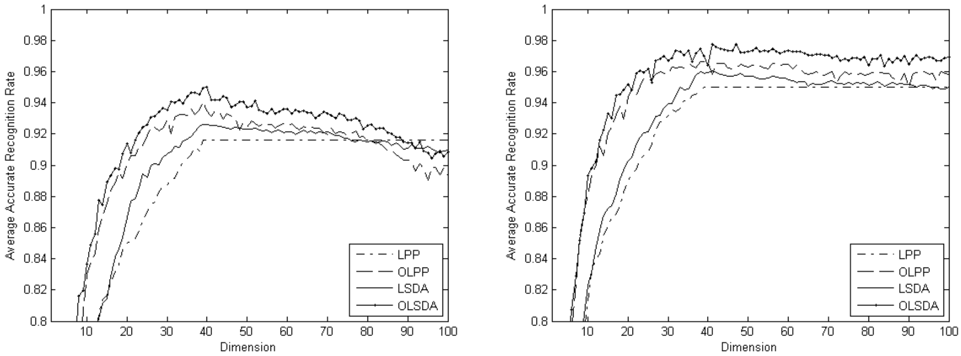
For each person, n_p = 3, 5 sample images are randomly selected for training and the rest are used for testing. Firstly, the face subspace learned from the training set is obtained by calculating the projective matrix. Secondly, the new face image is projected into the d-dimensional subspace. Finally, the recognition is performed in the new subspace by nearest neighbor classifier. In OLSDA algorithm, there are three parameters that need to be determined, the coefficient α, the nearest neighbor n and the dimension of OLSDA k. Since it is very complicated to determine the parameters, we adopt a stepwise selection strategy. In our experiment, the coefficient α is 0.4 and 0.1; the nearest neighbor n is 2 and 4 separately when n_p is 3 and 5. For each n_p, an average recognition rate is gained by repeating this experiment 50 random splits.

In order to show the effectiveness of the proposed OLSDA algorithm, a set of experiments is carried out to compare the performance of the OLSDA algorithm with Eigenfaces (PCA), Fisherfaces (LDA) and Laplacianfaces (LPP). Fig. 2 shows how the average recognition rate varies with the first 100 dimensions of the four face subspace when n_p = 3, 5 sample images are selected respectively.



(a) n_p = 3. (b) n_p = 5.

Fig. 2. Comparison of average recognition rates with four different algorithm (PCA, LDA, LPP, LSDA) under different dimension on the ORL database.



(a) n_p = 3. (b) n_p = 5.

Fig. 3. Comparison of average recognition rates with four different algorithm (LPP, OLPP, LSDA, OLSDA) under different dimension on the ORL database.

Table 1. Comparison of recognition rates with different algorithms when n_p = 3, 5 sample images are randomly selected on the ORL database.

(a) n_p = 3.

Method	Average Recognition Rate	Dimension
PCA	90.05%	82
LDA	92.98%	39
LPP	91.61%	39
MFA	92.14%	39
OLPP	94.09%	39
LSDA	92.59%	39
OLSDA	94.98%	39

(b) n_p = 5.

Method	Average Recognition Rate	Dimension
PCA	94.18%	66
LDA	96.60%	39
LPP	94.97%	39
MFA	95.97%	39
OLPP	96.65%	38
LSDA	96.05%	41
OLSDA	97.72%	39

As it is shown in Fig. 2, our algorithm outperformed all the other three methods. Results show that the OLSDA algorithm can get the highest recognition rate of the three, followed by LDA and LPP, and PCA performance being the worst.

Next, to show the efficacy of the orthogonal basis functions, another set of experiments is performed to compare the orthogonal basis function based methods and the non-orthogonal methods. In Fig. 3, the OLSDA method is compared with LSDA, LPP and OLPP methods and it shows how the average recognition rate varies with the dimensions of the four algorithms.

As we can see in Fig. 3, the orthogonal basis function based methods outperform the non-orthogonal methods in general, and the OLSDA algorithm again achieves the top average recognition rate of them all. Table 1 compares the proposed new method with PCA, LDA, LPP, MFA *etc*. It shows the average recognition rate and the dimension with the top average recognition rate between different algorithms.

It can be seen that the average recognition rate obtained by the new OLSDA algorithm is 94.98% and 97.72% with n_p = 3, 5 respectively. Both are higher than those obtained using any of the other six algorithms.

4.2 Experiments on the Yale Database

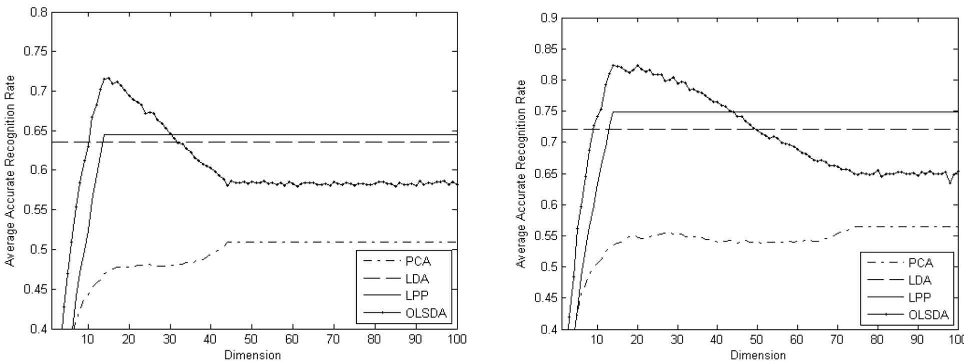
In this section, the Yale face database is utilized to test the performance of the OLSDA algorithm. The Yale face database is constructed at the Yale Center and contains 165 grayscale images of 15 people. The images show variations in lighting conditions, facial expressions and the use or non-use of glasses. Fig. 4 shows 10 sample images of one individual of the Yale face database.

Fig. 4. Ten sample images from Yale database.

The methodology of the experiment is the same as described in the case of the ORL database. For each individual, there are n_p = 3, 5 sample images randomly selected for training and the rest are used for testing. And a stepwise selection strategy is performed in our experiments. The optimal parameters are achieved as the values of the coefficient α are 0.5 and 0.4; the nearest neighbor n is 1 and 4 respectively when n_p is 3 and 5. For each n_p, the experiments are also randomly repeated 50 times and an average recognition rate is gained.

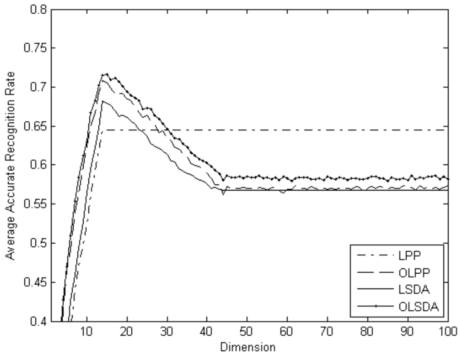
As with the experiments with the ORL database, there are also two sets of experiments to test the effectiveness of the proposed OLSDA algorithm. Firstly, experiments are carried out to compare the performance of the OLSDA algorithm with Eigenfaces (PCA), Fisherfaces (LDA) and Laplacianfaces (LPP). Fig. 5 shows the average recognition rate varies with the first 100 dimensions of the four face subspace when n_p = 3, 5 sample images are selected respectively.

Next, we will compare the performance levels of the different orthogonal basis function based methods. Face recognition is implemented on the Yale face database using four different algorithms, LPP, OLPP, LSDA and OLSDA. As it can be seen in Fig. 5, the OLSDA method is compared with LSDA, LPP and OLPP and it shows how the average recognition rate varies with the dimensions of face subspace.

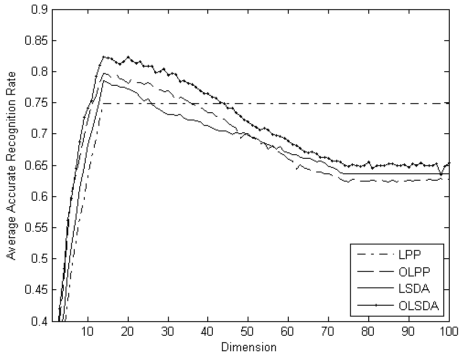


(a) $n_p = 3$. (b) $n_p = 5$.

Fig. 5. Comparison of average recognition rates with four different algorithm (PCA, LDA, LPP, LSDA) under different dimension on the Yale database.



(a) n_p = 3.



(b) n_p = 5.

Fig. 6. Comparison of average recognition rates with four different algorithm (LPP, OLPP, LSDA, OLSDA) under different dimension on the Yale database.

Table 2. Comparison of recognition rates with different algorithm when n_p = 3, 5 sample images are randomly selected on the Yale database.

(a) n_p = 3.

Method	Average Recognition Rate	Dimension
PCA	50.09%	44
LDA	63.53%	14
LPP	64.50%	14
MFA	67.62%	14
OLPP	70.81%	14
LSDA	68.22%	14
OLSDA	71.60%	14

(b) n_p = 5.

Method	Average Recognition Rate	Dimension
PCA	56.42%	74
LDA	72.16%	14
LPP	74.89%	14
MFA	78.18%	14
OLPP	79.73%	14
LSDA	78.51%	14
OLSDA	82.26%	14

In Fig. 6, recognition results on the Yale database also proves that the orthogonal basis function based methods outperform the non-orthogonal methods in general, and the OLSDA algorithm in particular achieves the top average recognition rate of the four.

Table 2 compares the proposed OLSDA method with PCA, LDA, LPP, MFA *etc*. It shows the details about the average recognition rate and the dimension with the top average recognition rate between different algorithms.

The average recognition rate achieved by the new OLSDA algorithm is 71.60% and 82.26% on the Yale database with n_p = 3, 5 respectively. Both are higher than those obtained using any the other six algorithms.

4.3 Discussion

Several experiments on two standard face databases have been conducted. There are some observations that need to be pointed out:

1. As a new manifold learning-based method, the Orthogonal Locality Sensitive Dis-

criminant Analysis (OLSDA) aims to detect the intrinsic structure from the face data and the results also demonstrate its superiority in performance. This is property due to the fact that OLSDA produces orthogonal basis functions and can have therefore more local structure preserving power.

2. The average recognition rate obtained by orthogonal basis functions based methods is much higher than that of the non-orthogonal methods in general. This shows that orthogonal basis functions based methods are effective in preserving the metric structure of the face space to a certain extent. However, Tables 1 and 2 shows that the average recognition rate can be improved differently based on the two different face databases. It is likely that the new algorithm is more robust against variations of face position than it is to variations in illumination conditions or different facial expressions.
3. In all these experiments on the two databases, the optimal dimensionality obtained by the OLSDA algorithm basically appears at the $c - 1$, where c is the number of classes. Consequently, more experiments in feature need to show whether we could simply project the face images into a $c - 1$ dimensional sub-manifold in practice in order to reduce the computational complexity.

5. CONCLUSION

In this paper, a supervised algorithm named Orthogonal Locality Sensitive Discriminant Analysis (OLSDA) has been developed for feature selection and dimensionality reduction of face recognition. In the new algorithm, the local and discriminant information is strengthened by adding an orthogonal basis functions based constraint to LSDA. Thus, OLSDA shares the same discriminant and local geometrical preserving character as LSDA, but at the same time it preserves the metric structure of the face space and as a consequence is more powerful than the LSDA algorithm for classification or clustering tasks. Experimental results on the two popular face databases demonstrate that the OLSDA is more effective than some popular methods in use today, such as PCA, LDA, LPP, LSDA *etc*. The results also show the superiority of the orthogonal basis functions based methods in comparison with the non-orthogonal based methods.

REFERENCES

1. M. Turk and A. Pentland, "Eigenfaces for recognition," *Journal of Cognitive Neuroscience*, Vol. 3, 1991, pp. 72-86.
2. P. N. Belhumeur, J. P. Hespanha, and D. J. Kriegman, "Eigenfaces vs. Fisherfaces: recognition using class specific linear projection," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 19, 1997, pp. 711-720.
3. M. S. Bartlett, H. M. Lades, and T. J. Sejnowski, "Independent component representations for face recognition," in *Proceedings of SPIE Symposiums on Electronic Imaging: Science and Technology, Conference on Human Vision and Electronic Imaging III*, Vol. 3299, 1998, pp. 528-539.
4. H. S. Seung and D. D. Lee, "The manifold ways of perception," *Science*, Vol. 290, 2000, pp. 2258-2259.
5. J. Zhang, S. Z. Li, and J. Wang, "Manifold learning and applications in recognition,"

Intelligent Multimedia Processing with Soft Computing, 2004, pp. 281-300.
6. G. Shakhnarovich and B. Moghaddam, *Face Recognition in Subspaces*, Springer, New York, 2004.
7. Y. Bengio, J. Palement, P. Vincent, O. Delalleau, N. L. Roux, and M. Ouimet, “Out-of-sample extensions for LLE, Isomap, MDS, Eigenmaps, and spectral clustering,” in *Proceedings of the 18th Annual Conference on Neural Information Processing Systems*, Vol. 16, 2003, pp. 177-184.
8. S. T. Roweis and L. K. Saul, “Nonlinear dimensionality reduction by locally linear embedding,” *Science*, Vol. 290, 2000, pp. 2323-2326.
9. M. Belkin and P. Niyogi, “Laplacian Eigenmaps and spectral techniques for embedding and clustering,” in *Proceedings of Conference on Advances in Neural Information Processing System*, Vol. 15, 2001, pp. 585-591.
10. X. He, S. Yan, Y. Hu, and H. J. Zhang, “Learning a locality preserving subspace for visual recognition,” in *Proceedings of the 9th IEEE International Conference on Computer Vision*, Vol. 1, 2003, pp. 385-392.
11. 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, 2005, pp. 328-340.
12. D. Cai, X. He, J. Han, and H. J. Zhang, “Orthogonal laplacianfaces for face recognition,” *IEEE Transactions on Image Processing*, Vol. 15, 2006, pp. 3608-3614.
13. D. Cai and X. He, “Orthogonal locality preserving indexing,” in *Proceedings of the 28th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, 2005, pp. 3-10.
14. K. Effrosyni and S. Yousef. “Orthogonal neighborhood preserving projections: A projection-based dimensionality reduction technique,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 29, 2007, pp. 2143-2156.
15. W. Yu, X. Teng, and C. Liu, “Face recognition using discriminant locality preserving projections,” *Image and Vision Computing*, Vol. 24, 2006, pp. 239-248.
16. D. Cai, X. He, K. Zhou, J. Han, and H. Bao, “Locality sensitive discriminant analysis,” in *Proceedings of the 20th International Joint Conference on Artificial Intelligence*, 2007, pp. 708-713.
17. H. T. Chen, H. W. Chang, and T. L. Liu, “Local discriminant embedding and its variants,” in *Proceedings of International Conference on Computer Vision and Pattern Recognition*, Vol. 2, 2005, pp. 846-853.
18. S. Yan, D. Xu, B. Zhang, and H. J. Zhang, “Graph embedding: A general framework for dimensionality reduction,” in *Proceedings of Internationl Conference on Computer Vision and Pattern Recognition*, Vol. 2, 2005, pp. 830-837.

Yi Jin (金一) was born in Heibei, P.R. China, 1982. She received her B.S. degree in Communication Engineering from China University of Geology. She is currently pursuing the Ph.D. degree at Institute of Information Science, Beijing Jiaotong University. The main research interests include image recognition and image processing.

Qiuqi Ruan (阮秋琦) was born in 1944. He received the B.S. and M.S. degrees from Beijing Jiaotong University, P.R. China in 1969 and 1981, respectively. From January 1987 to May 1990, he was a visiting scholar in the University of Pittsburgh, and the University of Cincinnati. Subsequently, he has been a visiting professor in USA for several times. He has published two books and more than 100 papers, and achieved a national patent. Now he is a professor, doctorate supervisor. He is a senior member of IEEE. His main research inte\rests include digital signal processing, computer vision, pattern recognition, and virtual reality, *etc*.