Discriminant Low-dimensional Subspace Analysis for Face Recognition with Small Number of Training Samples

Hui Kong[†], Xuchun Li[†], Jian-Gang Wang[‡], Eam Khwang Teoh[†] and Chandra Kambhamettu[§]
School of Electrical and Electronic Engineering
Nanyang Technological University[†], Singapore 639798
Institute for Infocomm Research[‡], 21 Heng Mui Keng Terrace, Singapore 119613
Department of Computer and Information Science
University of Delaware[§], Newark, DE 19716-2712
konghui@pmail.ntu.edu.sg

Abstract

In this paper, a framework of Discriminant Low-dimensional Subspace Analysis (DLSA) method is proposed to deal with the Small Sample Size (SSS) problem in face recognition area. Firstly, it is rigorously proven that the null space of the total covariance matrix, $\mathbf{S}_t$, is useless for recognition. Therefore, a framework of Fisher discriminant analysis in a low-dimensional space is developed by projecting all the samples onto the range space of $\mathbf{S}_t$. Two algorithms are proposed in this framework, i.e., Unified Linear Discriminant Analysis (ULDA) and Modified Linear Discriminant Analysis (MLDA). The ULDA extracts discriminant information from three subspaces of this low-dimensional space. The MLDA adopts a modified Fisher criterion which can avoid the singularity problem in conventional LDA. Experimental results on a large combined database have demonstrated that the proposed ULDA and MLDA can both achieve better performance than the other state-of-the-art LDA-based algorithms in recognition accuracy.

1 Introduction

Linear Discriminant Analysis [3] is a well-known scheme for feature extraction and dimension reduction. It has been used widely in many applications such as face recognition [1], image retrieval [7], etc. Classical LDA projects the data onto a lower-dimensional vector space such that the ratio of the between-class scatter to the within-class scatter is maximized, thus achieving maximum discrimination. The optimal projection (transformation) can be readily computed by solving a generalized eigenvalue problem.

However, the intrinsic limitation of classical LDA is that its objective function requires the within-class covariance matrix to be nonsingular. For many applications, such as face recognition, all scatter matrices in question can be singular since the data vectors lie in a very high-dimensional space, and in general, the feature dimension far exceeds the number of data samples. This is known as the *Small Sample Size* problem or singularity

BMVC 2005 doi:10.5244/C.19.70

problem [3]. In recent years, many approaches have been proposed to deal with this problem. Among these LDA extensions, PCA+LDA has received a lot of attention, especially for face recognition [1]. In this method, an intermediate PCA step is implemented before the LDA step and then LDA is performed in the PCA subspace. Although the scatter matrix in question can be of full-rank after the PCA step, the removed subspace may also contain some useful information, and this removal may result in a loss of discriminative information.

Chen et al. [2] suggested that the null space of $\mathbf{S}_w$ contains the most discriminative information, hence an LDA method in the null space of $\mathbf{S}_w$ was proposed, called N-LDA. However, when the number of training samples is large, the null space of $\mathbf{S}_w$ becomes small, and much discriminative information outside this null space will be lost. Huang et al. [4] introduced a more efficient null space approach by removing the null space of $\mathbf{S}_t$ and then applies Chen's original N-LDA [2] method to the reduced subspace of $\mathbf{S}_t$, but they didn't give the explicit reason why it works in the range space of the total covariance matrix. Yu and Yang [11] proposed the Direct LDA (D-LDA) algorithm which also incorporates the concept of null space.

Wang and Tang [9] presented a random sampling LDA for face recognition with small number of training samples. This paper concludes that both Fisherface and N-LDA encounter respective over-fitting problem for different reasons. A random subspace method and a random bagging approach are proposed to solve them. A fusion rule is adopted to combine these random sampling based classifiers. A dual-space LDA approach [8] for face recognition was proposed to simultaneously apply discriminant analysis in the principal and null subspaces of the within-class covariance matrix. The two sets of discriminative features are then combined for recognition.

Recently, a straightforward strategy was proposed for face recognition and representation, i.e., Two-Dimensional Principal Component Analysis (2D-PCA) [10]. Different from conventional PCA where data are represented as vectors, 2D-PCA adopts the matrix-based data representation model. That is, the image matrix does not need to be transformed into a vector beforehand. Instead, the covariance matrix is evaluated directly using the 2D image matrices. Inspired by 2D-PCA, discriminant analysis with the matrix-based data representation model [5], i.e., Two-Dimensional Fisher Discriminant Analysis (2D-FDA), has achieved more promising results than conventional LDA-based methods. In contrast to the $\mathbf{S}_b$ and $\mathbf{S}_w$ of conventional LDA, the covariance matrices obtained by 2D-FDA are generally not singular.

To overcome the *Small Sample Size* problem in LDA, a framework of Discriminant Low-dimensional Subspace Analysis is proposed in this paper for face recognition with each subject having small number of training images. In this framework, it is mathematically proven that the null space of the total covariance matrix, $\mathbf{S}_t$, is useless for recognition. Therefore, a low-dimensional feature space, $\mathscr{Q}$, is formed by discarding the null space of $\mathbf{S}_t$ and then projecting all the image samples onto the range space of $\mathbf{S}_t$. Two algorithms are proposed in this low-dimensional space, i.e., Unified Linear Discriminant Analysis (ULDA) and Modified Linear Discriminant Analysis (MLDA). The ULDA extracts discriminant information from three subspaces of $\mathscr{Q}$ respectively, and then a fusion scheme is given to combine the three kinds of extracted discriminant features for face recognition. The MLDA adopts a modified Fisher criterion which takes a *Deduction* form rather than *Quotient* form. Therefore, MLDA can avoid the singularity problem in conventional LDA.

The rest of the paper is organized as follows: Section 2 gives the theoretical foundation of the DLSA method. Section 3 describes the details of the ULDA. Section 4 introduces the MLDA algorithm. Experiments and discussions are presented in Section 5. We draw the conclusion in Section 6.

2 Unified Linear Discriminant Analysis

The conventional two-stage LDA-based algorithms, such as Fisherface [1], Null-space based LDA [2, 4], Direct-LDA [11], etc., all lose some useful information to ensure the nonsingularity of the within-class covariance matrix by discarding part of the discriminant subspaces. In this part, our task is to unify the specific discriminant analysis methods by splitting subspace and fusing extracted features from them.

It is well known that the optimal projection (transformation), $\mathbf{W} = [\mathbf{w}_1, \mathbf{w}_2, ..., \mathbf{w}_{L-1}]$, in LDA can be readily computed by solving a generalized eigenvalue problem, where $\mathbf{w}_i, i = 1, ..., L-1$ satisfy the following conventional Fisher criterion,

$$\mathbf{w} = \arg\max \frac{\mathbf{w}^T \mathbf{S}_b \mathbf{w}}{\mathbf{w}^T \mathbf{S}_w \mathbf{w}} \tag{1}$$

If $\mathbf{S}_w$ is invertible, $\mathbf{w}^T \mathbf{S}_w \mathbf{w}$ is always positive for every nonzero $\mathbf{w}$ since $\mathbf{S}_w$ is positive definite. In such a case, Eq.1 can be used directly to extract a set of optimal discriminant projection vectors. However, it is almost impossible in real-world application, such as face recognition, that $\mathbf{S}_w$ is of full rank. Therefore, there always exist vectors (these vectors are from $Null(\mathbf{S}_w)$) making $\mathbf{w}^T \mathbf{S}_w \mathbf{w}$ be zero. These vectors turn out to be very effective for classification if they satisfy $\mathbf{w}^T \mathbf{S}_b \mathbf{w} > 0$ meantime. This is the idea of Null-space based LDA methods [2], [4] where they adopt the modified criterion as follows,

$$\mathbf{w} = \arg\max_{\mathbf{w}^T \mathbf{S}_w \mathbf{w} = 0} \mathbf{w}^T \mathbf{S}_b \mathbf{w} \tag{2}$$

Before the detailed description of ULDA, we provide some theoretical foundation to it.

Theorem 1: The null space of $\mathbf{S}_t$ is the common null space of both $\mathbf{S}_b$ and $\mathbf{S}_w$.

Proof: Let $\mathbf{W}_{null}$ be the null space of $\mathbf{S}_t$, that is,

$$\mathbf{W}_{null}^T \mathbf{S}_t \mathbf{W}_{null} = 0$$

It is trivial to check that $\mathbf{S}_t = \mathbf{S}_b + \mathbf{S}_w$. Therefore, we have,

$$\mathbf{W}_{null}^T (\mathbf{S}_b + \mathbf{S}_w) \mathbf{W}_{null} = 0$$

Since both $\mathbf{S}_b$ and $\mathbf{S}_w$ are positive semi-definite, we have,

$$\mathbf{W}_{null}^T \mathbf{S}_b \mathbf{W}_{null} = 0 \quad and \quad \mathbf{W}_{null}^T \mathbf{S}_w \mathbf{W}_{null} = 0$$

Therefore, we can conclude that the null space of $\mathbf{S}_t$ is the common null space of $\mathbf{S}_b$ and $\mathbf{S}_w$. □

Let $\mathscr{F}$ be the vector space where all the face-image vectors lie. Since $\mathbf{S}_t$ is symmetric and positive semi-definite, its eigenvectors that correspond to the nonzero eigenvalues forms a set of orthonormal basis for $\mathscr{F}$. Generally, $Rank(\mathbf{S}_t) = n-1$. Suppose

$\mathbf{u}_1, \mathbf{u}_2, ..., \mathbf{u}_{n-1}$ are the set of eigenvectors, then $Range(\mathbf{S}_t)$=span$\{\mathbf{u}_1, \mathbf{u}_2, ..., \mathbf{u}_{n-1}\}$ and $\mathscr{F} = Range(\mathbf{S}_t) \bigoplus Null(\mathbf{S}_t)$. For any vector $\mathbf{z} \in \mathscr{F}$, $\mathbf{z}$ can be uniquely represented by $\mathbf{z} = \mathbf{g} + \mathbf{h}$ with $\mathbf{g} \in Range(\mathbf{S}_t)$ and $\mathbf{h} \in Null(\mathbf{S}_t)$.

Lemma 1 [6]: If matrix $\mathbf{A}$ is positive, $\mathbf{x}^T\mathbf{A}\mathbf{x} = 0$ if and only if $\mathbf{A}\mathbf{x} = 0$.

Proof: It is trivial to see that if $\mathbf{A}\mathbf{x} = 0$, $\mathbf{x}^T\mathbf{A}\mathbf{x} = 0$. Therefore, what we need is only to prove that $\mathbf{x}^T\mathbf{A}\mathbf{x} = 0$ will result in $\mathbf{A}\mathbf{x} = 0$. Since $\mathbf{A}$ is positive, it must have a positive square root $\mathbf{R}$ such that $\mathbf{A} = \mathbf{R}^2$. Thus, $\langle \mathbf{R}\mathbf{x}, \mathbf{R}\mathbf{x} \rangle$=$\langle \mathbf{A}\mathbf{x}, \mathbf{x} \rangle$=$\mathbf{x}^T\mathbf{A}\mathbf{x} = 0$. Therefore, $\mathbf{R}\mathbf{x} = 0$. We can further derive that $\mathbf{A}\mathbf{x} = \mathbf{R}(\mathbf{R}\mathbf{x}) = 0$. □

Theorem 2: The null space of $\mathbf{S}_t$, $Null(\mathbf{S}_t)$, is useless for recognition.

Proof: For any vector $\mathbf{h} \in Null(\mathbf{S}_t)$, we have $\mathbf{h}^T\mathbf{S}_t\mathbf{h} = 0$. From Theorem 1, we have $\mathbf{h}^T\mathbf{S}_b\mathbf{h} = 0$. Since $\mathbf{S}_b$ is positive, according to Lemma 1, we have $\mathbf{S}_b\mathbf{h} = 0$. Hence, for any vector $\mathbf{z} \in \mathscr{F} = \mathbf{g} + \mathbf{h}$, where $\mathbf{g} \in Range(\mathbf{S}_t)$ and $\mathbf{h} \in Null(\mathbf{S}_t)$,

$$\begin{aligned} \mathbf{z}^T\mathbf{S}_b\mathbf{z} &= (\mathbf{g}+\mathbf{h})^T\mathbf{S}_b(\mathbf{g}+\mathbf{h}) \\ &= \mathbf{g}^T\mathbf{S}_b\mathbf{g} + 2\mathbf{g}^T\mathbf{S}_b\mathbf{h} + \mathbf{h}^T\mathbf{S}_b\mathbf{h} \\ &= \mathbf{g}^T\mathbf{S}_b\mathbf{g} \end{aligned}$$

Similarly, we also have

$$\begin{aligned} \mathbf{z}^T\mathbf{S}_w\mathbf{z} &= (\mathbf{g}+\mathbf{h})^T\mathbf{S}_w(\mathbf{g}+\mathbf{h}) \\ &= \mathbf{g}^T\mathbf{S}_w\mathbf{g} + 2\mathbf{g}^T\mathbf{S}_w\mathbf{h} + \mathbf{h}^T\mathbf{S}_w\mathbf{h} \\ &= \mathbf{g}^T\mathbf{S}_w\mathbf{g} \end{aligned}$$

Therefore, the conventional Fisher criterion in Eq.1 and modified criterion of Null-space based LDA in Eq.2 will be converted into

$$\begin{aligned} \mathbf{g} &= \arg \max_{\mathbf{g} \in Range(\mathbf{S}_t)} \frac{(\mathbf{g}+\mathbf{h})^T\mathbf{S}_b(\mathbf{g}+\mathbf{h})}{(\mathbf{g}+\mathbf{h})^T\mathbf{S}_w(\mathbf{g}+\mathbf{h})} \\ &= \arg \max_{\mathbf{g} \in Range(\mathbf{S}_t)} \frac{\mathbf{g}^T\mathbf{S}_b\mathbf{g}}{\mathbf{g}^T\mathbf{S}_w\mathbf{g}} \end{aligned}$$

and

$$\begin{aligned} \mathbf{g} &= \arg \max_{(\mathbf{g}+\mathbf{h})^T\mathbf{S}_w(\mathbf{g}+\mathbf{h})=0} (\mathbf{g}+\mathbf{h})^T\mathbf{S}_b(\mathbf{g}+\mathbf{h}) \\ &= \arg \max_{\mathbf{g}^T\mathbf{S}_w\mathbf{g}=0} \mathbf{g}^T\mathbf{S}_b\mathbf{g} \end{aligned}$$

Therefore, we can draw the conclusion that the discriminant projection vectors can be extracted only from $Range(\mathbf{S}_t) = span\{\mathbf{u}_1, \mathbf{u}_2, ..., \mathbf{u}_{n-1}\}$, i.e., $Null(\mathbf{S}_t)$ is not useful for recognition and can be safely discarded without losing any discriminant information. □

Based on Theorem 2, the face-image vectors are all projected onto a low-dimensional (n-1) space, called $\mathscr{Q}$, determined by the eigenvectors, $\mathbf{U} = [\mathbf{u}_1, \mathbf{u}_2, ..., \mathbf{u}_{n-1}]$, of $\mathbf{S}_t$ corresponding to those nonzero eigenvalues. Therefore, the between-class and within-class covariance matrices are transformed into $\widehat{\mathbf{S}}_b = \mathbf{U}^T\mathbf{S}_b\mathbf{U}$ and $\widehat{\mathbf{S}}_w = \mathbf{U}^T\mathbf{S}_w\mathbf{U}$ respectively in $\mathscr{Q}$. The transformed total covariance matrix $\widehat{\mathbf{S}}_t = \widehat{\mathbf{S}}_b + \widehat{\mathbf{S}}_w$. It is easy to check that $Rank(\widehat{\mathbf{S}}_t) = n-1$, $Rank(\widehat{\mathbf{S}}_b) = L-1$ and $Rank(\widehat{\mathbf{S}}_w) = n-L$. The standard LDA method

Table 1: ULDA for discriminant projection-matrix extraction from three subspaces

Algorithm 1: The ULDA Algorithm

Input: All gallery data $\mathbf{A}$

Output: Discriminant projection matrix $\mathbf{W}_{opt}^{LDA}$, $\mathbf{W}_{opt}^{NLDA}$ and $\mathbf{W}_{opt}^{DLDA}$

1. Compute the mean of each class, $\mathbf{m}_i$, and the mean of all the classes, $\mathbf{m}$. Construct $\mathbf{BM} = [(\mathbf{m}_1 - \mathbf{m}), (\mathbf{m}_2 - \mathbf{m}), ..., (\mathbf{m}_L - \mathbf{m})]$, let $\tilde{\mathbf{S}}_b = \mathbf{BM}^T\mathbf{BM}$
2. Construct $\mathbf{WM} = [(\mathbf{x}_1^1 - \mathbf{m}_1), (\mathbf{x}_1^2 - \mathbf{m}_1), ..., (\mathbf{x}_L^{n_L} - \mathbf{m}_L)]$. Let $\tilde{\mathbf{S}}_w = \mathbf{WM}^T\mathbf{WM}$. Similarly, construct $\mathbf{TM} = [(\mathbf{x}_1^1 - \mathbf{m}), (\mathbf{x}_1^2 - \mathbf{m}), ..., (\mathbf{x}_L^{n_L} - \mathbf{m})]$. Let $\tilde{\mathbf{S}}_t = \mathbf{TM}^T\mathbf{TM}$
3. Compute the eigenvectors of $\mathbf{S}_t$ corresponding to the top (n-1) eigenvalues as $\mathbf{U} = [\mathbf{u}_1, \mathbf{u}_2, ..., \mathbf{u}_{n-1}]$ using the trick proposed in Eigenface method
4. Compute the $\widehat{\mathbf{S}}_b = \mathbf{U}^T\mathbf{BMBM}^T\mathbf{U}$ and $\widehat{\mathbf{S}}_w = \mathbf{U}^T\mathbf{WMWM}^T\mathbf{U}$
5. Perform standard LDA in $Range(\widehat{\mathbf{S}}_w)$ and get the $\mathbf{W}^{LDA} = [\mathbf{w}_1^{LDA}, \mathbf{w}_2^{LDA}, ..., \mathbf{w}_{L-1}^{LDA}]$
6. Perform N-LDA in $Null(\widehat{\mathbf{S}}_w)$ and get $\mathbf{W}^{NLDA} = [\mathbf{w}_1^{NLDA}, \mathbf{w}_2^{NLDA}, ..., \mathbf{w}_{L-1}^{NLDA}]$
7. Perform D-LDA in $Range(\widehat{\mathbf{S}}_b)$ and get $\mathbf{W}^{DLDA} = [\mathbf{w}_1^{DLDA}, \mathbf{w}_2^{DLDA}, ..., \mathbf{w}_{L-1}^{DLDA}]$
8. Construct $\mathbf{W}_{opt}^{LDA}$, $\mathbf{W}_{opt}^{NLDA}$ and $\mathbf{W}_{opt}^{DLDA}$
   8.1 $\mathbf{W}_{opt}^{LDA} = \mathbf{U} * \mathbf{V}_{Range} * \mathbf{W}^{LDA}$
   8.2 $\mathbf{W}_{opt}^{NLDA} = \mathbf{U} * \mathbf{V}_{Null} * \mathbf{W}^{NLDA}$
   8.3 $\mathbf{W}_{opt}^{DLDA} = \mathbf{U} * \mathbf{P}_{Range} * \mathbf{W}^{DLDA}$

remains inapplicable since $\widehat{\mathbf{S}}_w$ is still singular in $\mathscr{Q}$. To extract the discriminant information from $\mathscr{Q}$ as complete as possible, we use a unified discriminant subspace analysis method in a divide and conquer way. Our strategy is to split the $\mathscr{Q}$ space first, then a specific algorithm will be adopted in the individually split subspace of $\mathscr{Q}$ to deal with different situations. We split the $\mathscr{Q}$ space in two ways. The first way is to split $\mathscr{Q}$ into two subspaces: $Null(\widehat{\mathbf{S}}_w)$ and $Range(\widehat{\mathbf{S}}_w)$. The other way is to split $\mathscr{Q}$ into another two subspaces: $Null(\widehat{\mathbf{S}}_b)$ and $Range(\widehat{\mathbf{S}}_b)$. Therefore, we have four subspaces available for feature extraction. However, $Null(\widehat{\mathbf{S}}_b)$ is not used considering its little contribution to the discriminant feature extraction. Then we use three different LDA methods for the three subspaces, i.e., the standard LDA is used in the $Range(\widehat{\mathbf{S}}_w)$, Null-space based LDA is used in the $Null(\widehat{\mathbf{S}}_w)$ and Direct-LDA is used in the $Range(\widehat{\mathbf{S}}_b)$. Generally, we have the following steps for the ULDA algorithm. They are listed in Table 1.

After obtaining the $\mathbf{W}_{opt}^{LDA}$, $\mathbf{W}_{opt}^{NLDA}$ and $\mathbf{W}_{opt}^{DLDA}$, we can extract three $(L-1)$ dimensional discriminant feature vectors for any given sample $\mathbf{x}$ via three linear transformations, $\mathbf{y}^{LDA} = (\mathbf{W}_{opt}^{LDA})^T\mathbf{x}$, $\mathbf{y}^{NLDA} = (\mathbf{W}_{opt}^{NLDA})^T\mathbf{x}$ and $\mathbf{y}^{DLDA} = (\mathbf{W}_{opt}^{DLDA})^T\mathbf{x}$. Therefore, it is possible to fuse them in a decision level. Here, we propose a fusion strategy and use it for classification in this paper.

Suppose the distance between any two data, $\mathbf{x}_i$ and $\mathbf{x}_j$, is given by the Euclidean distance, i.e., $d(\mathbf{x}_i, \mathbf{x}_j) = \|\mathbf{x}_i - \mathbf{x}_j\|_2$. Let us denote a pattern $\mathbf{y} = [\mathbf{y}^{LDA}, \mathbf{y}^{NLDA}, \mathbf{y}^{DLDA}]$. The summed normalized-distance between $\mathbf{y}$ and those of the training samples $\mathbf{y}_i = [\mathbf{y}_i^{LDA}, \mathbf{y}_i^{NLDA}, \mathbf{y}_i^{DLDA}], i = 1, 2, ..., n$, is given by,

$$d(\mathbf{y}, \mathbf{y}_i) = d^{LDA} + \alpha * d^{NLDA} + \beta * d^{DLDA} \tag{3}$$

where α and β are used to adjust the relative importance of each distance, $d^{LDA} =$

Table 2: Modified Linear Discriminant Analysis

Algorithm 2: The MLDA Algorithm
Input: All gallery data $\mathbf{A}$
Output: Discriminant vectors $\{\mathbf{w}_i\}$'s of MLDA
1. Compute the eigenvectors of $\mathbf{S}_t$ corresponding to the top (n-1) eigenvalues as $\mathbf{U} = [\mathbf{u}_1, \mathbf{u}_2, ..., \mathbf{u}_{n-1}]$ using the method in Algorithm 1.
2. Compute the $\widehat{\mathbf{S}}_b = \mathbf{U}^T \mathbf{S}_b \mathbf{U}$ and $\widehat{\mathbf{S}}_w = \mathbf{U}^T \mathbf{S}_w \mathbf{U}$
3. Solve the eigen-equation: $(\widehat{\mathbf{S}}_b - \mu \widehat{\mathbf{S}}_w)\widehat{\mathbf{w}}_i = \lambda_i \widehat{\mathbf{w}}_i$. Let $\widehat{\mathbf{W}} = [\widehat{\mathbf{w}}_1, \widehat{\mathbf{w}}_2, ..., \widehat{\mathbf{w}}_l]$, where l generally equals $(L-1)$
4. $\mathbf{W} = [\mathbf{w}_1, \mathbf{w}_2, ..., \mathbf{w}_{L-1}] \longleftarrow \mathbf{U}\widehat{\mathbf{W}}$

$\frac{\|\mathbf{y}^{LDA} - \mathbf{y}_i^{LDA}\|}{\sum_{j=1}^n \|\mathbf{y}^{LDA} - \mathbf{y}_j^{LDA}\|}$, $d^{NLDA} = \frac{\|\mathbf{y}^{NLDA} - \mathbf{y}_i^{NLDA}\|}{\sum_{j=1}^n \|\mathbf{y}^{NLDA} - \mathbf{y}_j^{NLDA}\|}$ and $d^{DLDA} = \frac{\|\mathbf{y}^{DLDA} - \mathbf{y}_i^{DLDA}\|}{\sum_{j=1}^n \|\mathbf{y}^{DLDA} - \mathbf{y}_j^{DLDA}\|}$.

In classification, the nearest-neighborhood classifier is used.

3 Modified Linear Discriminant Analysis

In this section, a Modified Linear Discriminant Analysis (MLDA) algorithm is proposed. In the conventional LDA, the Fisher optimization criterion is in a *Quotient* form. From the analysis mentioned above, the *Quotient* form will induce the numerical problem when there only exist small number of gallery images for each subject in the databases. To overcome this problem and meanwhile to be consistent with the Fisher rule (maximizing the between-class scatter and minimizing the within-class scatter), a modified Fisher optimization criterion in the *deduction* form is proposed. That is,

$$\mathbf{W}_{fld} = \arg\max_{\mathbf{W}} tr\{\mathbf{W}^T(\mathbf{S}_b - \mu \mathbf{S}_w)\mathbf{W}\} \tag{4}$$

where $tr(\bullet)$ is the trace operation. It is trivial to check that this modified criterion is consistent with the Fisher rule.

To get the $\mathbf{W}_{fld}$, we only need to get the solutions to the following eigenvalue problem,

$$(\mathbf{S}_b - \mu \mathbf{S}_w)\mathbf{w}_i = \lambda_i \mathbf{w}_i \tag{5}$$

However, because of the high-dimensionality of image vectors, it is difficult to directly solve Eq.5 where $\mathbf{S}_b$ and $\mathbf{S}_w$ are pretty large matrix. Fortunately, we have analyzed in Theorem 2 that the null space of $\mathbf{S}_t$ is useless for recognition, therefore, the $Null(\mathbf{S}_t)$ is discarded beforehand. The proposed MLDA is listed as Algorithm 2 in Table 2.

Once given the projection matrix $\mathbf{W}_{fld}$, the projection of any data, $\mathbf{x}$, onto this subspace is given by $\mathbf{y} = \mathbf{W}_{fld}^T \mathbf{x}$, the nearest-neighborhood classifier is used for classification.

4 Experimental results

To test the proposed algorithms in the situation where there are a large number of subjects in the database and very few training samples are available for each subject, meanwhile,

Figure 1: Sample images in the combined database

there are pose, illumination and expression variations for the subjects, we construct a large database by combining several existing public databases, i.e., ORL, Yale, YaleB, CMU PIE, UMIST, CMU AMP Expression, and XM2VTS databases.

In the combined database, we have 10 images for each subject. The database is combined in the following way. All the images in ORL database are used (400 images for 40 subjects). 150 images of Yale database are used (15 subjects). Altogether 100 images of YaleB database for 10 subjects are used (10 mild illumination conditions under the same frontal pose). Altogether 450 images of CMU PIE database for 45 subjects are used (10 mild illumination conditions under the same frontal pose). All the images from YaleB and PIE databases are preprocessed and normalized by translation, rotation, and scaling such that the two outer eye corners are in fixed positions. The UMIST face database consists of 564 images of 20 people with large pose variations. In our experiment, 200 images with each subject having 10 samples are used to ensure that face appearance changes from profile to frontal orientation with a step of about 10° separation. In the CMU face expression database, the images are collected only with expression variations. There are 13 subjects with each one has 75 images showing expression variations. In our experiments, 130 images are used for 13 subjects with each subject has 10 different expressions. We use the 3D VRML face model to render 10 images for 30 subjects in XM2VTS database, therefore, 300 images are used for XM2VTS database. Some sample images are shown in Fig.1.

Therefore, we have 1730 images for altogether 173 subjects with each person having 10 images in the large combined database. All the images are normalized into the same size of 92×112. To check the performance of the two proposed algorithms, we compare our algorithms with the state-of-the-art methods, e.g., N-LDA, D-LDA, 2D-FDA, 2D-PCA, Fisherface, etc. The comparison results are shown in Fig.2.

4.1 Random Grouping of Training and Testing Sets

To test the recognition performance with different training numbers, k ($2 \leq k \leq 5$) images of each subject are randomly selected for training and the remaining *10-k* images of each subject for testing. For each number k, 50 times of random selections are performed the combined database. The final recognition rate is the average of all.

4.2 Comparison and Discussion

In Fisherface, N-LDA and D-LDA, the dimension of the feature vector for classification is all reserved to $L-1$. For 2D-PCA and 2D-FDA, the results shown in Fig.2 are both from their optimal one, the reserved feature matrix is about 112×3 or 112×4. For the pro-

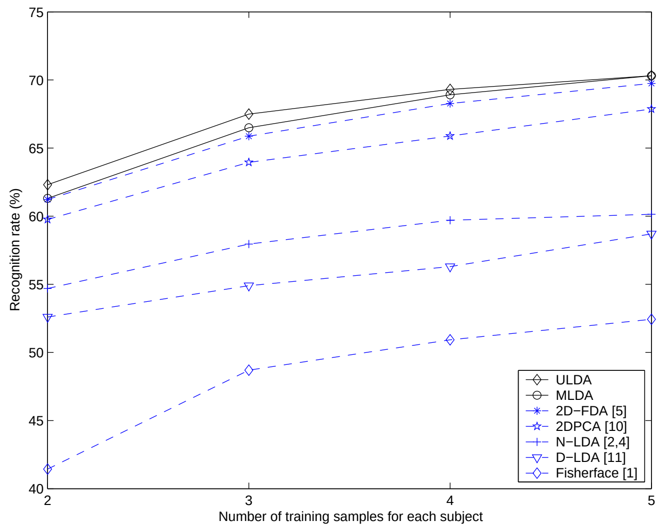


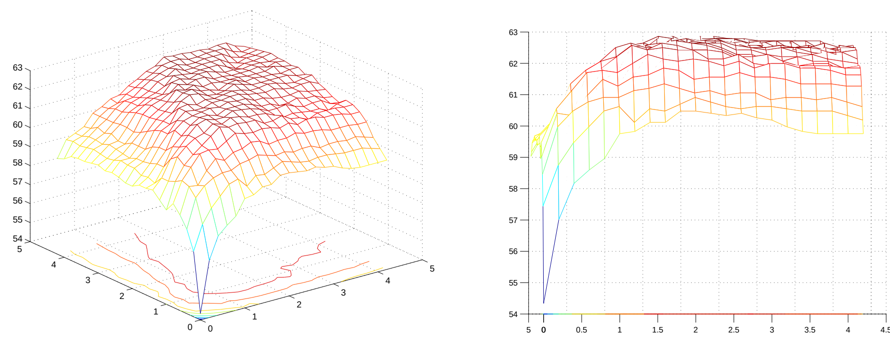
Figure 2: Experimental results on the combined database

posed algorithms in this paper, the dimensions reserved for classification are both $L-1$. From Fig.2, we may find that 2D-PCA can achieve a much better performance than Fisherface, N-LDA and D-LDA. The underlying reason is that the images in the combined database only contain mild illumination and pose variation, 2D-PCA can achieve a good generalization. Meantime, the number of training samples is small, the performance of the LDA-based algorithm (Fisherface, N-LDA and D-LDA) often degrades sharply (making the within-class covariance matrix nonsingular at the price of losing part of useful subspaces). Therefore, their performance is often inferior to that of the 2D-PCA. In addition, consistent with the experimental results of the previous papers, we do find that N-LDA and D-LDA is better than Fisherface method in recognition accuracy. N-LDA is superior to D-LDA. 2D-FDA can have better performance than 2D-PCA since the SSS problem does not exist anymore in both of them, and 2D-PCA is just good at image representation rather than discrimination.

The proposed MLDA can bypass the SSS problem by avoiding the direct use of the inverse of the within-class covariance matrix, and it is also consistent with the major idea of Fisher criterion. For this reason, it has better discriminant ability, especially when there are only few training samples for each subject, over 2D-FDA, 2D-PCA, Fisherface, N-LDA and D-LDA. The proposed ULDA takes full advantage of all the discriminant information from the possible useful subspace. Therefore it can achieve higher recognition rate than the other LDA-based algorithms. In addition, ULDA can both achieve better recognition performance than 2D-FDA, and ULDA is better than MLDA.

4.3 The Effect of α and β on Recognition Rate

In ULDA, the performance will be different if α and β take different values. To show the effect of different α and β on recognition accuracy, we use one specific example to



(a) (b)

Figure 3: The effect of α and β on recognition performance

illustrate it where only two images are available (the 2-th and 6-th images of each subject) for training. We change the values of α and β both from $0 \sim 4.2$ with a step of 0.2, Fig.3 shows the experimental results, where (a) is viewed from above and (b) is the side view.

4.4 The Effect of μ on Recognition Rate

In MLDA, the performance will be different if μ takes different values. To show the effect of different μ on recognition accuracy, a coarse adjustment of μ is followed by a fine adjustment. The number of training samples for each subject is 2, 3 and 4 for this experiment. For the coarse adjustment, the range is from 0 to 500 with a step of 5. For the fine adjustment, the range is from 0 to 20 with a step of 0.2, Fig.4 shows the experimental results, where the results in the first row is for the coarse adjustment and the second row denotes the fine adjustment.

5 Conclusions

A framework of low-dimensional Fisher discriminant analysis is developed in two different ways by discarding the null space of $\mathbf{S}_t$. Two different algorithms are proposed therein, i.e., Unified Linear Discriminant Analysis (ULDA) and Modified Linear Discriminant Analysis (MLDA). Experimental results on a large combined face database have demonstrated that the proposed two linear schemes in this framework can both achieve better performance than the state-of-the-art LDA-based algorithms in recognition accuracy.

References

[1] P.N. Belhumeur, J. Hespanha, and D.J. Kriegman. Eigenfaces vs. fisherfaces: Recognition using class specific linear projection. *IEEE Trans. on PAMI*, 19(7):711–720, 1997.

[2] L. Chen, H. Liao, M. Ko, J. Lin, and G. Yu. A new lda-based face recognition system which can solve the samll sample size problem. *Pattern Recognition*, 2000.

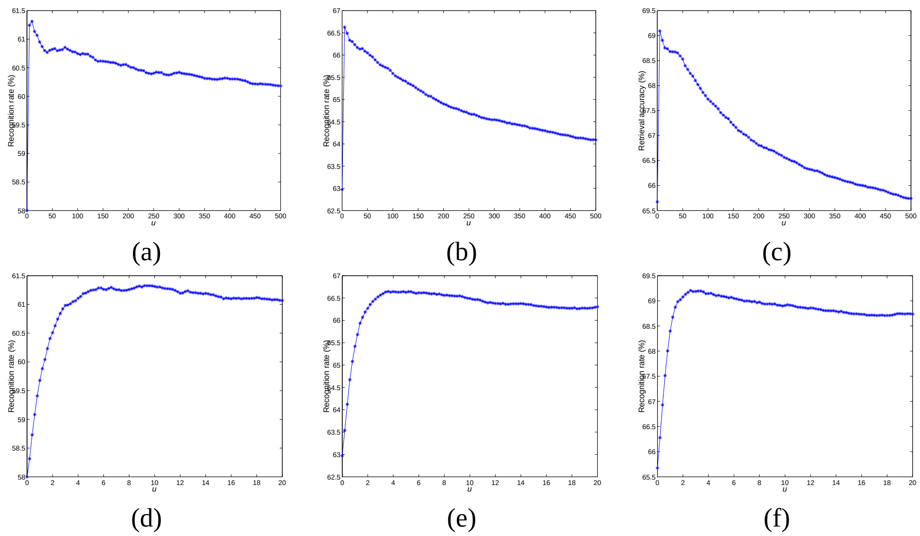


Figure 4: Recognition performance of MLDA with different μ value. (a)-(c): Coarse adjustment of μ with two, three and four training samples respectively for each subject; (d)-(f): Fine adjustment of μ with two, three and four training samples respectively for each subject.

[3] K. Fukunnaga. *Introduction to Statistical Pattern Recognition*. Academic Press, 1991.

[4] R. Huang, Q.S. Liu, H.Q. Lu, and S.D. Ma. Solving the small sample size problem of lda. In *International Conference on Pattern Recognition*, 2002.

[5] H. Kong, L. Wang, E.K. Teoh, J.G. Wang, and R. Venkateswarlu. A framework of 2d fisher discriminant analysis: Application to face recognition with small number of training samples. In *IEEE Int. Conf. on CVPR*, 2005.

[6] C.D. Meyer. *Matrix Analysis and Applied Linear Algebra*. Society for Industrial & Applied Mathematics (SIAM), 2003.

[7] D.L. Swets and J. Weng. Using discriminant eigenfeatures for image retrieval. *IEEE Trans. on PAMI*, 18(8):831–836, 1996.

[8] X. Wang and X. Tang. Dual-space linear discriminant analysis for face recognition. In *IEEE Int. Conf. on CVPR*, 2004.

[9] X. Wang and X. Tang. Random sampling lda for face recognition. In *IEEE Int. Conf. on CVPR*, 2004.

[10] J. Yang, D. Zhang, A.F. Frangi, and J. Yang. Two-dimensional pca: A new approach to appearance-based face representation and recognition. 2004.

[11] H. Yu and J. Yang. A direct lda algorithm for high-dimensional data with application to face recognition. *Pattern Recognition*, 2001.