

Research Article

Direct Neighborhood Discriminant Analysis for Face Recognition

Miao Cheng, Bin Fang, Yuan Yan Tang, and Jing Wen

Department of Computer Science, Chongqing University, Chongqing 400030, China

Correspondence should be addressed to Bin Fang, fb@cqu.edu.cn

Received 21 March 2008; Accepted 16 April 2008

Recommended by Ming Li

Face recognition is a challenging problem in computer vision and pattern recognition. Recently, many local geometrical structure-based techniques are presented to obtain the low-dimensional representation of face images with enhanced discriminatory power. However, these methods suffer from the small sample size (SSS) problem or the high computation complexity of high-dimensional data. To overcome these problems, we propose a novel local manifold structure learning method for face recognition, named direct neighborhood discriminant analysis (DNDA), which separates the nearby samples of interclass and preserves the local within-class geometry in two steps, respectively. In addition, the PCA preprocessing to reduce dimension to a large extent is not needed in DNDA avoiding loss of discriminative information. Experiments conducted on ORL, Yale, and UMIST face databases show the effectiveness of the proposed method.

Copyright © 2008 Miao Cheng et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

1. Introduction

Many pattern recognition and data mining problems involve data in very high-dimensional spaces. In the past few decades, face recognition (FR) has become one of the most active topics in machine vision and pattern recognition, where the feature dimension of data usually can be very large and hardly handled directly. To get a high recognition rate for FR, numerous feature extraction and dimension reduction methods have been proposed to find the low-dimensional feature representation with enhanced discriminatory power. Among these methods, two state-of-the-art FR methods, principle component analysis (PCA) [1], and linear discriminant analysis (LDA) [2] have been proved to be useful tools for dimensionality reduction and feature extraction.

LDA is a popular supervised feature extraction technique for pattern recognition, which intends to find a set of projective direction to maximize the between-class scatter matrix S_b and minimize the within-class scatter matrix S_w simultaneously. Although successful in many cases, many LDA-based algorithms suffer from the so-called "small sample size" (SSS) problem that exists when the number of available samples is much smaller

than the dimensionality of the samples, which is particularly problematic in FR applications. To solve this problem, many extensions of LDA have been developed in the past. Generally, these approaches to address SSS problem can be divided into three categories, namely, Fisherface method, Regularization methods, and Subspace methods. Fisherface methods incorporate a PCA step into the LDA framework as a preprocessing step. Then LDA is performed in the lower dimensional PCA subspace [2], where the within-class scatter matrix is no longer singular. Regularization methods [3, 4] add a scaled identity matrix to scatter matrix so that the perturbed scatter matrix becomes nonsingular. However, Chen et al. [5] have proved that the null space of S_w contains the most discriminate information, while the SSS problem takes place, and proposed the null space LDA (NLDA) method which only extracts the discriminant features present in the null space of the S_w . Later, Yu and Yang [6] utilized discriminatory information of both S_b and S_w , and proposed a direct-LDA (DLDA) method to solve SSS problem.

Recently, the motivation for finding the manifold structure in high-dimensionality data elevates the wide application of manifold learning in data mining and machine learning. Among these methods, Isomap [7], LLE [8], and Laplacian eigenmaps [9, 10] are representative techniques. Based on the locality preserving concept, some excellent local embedding analysis techniques are proposed to find the manifold structure based on local nearby data [11, 12]. However, these methods are designed to preserve the local geometrical structure of original high-dimensional data in the lower dimensional space rather than good discrimination ability. In order to get a better classification effect, some supervised learning techniques are proposed by incorporating the discriminant information into the locality preserve learning techniques [13–15]. Moreover, Yan et al. [15] explain the manifold learning techniques and the traditional dimensionality reduction methods as a unified framework that can be defined in a graph embedding way instead of a kernel view [16]. However, the SSS problem is still exists in the graph embedding-based discriminant techniques. To deal with such problem, PCA is usually performed to reduce dimension as a preprocessing step in such environment [11, 15].

In this paper, we present a two-stage feature extraction technique named direct neighborhood discriminant analysis (DNDA). Compared to other geometrical structure learning work, the PCA step is not needed to be done in our method. Thus, more discriminant information can be kept for FR purpose, and as a result improved performance is expected. The rest of the paper is structured as follows: we give a brief review of LDA and DLDA in Section 2. We then introduce in Section 3 the proposed method for dimensionality reduction and feature extraction in FR. The effectiveness of our method is evaluated in a set of FR experiments in Section 4. Finally, we give concluding remarks in Section 5.

2. Review of LDA and DLDA

2.1. LDA

LDA is a very popular technique for linear feature extraction and dimensionality reduction [2], which chooses the basis vectors of the transformed space as those directions of the original space to make the ratio of the between-class scatter and the within-class scatter are maximized. Formally, the goal of LDA is to seek the optimal orthogonal matrix w , such that maximizing the following quotient, the Fisher Criterion:

$$J(W) = \arg \max_w \frac{w^T S_b w}{w^T S_w w}, \quad (2.1)$$

where S_b is the between-class scatter matrix, S_w is the within-class scatter matrix, such that w can be formed by the set of generalized eigenvectors corresponding to following eigenanalysis problem:

$$S_b w = \lambda S_w w. \quad (2.2)$$

When the inverse of S_w exists, the generalize vectors can be obtained by eigenvalue decomposition of $S_w^{-1} S_b$. However, one usually confronts the difficulty that the within-class scatter matrix S_w is singular (SSS) in FR problem. The so-called PCA plus LDA approach [2] is a very popular technique which intends to overcome such circumstances.

2.2. DLDA

To take discriminant information of both S_b and S_w into account without conducting PCA, a direct LDA (DLDA) technique has been presented by Yu and Yang [6]. The basic idea behind the approach is that no significant information will be lost if the null space of S_b is discarded. Based on the assumption, it can be concluded that the optimal discriminant features exist in the range space of S_b .

Let multiclass classification be considered, given a data matrix $X \in R^{d \times N}$, where each column x_i represents a sample data. Suppose X is composed of c classes and total number of samples is denoted by $\sum_{i=1}^c N_i = N$, for the i th class consists of N_i samples. Then, the between-class scatter matrix is defined as

$$S_b = \frac{1}{N} \sum_{i=1}^c N_i (\mu_i - \mu) (\mu_i - \mu)^T = G_b G_b^T, \quad (2.3)$$

where

$$G_b = \left[\frac{\sqrt{N_1}}{\sqrt{N}} (\mu_1 - \mu), \frac{\sqrt{N_2}}{\sqrt{N}} (\mu_2 - \mu), \dots, \frac{\sqrt{N_c}}{\sqrt{N}} (\mu_c - \mu) \right], \quad (2.4)$$

$$\mu_i = \frac{1}{N_i} \sum_{m=1}^{N_i} x_m \quad (2.5)$$

are the class mean sample, and

$$\mu = \frac{1}{N} \sum_{i=1}^c N_i \mu_i \quad (2.6)$$

denotes the total mean sample. Similarly, the within-class scatter matrix is defined as

$$S_w = \frac{1}{N} \sum_{i=1}^c \sum_{j=1}^{N_i} (x_j - \mu_i) (x_j - \mu_i)^T = G_w G_w^T, \quad (2.7)$$

where,

$$G_w = \left[\frac{1}{\sqrt{N}}(x_1 - \mu_{c_1}), \frac{1}{\sqrt{N}}(x_2 - \mu_{c_2}), \dots, \frac{1}{\sqrt{N}}(x_N - \mu_{c_N}) \right]. \quad (2.8)$$

In DLDA, eigenvalue decomposition is performed on the between-class matrix S_b , firstly. Suppose the rank of S_b is t , and let $D_b = \text{diag}(\lambda_1, \lambda_2, \dots, \lambda_t)$ be a diagonal matrix with the t largest eigenvalue on the main diagonal in descending order, $Y = [v_1, v_2, \dots, v_t]$ is the eigenvector matrix that consists of t corresponding eigenvectors. Then, dimensionality of data x is reduced by using the projection matrix $Z = YD_b^{-1/2}$ from d to t , $Z^T x$. And eigenvalue decomposition is performed on the within-class scatter matrix of the projected samples, $\widetilde{S}_w = Z^T S_w Z$. Let $D_w = \text{diag}(\eta_1, \eta_2, \dots, \eta_t)$ be the ascending order eigenvalue matrix of \widetilde{S}_w and $U = [u_1, u_2, \dots, u_t]$ be the corresponding eigenvector matrix. Therefore, the final transformation matrix is given by $W = YD_b^{-1/2}UD_w^{-1/2}$.

To address the computation complexity problem of high dimensional data, the eigenanalysis method presented by Turk and Pentland [1] is applied in DLDA, which makes the eigenanalysis of scatter matrices be progressed in an efficient way. For the eigenvalue decomposition of any symmetry matrix A with the form of $A = BB^T$, we can consider the eigenvectors v_i of $B^T B$ such that

$$B^T B v_i = \lambda_i v_i. \quad (2.9)$$

Premultiplying both sides by G , we have

$$BB^T B v_i = A B v_i = \lambda_i B v_i \quad (2.10)$$

from which it can be concluded that the eigenvectors of A is $B v_i$ with the corresponding eigenvalue λ_i .

3. Direct neighborhood discriminant analysis

Instead of mining the statistical discriminant information, manifold learning techniques try to find out the local manifold structure of data. Derived from the locality preserving idea [10, 11], graph embedding framework-based techniques extract the local discriminant features for classification. For a general pattern classification problem, it is expected to find a linear transformation, such that the compactness for the samples that belong to the same class and the separation for the samples of the interclass should be enhanced in the transformed space. As an example, a simple multiclass classification problem is illustrated in Figure 1. Suppose there are two nearest inter- and intraclass neighbors searched for classification. The inter- and intraclass nearby data points of five data points A–E is shown in Figures 1(b) and 1(c), respectively. For data point A, it is optimal that the distance from its interclass neighbors should be maximized to alleviate their bad influence for classification. On the other hand, the distance between data point A and its intraclass neighbors should be minimized to make A be classified correctly.

Based on the consideration, two graphs, that is, the between-class graph G and the within-class graph G' are constructed to discover the local discriminant structure [13, 15]. For each data point x_i , its sets of inter- and intraclass neighbors are indicated by $kNN^b(x_i)$ and

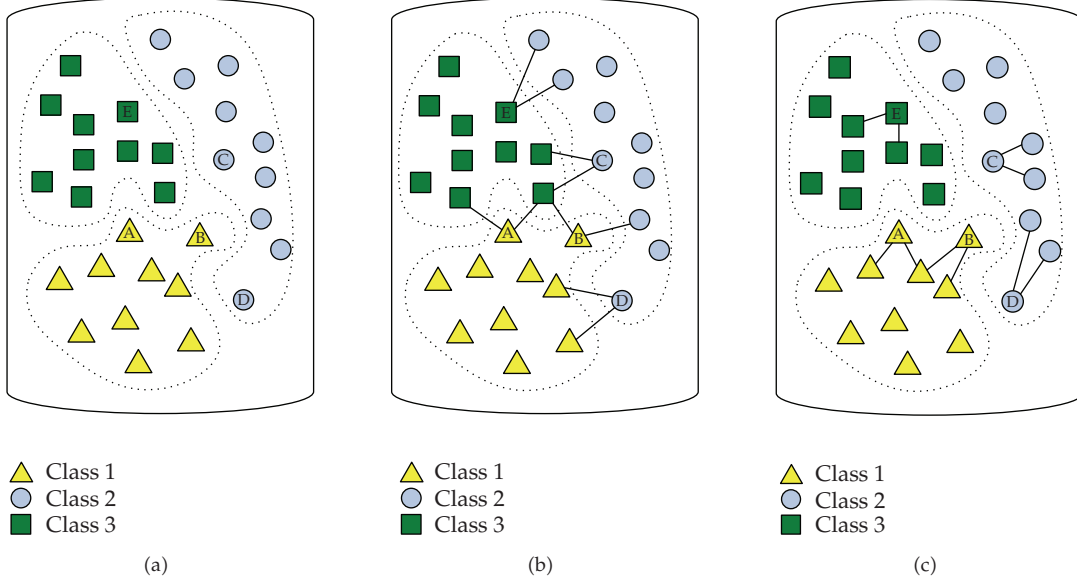


Figure 1: Local discriminant neighbors. (a) Multi-class classification (b) Two interclass neighbors (c) Two intraclass neighbors.

$kNN^w(x_i)$, respectively. Then, the weight W_{ij} reflects the weight of the edge in the between-class graph G is defined as

$$W_{ij}^b = \begin{cases} 1 & \text{if } x_i \in kNN^b(x_j) \text{ or } x_j \in kNN^b(x_i), \\ 0 & \text{else,} \end{cases} \quad (3.1)$$

and similarly define within-class affinity weight as

$$W_{ij}^w = \begin{cases} 1 & \text{if } x_i \in kNN^w(x_j) \text{ or } x_j \in kNN^w(x_i), \\ 0 & \text{else.} \end{cases} \quad (3.2)$$

Let the transformation matrix be denoted by $P \in R^{d \times d'}$ ($d' \ll d$), which transforms the original data x from high-dimensional space R^d into a low-dimensional space $R^{d'}$ by $y = P^T x$. The separability of interclass samples in the transformed low-dimensional space can be defined as

$$\begin{aligned}
 F_b &= \sum_{i,j} \|P^T x_i - P^T x_j\|^2 W_{ij}^b \\
 &= \sum_{i,j} \text{tr}[(P^T x_i - P^T x_j)(P^T x_i - P^T x_j)^T W_{ij}^b] \\
 &= \sum_{i,j} \text{tr}[P^T (x_i - x_j) W_{ij}^b (x_i - x_j)^T P] \\
 &= \text{tr} \left(2 \sum_i P^T x_i D_{ii}^b x_i^T P - 2 \sum_{i,j} P^T x_i W_{ij}^b x_j^T P \right) \\
 &= \text{tr}(P^T X (2D^b - 2W^b) X^T P),
 \end{aligned} \quad (3.3)$$

where $\text{tr}(\cdot)$ is the trace of matrix, $X = [x_1, x_2, \dots, x_N]$ is the data matrix, and D^b is a diagonal matrix, of which entries are column (or row, since W^b is symmetric) sum of W^b , $D_{ii}^b = \sum_j W_{ij}^b$. Similarly, the compactness of intraclass samples can be characterized as

$$\begin{aligned}
F_w &= \sum_{i,j} \|P^T x_i - P^T x_j\|^2 W_{ij}^w \\
&= \sum_{i,j} \text{tr}[(P^T x_i - P^T x_j)(P^T x_i - P^T x_j)^T W_{ij}^w] \\
&= \sum_{i,j} \text{tr}[P^T(x_i - x_j)W_{ij}^w(x_i - x_j)^T P] \\
&= \text{tr}\left(2\sum_i P^T x_i D_{ii}^w x_i^T P - 2\sum_{i,j} P^T x_i W_{ij}^w x_j^T P\right) \\
&= \text{tr}(P^T X(2D^w - 2W^w)X^T P).
\end{aligned} \tag{3.4}$$

Here, D^w is a diagonal matrix of which entries are column (or row) sum of W^w on the main diagonal, $D_{ii}^w = \sum_j W_{ij}^w$. Then, the optimal transformation matrix P can be obtained by solving the following problem:

$$\begin{aligned}
P^* &= \arg \max_P \frac{P^T S_s P}{P^T S_c P}, \\
S_s &= X(2D^b - 2W^b)X^T, \\
S_c &= X(2D^w - 2W^w)X^T.
\end{aligned} \tag{3.5}$$

Here, S_c is always singular with small training sample set leading problem to get projective matrix P directly, thus previous local discriminant techniques still suffer from the curse of high dimensionality. Generally, PCA is usually performed to reduce dimension as a preprocessing step in such environment [15], however, possible discriminant information may be ignored. Inspired by DLDA, we can perform eigenanalysis on S_s and S_c successively to extract the complete local geometrical structure directly without PCA preprocessing. To alleviate the burden of computation, we reformulate S_s and S_c so that Turk's eigenanalysis method can be employed. For each nonzero element of W^b , W_{ij}^b , we build an N dimensional interclass index vector $h^{(i,j)}$ of all zeroes except the i th and j th element is set to be 1 and -1 , respectively:

$$h^{(i,j)} = \left[\overbrace{0 \dots 0}^{i-1}, 1, 0 \dots 0, \overbrace{-1, 0 \dots 0}^{N-j} \right]^T. \tag{3.6}$$

Suppose there are N_b nonzero elements in W^b , let $H_s = [h_1, h_2, \dots, h_{N_b}]$ be the interclass index matrix made up of N_b interclass index vectors. It can be easily obtained that $2D^b - 2W^b = H_s H_s^T$, which we prove in Appendix A. Therefore, S_s can be reformulated as

$$S_s = X H_s H_s^T X^T. \tag{3.7}$$

Input: Data matrix $X \in R^{d \times N}$, class label L

Output: Transformed matrix P^*

1. Construct the between-class and the within-class affinity weight matrix W^b, W^w .
2. Construct the interclass and the intraclass index matrix H_s, H_c according to the nonzero elements of W^b, W^w .

For the k th nonzero element of $W^b (W^w)$, $W_{ij}^b (W_{ij}^w)$, the corresponding k th column in $H_s (H_c)$ is constructed as

$$\left[\underbrace{0 \cdots 0}_{i-1}, \underbrace{1, 0 \cdots 0}_i, \underbrace{-1, 0 \cdots 0}_j, \underbrace{0 \cdots 0}_{N-j} \right]^T.$$

3. Apply eigenvalue decomposition to S_s and keep the largest t nonzero eigenvalues $\lambda = [\lambda_1, \lambda_2, \dots, \lambda_t]$ and corresponding eigenvectors $U = [u_1, u_2, \dots, u_t]$ after sorted in decreasing order, where $t = \text{rank}(S_s)$.
4. Compute P_s as $P_s = UD_s^{-1/2}$, where $D_s = \text{diag}(\lambda_1, \lambda_2, \dots, \lambda_t)$ is diagonal matrix with λ_i on the main diagonal.
5. Perform eigenvalue decomposition on $\widetilde{S}_c = P_s^T S_c P_s$. Let $D_c = \text{diag}(\mu_1, \mu_2, \dots, \mu_n)$ be the eigenvalue matrix of \widetilde{S}_c in ascending order and $V = [v_1, v_2, \dots, v_n]$ be the corresponding eigenvector matrix. Calculate P_c as $P_c = VD_c^{-1/2}$.
6. $P^* \leftarrow P_s P_c$.

Algorithm 1: DNDA algorithm.

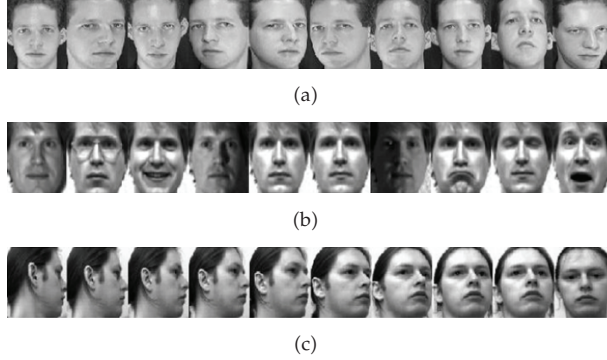


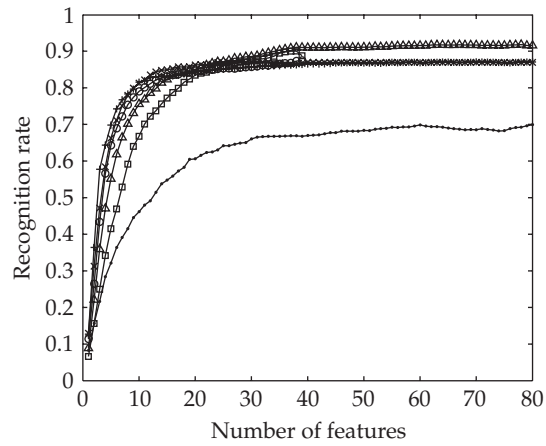
Figure 2: Sample images from ORL, Yale, and UMIST face database. (a) ORL, (b) Yale, and (c) UMIST.

As each column in H_s has only two nonzero elements 1 and -1 , we can make the first row in H_s be a null row by adding all rows but the first to the first row. On the other hand, for each column $h^{(j,i)}$ in H_s , there is another column $h^{(j,i)}$ with contrary sign. Then, it is clear that

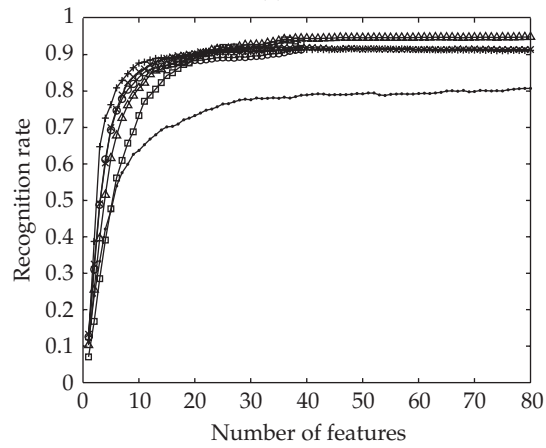
$$\text{rank}(H_s) = \min \left\{ N - 1, \frac{N_b}{2} \right\}, \quad (3.8)$$

where N_b is the number of nonzero elements in W^b . Due to the properties of matrix trace [17], we can get

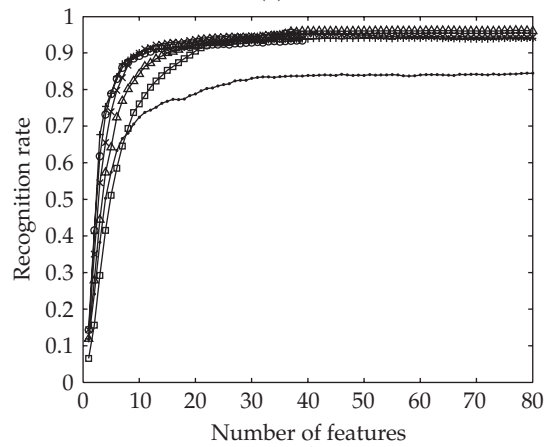
$$\text{rank}(S_s) = \text{rank}(XH_s) \leq \min \{ \text{rank}(X), \text{rank}(H_s) \}. \quad (3.9)$$



(a)



(b)



(c)

+ Eigenface + LPP
 o Fisherface x MFA
 □ DLDA △ DNDA

(c)

Figure 3: Recognition rate against the number of features used in the matching on the ORL database: (a) 3 training samples, (b) 4 training samples, and (c) 5 training samples.

Table 1: Comparison of recognition rates of Eigenface, Fisherface, DLDA, LPP, MFA, and DNDA on the ORL database.

Method	3 Training samples	4 Training samples	5 Training samples
Eigenface	86.64% (121)	91.65% (112)	94.05% (123)
Fisherface	87.46% (39)	91.42% (39)	93.35% (38)
DLDA	90.04% (38)	94.04% (39)	95.6% (37)
LPP	73.54% (91)	82.73% (98)	86.62% (99)
MFA	87.13% (27)	92% (39)	95.28% (41)
DNDA	91.07% (44)	94.69% (46)	96.12% (77)

Table 2: Comparison of recognition rates of Eigenface, Fisherface, DLDA, LPP, MFA, and DNDA on the Yale database.

Method	3 Training samples	4 Training samples	5 Training samples
Eigenfaces	76.79% (39)	80.14% (50)	82.39% (60)
Fisherfaces	80.96% (14)	84.27% (14)	90% (14)
DLDA	79.62% (12)	84.52% (11)	89.56% (14)
LPP	77.38% (44)	80.48% (59)	84.33% (59)
MFA	79.42% (26)	86.48% (23)	88.94% (24)
DNDA	82.42% (22)	88.62% (35)	90.61% (29)

In many FR cases, the number of pixels in a facial image is much larger than the number of available samples, that is, $d \gg N$. It tells us that the rank of S_s is at most $\min\{N - 1, N_b/2\}$. Similarly, S_c can also be reformulated as

$$S_c = XH_cH_c^T X^T. \quad (3.10)$$

Here, $H_c \in R^{N \times N_w}$ is the intraclass index matrix consisting of all the N_w intraclass index vectors as columns, which is constructed according to the N_w nonzero elements in W^w . Similar to S_s , the rank of S_c is up to $\min\{N - 1, N_w/2\}$. Based on the modified formulation, the optimal transformation matrix P can be obtained as

$$P^* = \arg \max_P \frac{P^T S_s P}{P^T S_c P} = \arg \max_P \frac{P^T G_s G_s^T P}{P^T G_c G_c^T P}, \quad G_s = XH_s, G_c = XH_c. \quad (3.11)$$

As the null space of S_s contributes little to classification, it is feasible to remove such subspace by projecting S_s into its range space. We apply the eigenvalue decomposition to S_s and unitize it through Turk's eigenanalysis method, while discarding those eigenvectors whose corresponding eigenvalues are zero, which do not take much power for discriminant analysis. Then, the discriminant information in S_c can be obtained by performing eigenanalysis on \tilde{S}_c , which is gotten by transforming S_c into the range subspace of S_s . This algorithm can be implemented by the pseudocode shown in Algorithm 1.

Table 3: Comparison of recognition rates of Eigenface, Fisherface, DLDA, LPP, MFA, and DNDA on the UMIST database.

	Dimensionality	Recognition rate
Eigenfaces	99	89.84%
Fisherfaces	18	93.04%
DLDA	13	93.65%
LPP	93	92.95%
MFA	72	94.82%
DNDA	48	96.01%

DNDA has a computational complexity of $o(N_b^3)$ (N_b is the number of nonzero elements in W^b), as it preserves a similar procedure to DLDA ($o(c^3)$). Compared with Eigenface ($o(N^2d)$) and Fisherface ($o(N^2d)$), DNDA is still more efficient for feature extraction in high dimensionality if $d \gg N$.

4. Experiments

In this section, we investigate the performance of the proposed DNDA method for face recognition. Three popular face databases, ORL, Yale, and UMIST are used in the experiments. To verify the performance of DNDA, each experiment is compared with classical approaches: Eigenface [1], Fisherface [2], DLDA [6], LPP [11], and MFA [15]. The three nearest-neighbor classifier with Euclidean distance metric is employed to find the image in the database with the best match.

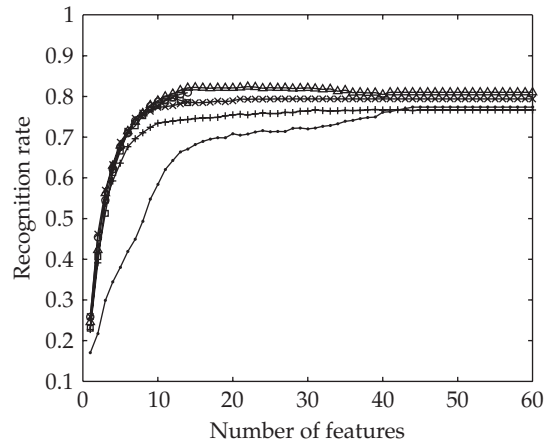
4.1. ORL database

In ORL database [18], there are 10 different images for each of 40 distinct subjects. For some subjects, the images were taken at different times, varying the lighting, facial expressions (open/closed eyes, smiling/not smiling), and facial details (glasses/no glasses). All the images are taken against a dark homogeneous background with the subjects in an upright, frontal position (with tolerance for some side movement). The original images have size of 92×112 pixels with 256 gray levels; such one subject is illustrated in Figure 2(a).

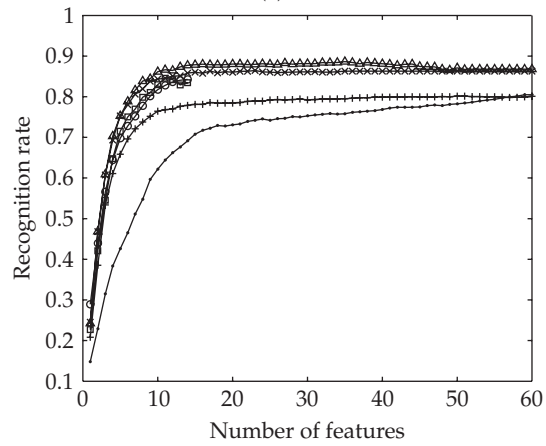
The experiments are performed with different numbers of training samples. As there are 10 images for each subject, n ($n = 3, 4, 5$) of them are randomly selected for training and the remaining are used for testing. For each n , we perform 20 times to choose randomly the training set and the average recognition rate is calculated. Figure 3 illustrates the plot of recognition rate versus the number of features used in the matching for Eigenface, Fisherface, DLDA, LPP, MFA, and DNDA. The best performance obtained by each method and the corresponding dimension of reduced space in the bracket are shown in Table 1.

4.2. Yale database

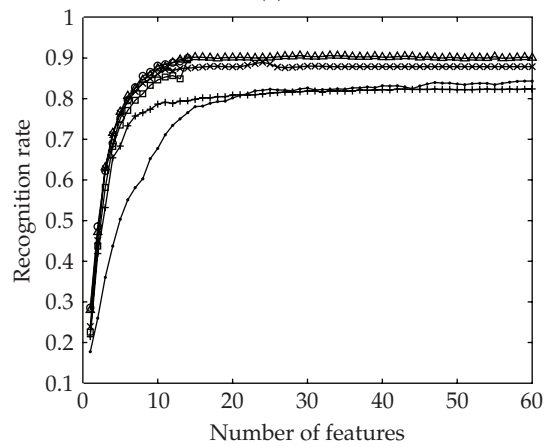
The Yale Face Database [19] contains 165 grayscale images of 15 individuals. There are 11 images per subject, one per different lighting condition (left-light, center-light, right-light), facial expression (normal, happy, sad, sleepy, surprised, wink), and with/without glasses. Each images used in the experiments is 92×112 pixels with 256 gray levels. The facial images of one individual are illustrated in Figure 2(b).



(a)



(b)



(c)

- +— Eigenface
- o— Fisherface
- DLDA
- *— LPP
- x— MFA
- △— DNDA

Figure 4: Recognition rate against the number of features used in the matching on the Yale database: (a) 3 training samples, (b) 4 training samples and (c) 5 training samples.

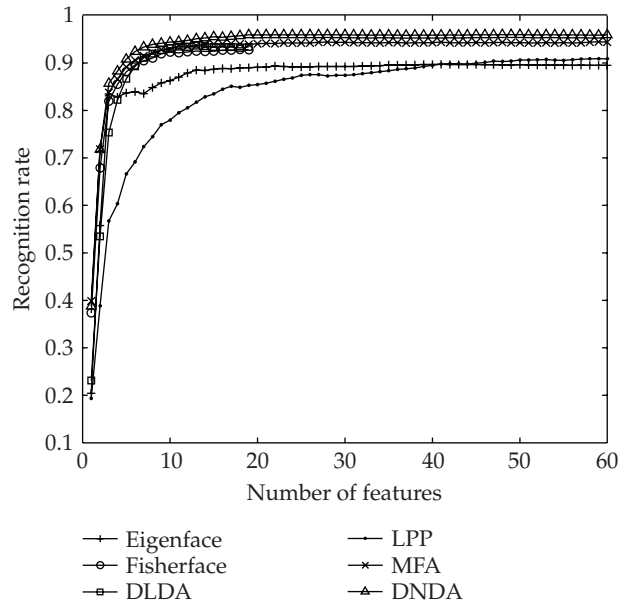


Figure 5: Recognition rate against the number of features used in the matching on the UMIST database.

The experimental implementation is the same as before. For each individual, n ($n = 3, 4, 5$) images are randomly selected for training and the rest are used for testing. For each given n , we average the results over experiments repeated 20 times independently. Figure 4 illustrates the plot of recognition rate versus the number of features used in the matching for Eigenface, Fisherface, DLDA, LPP, MFA, and DNDA. The best results obtained in the experiments and the corresponding reduced dimension for each method is shown in Table 2.

4.3. UMIST database

The UMIST face database [20] consists of 564 images of 20 people. For simplicity, the Precropped version of the UMIST database is used in this experiment, where each subject covers a range of poses from profile to frontal views and a range of race/sex/appearance. The size of cropped image is 92×112 pixels with 256 gray levels. The facial images of one subject with different views are illustrated in Figure 2(c).

For each individual, we chose 8 images of different views distributed uniformly in the range $0-90^\circ$ for training, and the rest are used for training. Figure 5 illustrates the plot of recognition rate versus the number of features used in the matching for Eigenface, Fisherface, DLDA, LPP, MFA, and DNDA. The best performance and the corresponding dimensionalities of the projected spaces for each method are shown in Table 3.

From the experiment results, it is very obvious that DNDA achieves higher accuracy than the other methods. This is probably due to the fact that DNDA is a two-stage local discriminant technique, different from LPP and MFA. Moreover, PCA is removed in DNDA preserving more discriminant information compared with others.

5. Conclusions

Inspired by DLDA, we propose in this paper a novel local discriminant feature extraction method called direct neighborhood discriminant analysis (DNDA). In order to avoid SSS

problem, DNDA performs a two-stage eigenanalysis approach, which can be implemented efficiently by using Turk's method. Compared with other methods, PCA preprocessing is left out in DNDA with the immunity from the SSS problem. Experiments on ORL, Yale, and UMIST face databases show the effectiveness and robustness of our proposed method for face recognition. To get a better classification result, the improvement and extension of DNDA are to be taken into account in our future work.

Appendix

A. Proof of $2D - 2W = HH^T$

Given the graph weight matrix W with l nonzero elements, consider two matrices $M, N \in R^{N \times l}$. For each nonzero element in W , there is corresponding column in M and N with common location, respectively. Let $Z = \{(i, j) \mid W_{ij} \neq 0\}$ be the index set of nonzero elements in W . For the k th ($1 \leq k \leq l$) nonzero element W_{ij} in W , the k th column of M, N is represented as

$$\begin{aligned} M_{(:,k)} &= \left[\overbrace{0 \cdots 0}^{i-1}, 1, 0 \cdots 0 \right]^T, \\ N_{(:,k)} &= \left[\overbrace{0 \cdots 0}^{j-1}, -1, 0 \cdots 0 \right]^T. \end{aligned} \quad (\text{A.1})$$

Then, it is easy to get

$$\begin{aligned} M_{(a,:)} M_{(b,:)}^T &= 0, \\ N_{(a,:)} N_{(b,:)}^T &= 0 \end{aligned} \quad (\text{A.2})$$

for $a \neq b$ ($1 \leq a, b \leq N$), and

$$\begin{aligned} M_{(a,:)} N_{(b,:)}^T &= 0 \quad \text{for } (a, b) \notin Z, \\ N_{(a,:)} M_{(b,:)}^T &= 0 \quad \text{for } (b, a) \notin Z, \end{aligned} \quad (\text{A.3})$$

where $M_{(k,:)}$ and $N_{(k,:)}$ denote the k th row of M and N , respectively. Therefore, we can get

$$\begin{aligned} (MM^T)_{ij} &= \sum_{k=1}^l M_{ik} M_{jk} = \delta_{ij} \sum_{k=1}^l M_{ik} = \delta_{ij} \sum_{q=1}^N W_{iq}, \\ (NN^T)_{ij} &= \sum_{k=1}^l N_{ik} N_{jk} = \delta_{ij} \sum_{k=1}^l N_{ik} = \delta_{ij} \sum_{q=1}^N W_{qi}, \\ (MN^T)_{ij} &= \sum_{k=1}^l M_{ik} N_{jk} = W_{ij}, \\ (NM^T)_{ij} &= \sum_{k=1}^l N_{ik} M_{jk} = W_{ji}, \end{aligned} \quad (\text{A.4})$$

where δ_{ij} is the Kronecker delta. Note that both matrix D and W are symmetry matrices, based on the above equations, it is easy to find out

$$\begin{aligned}(M - N)(M - N)^T &= MM^T + NN^T - MN^T - NM^T \\ &= D + D - W - W \\ &= 2D - 2W.\end{aligned}\tag{A.5}$$

It is easy to check that $H = M - N$, which completes the proof.

Acknowledgments

This work is supported in part by the Program for New Century Excellent Talents of Educational Ministry of China (NCET-06-0762) and Specialized Research Fund for the Doctoral Program of Higher Education (20060611009), and in part by Natural Science Foundations of Chongqing CSTC (CSTC2007BA2003 and CSTC2006BB2003), the National Natural Science Foundation of China under the Project Grant no. 60573125.

References

- [1] M. Turk and A. Pentland, "Eigenfaces for recognition," *Journal of Cognitive Neuroscience*, vol. 3, no. 1, pp. 71–86, 1991.
- [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, no. 7, pp. 711–720, 1997.
- [3] J. Lu, K. N. Plataniotis, and A. N. Venetsanopoulos, "Regularized discriminant analysis for the small sample size problem in face recognition," *Pattern Recognition Letters*, vol. 24, no. 16, pp. 3079–3087, 2003.
- [4] W.-S. Chen, P. C. Yuen, and J. Huang, "A new regularized linear discriminant analysis method to solve small sample size problems," *International Journal of Pattern Recognition and Artificial Intelligence*, vol. 19, no. 7, pp. 917–935, 2005.
- [5] L.-F. Chen, H.-Y. M. Liao, M.-T. Ko, J.-C. Lin, and G.-J. Yu, "New LDA-based face recognition system which can solve the small sample size problem," *Pattern Recognition*, vol. 33, no. 10, pp. 1713–1726, 2000.
- [6] H. Yu and J. Yang, "A direct LDA algorithm for high dimensional data-with application to face recognition," *Pattern Recognition*, vol. 34, no. 10, pp. 2067–2070, 2001.
- [7] J. B. Tenenbaum, V. de Silva, and J. C. Langford, "A global geometric framework for nonlinear dimensionality reduction," *Science*, vol. 290, no. 5500, pp. 2319–2323, 2000.
- [8] S. T. Roweis and L. K. Saul, "Nonlinear dimensionality reduction by locally linear embedding," *Science*, vol. 290, no. 5500, pp. 2323–2326, 2000.
- [9] M. Belkin and P. Niyogi, "Laplacian eigenmaps and spectral techniques for embedding and clustering," in *Advances in Neural Information Processing Systems 14 (NIPS '01)*, pp. 585–591, MIT Press, Cambridge, Mass, USA, 2002.
- [10] X. He and P. Niyogi, "Locality preserving projections," in *Advances in Neural Information Processing Systems 16 (NIPS '03)*, pp. 153–160, MIT Press, Cambridge, Mass, USA, 2004.
- [11] X. He, S. Yan, Y. Hu, P. Niyogi, and H.-J. Zhang, "Face recognition using Laplacian faces," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 27, no. 3, pp. 328–340, 2005.
- [12] X. He, D. Cai, S. Yan, and H.-J. Zhang, "Neighborhood preserving embedding," in *Proceedings of the 10th IEEE International Conference on Computer Vision (ICCV '05)*, vol. 2, pp. 1208–1213, Beijing, China, October 2005.
- [13] H.-T. Chen, H.-W. Chang, and T.-L. Liu, "Local discriminant embedding and its variants," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR '05)*, vol. 2, pp. 846–853, San Diego, Calif, USA, June 2005.
- [14] W. Zhang, X. Xue, H. Lu, and Y.-F. Guo, "Discriminant neighborhood embedding for classification," *Pattern Recognition*, vol. 39, no. 11, pp. 2240–2243, 2006.

- [15] 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, no. 1, pp. 40–51, 2007.
- [16] J. Ham, D. D. Lee, S. Mika, and B. Schölkopf, "A kernel view of the dimensionality reduction of manifolds," in *Proceedings of the 21th International Conference on Machine Learning (ICML '04)*, pp. 369–376, Banff, Alberta, Canada, July 2004.
- [17] G. H. Golub and C. F. van Loan, *Matrix Computations*, Johns Hopkins Studies in the Mathematical Sciences, Johns Hopkins University Press, Baltimore, Md, USA, 3rd edition, 1996.
- [18] Olivetti & Oracle Research Laboratory, The Olivetti & Oracle Research Laboratory Face Database of Faces, <http://www.cl.cam.ac.uk/research/dtg/attarchive/facedatabase.html>.
- [19] Yale Face Database, 2002, <http://cvc.yale.edu/projects/yalefaces/yalefaces.html>.
- [20] D. B. Graham and N. M. Allinson, "Characterizing virtual eigensignatures for general purpose face recognition," in *Face Recognition: From Theory to Applications*, H. Wechsler, P. J. Phillips, V. Bruce, F. Fogelman-Soulie, and T. S. Huang, Eds., vol. 163 of *NATO ASI Series F, Computer and Systems Sciences*, pp. 446–456, Springer, Berlin, Germany, 1998.