Generalized 2D Fisher Discriminant Analysis

Hui Kong[†], Eam Khwang Teoh[†], Jian-Gang Wang[‡] 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

To solve the *Small Sample Size* (SSS) problem, the recent linear discriminant analysis using the 2D matrix-based data representation model has demonstrated its superiority over that using the conventional vector-based data representation model in face recognition [7]. But the explicit reason why the matrix-based model is better than vectorized model has not been given until now. In this paper, a framework of Generalized 2D Fisher Discriminant Analysis (G2DFDA) is proposed. Three contributions are included in this framework: 1) the essence of these '2D' methods is analyzed and their relationships with conventional '1D' methods are given, 2) a Bilateral and 3) a Kernel-based 2D Fisher Discriminant Analysis methods are proposed. Extensive experiment results show its excellent performance.

1 Introduction

Fisher Linear Discriminant (FLD), sometimes known as Linear Discriminant Analysis (LDA), has been widely used in pattern recognition [1] and image retrieval [13] for feature extraction. The objective of FLD is to find the optimal projection which maximizes the between-class scatter and meantime minimizes the within-class scatter of the projected samples. However, the within-class covariance is always singular due to the SSS problem [3], making the direct implementation of the classical FLD impossible.

To overcome the limitation of the SSS problem in LDA, a few techniques have been proposed such as the pseudo-inverse LDA [11], two-stage LDA [1, 13], regularized LDA [10] and generalized SVD based LDA [4, 16]. Among these approaches, the two-stage LDA has received much more attention than the other LDA extensions. In this method, an intermediate Principal Component Analysis (PCA) step is implemented before the LDA step. The high dimensional data are projected to a low dimensional subspace and then LDA is performed in this space. Although the scatter matrix in question can be of full-rank after the PCA step, the removed subspace contains some useful information, and this removal will result in a loss of discriminative information. To solve the same problem in the two-stage LDA, Direct-LDA (D-LDA) [17], Null-space based LDA (N-LDA) [5] and Discriminant Common Vector based LDA (DCV) [2] have been proposed.

BMVC 2005 doi:10.5244/C.19.71

However, all the above LDA techniques adopt the vector-based data representation model. The resulting feature vectors usually have a high dimensionality. The between-class covariance matrix, $\mathbf{S}_b$, and the within-class covariance matrix, $\mathbf{S}_w$, are generally singular due to the small number of training samples. Recently, inspired by the Two-Dimensional Principal Component Analysis (2DPCA) [14] and its generalized version [6]. The Two-Dimensional Fisher Discriminant Analysis (2DFDA) [7] has been proposed and achieved more promising results than conventional LDA-based methods. The Fisher's criterion is adopt in 2DFDA to find the optimal discriminative projection axes. The $\mathbf{S}_b$ and $\mathbf{S}_w$ in 2DFDA are generally not singular.

However, there exist several points that deserve to be further investigated. Firstly, the essence of 2DFDA has not been given explicitly to explain the reason why 2DFDA is better than conventional LDA. Secondly, 2DFDA is a method of extracting the optimal discriminant directions via a unilaterally left-multiplying operation. However, it can be found that in the left-multiplying U2DFDA, the computations of $\mathbf{S}_b$ and $\mathbf{S}_w$ solely emphasize the dependency (correlation) among the columns of the image matrices while neglects that among the rows. Thirdly, 2DFDA is a linear method, which neglects the higher-order statistics among the row/column vectors of the images. It is well known that the object/face appearances lie in a nonlinear low-dimensional manifold when there exist pose or/and illumination variations [8]. 2DFDA cannot effectively model such a nonlinearity, and this prevents it from higher recognition rate. Accordingly, a framework of generalized 2DFDA (G2DFDA) is proposed in this paper to overcome the drawbacks in the standard 2DFDA. Firstly, the essence of the 2DFDA is given. Secondly, a bilateral 2DFDA is proposed to consider both the correlation among the rows/columns of the image matrices. Thirdly, with the inspiration of the current kernel subspace representations, such as Kernel PCA (KPCA) [12], Kernel 2DPCA [6], Kernel FDA (KFDA) [15] and Kernel Direct Discriminant Analysis (KDDA) [9], a Kernel-based 2DFDA (K2DFDA) is investigated.

The rest parts of this paper are as follows: section 2 reviews the original 2DFDA [7] algorithms and give the essence of 2DFDA. The Bilateral 2DFDA and Kernel-based 2DFDA are proposed in Section 3 and 4. Experimental results in face recognition and discussions are presented in section 5. Conclusions are given in the last section.

2 2D Fisher Discriminant Analysis

Let $\mathbf{W} = [\mathbf{w}_1, \mathbf{w}_2, ..., \mathbf{w}_d]$ denote an $m \times d$ matrix, where $\mathbf{w}_i$ is a column vector. The idea is to project image $\mathbf{X}$, an $m \times n$ matrix, onto $\mathbf{W}$ by $\mathbf{Y} = \mathbf{W}^T\mathbf{X}$. Thus, we obtain a $d \times n$ feature matrix $\mathbf{Y}$ for $\mathbf{X}$. The discriminatory power of $\mathbf{W}$ can be measured by

$$J(\mathbf{W}) = \frac{det(\mathbf{PS}_b)}{det(\mathbf{PS}_w)} = \frac{det(\mathbf{W}^T\mathbf{S}_b\mathbf{W})}{det(\mathbf{W}^T\mathbf{S}_w\mathbf{W})} \tag{1}$$

where $\mathbf{PS}_b$ and $\mathbf{PS}_w$ are the between- and within-class covariance of the projected samples respectively. $\mathbf{S}_b$=$\sum_{i=1}^{L} L_i(\overline{\mathbf{M}}_i - \overline{\mathbf{M}})\,(\overline{\mathbf{M}}_i - \overline{\mathbf{M}})^T$ and $\mathbf{S}_w$=$\sum_{i=1}^{L}\sum_{j=1}^{L_i}(\mathbf{X}_i^j - \overline{\mathbf{M}}_i)\,(\mathbf{X}_i^j - \overline{\mathbf{M}}_i)^T$.

The vectors in $\mathbf{W}$ that maximize Eq.1 are called the optimal discriminating projection axes. Because the covariance matrices in 2DFDA are not singular anymore [7], the solution to Eq.1 can be obtained by solving a generalized eigenvalue problem.

2.1 The Essence of 2DFDA

Since the projection is a left-multiplying unilateral operation, the 2DFDA in this way is called Left-multiplying Unilateral 2D Fisher Discriminant Analysis (LU2DFDA). We notice that the covariance matrices in the LU2DFDA appears to be physically meaningful in the matrix space rather than in the vector space. However, Theorem 1 will give another perspective to make the LU2DFDA physically meaningful even in vector space.

Theorem 1: The LU2DFDA performed on the image matrices is essentially the conventional LDA method performed on the columns of the image matrices if each column is viewed as a computational unit.

Proof : Another form of $\mathbf{S}_w$ can be written as $\mathbf{S}_w = \Phi_{Sw}\Phi_{Sw}^T$. $\Phi_{Sw} = [\phi_1^{Sw}, \phi_2^{Sw}, \cdots, \phi_L^{Sw}]$, and $\phi_i^{Sw} = [(\mathbf{X}_i^1 - \overline{\mathbf{M}}_i), \cdots, (\mathbf{X}_i^{L_i} - \overline{\mathbf{M}}_i)]$=$[((\mathbf{X}_i^1(:,1) - \overline{\mathbf{M}}_i(:,1)), \cdots, ((\mathbf{X}_i^1(:,n) - \overline{\mathbf{M}}_i(:,n)))$ $, \cdots, ((\mathbf{X}_i^{L_i}(:,1) - \overline{\mathbf{M}}_i(:,1)), \cdots, (\mathbf{X}_i^{L_i}(:,n) - \overline{\mathbf{M}}_i(:,n)))]$, where $\mathbf{X}_i^j$ is the j-th training sample in the i-th class, and $\mathbf{A}(:,i)$ is the i-th column of matrix $\mathbf{A}$. Therefore, $\mathbf{S}_w$ is constructed directly by the columns of the centered training image matrices. Similarly, $\mathbf{S}_b$ is also constructed using the columns.

Therefore, the LU2DFDA performed on the image matrices can be viewed as the conventional FLD performed on the columns of all the training samples. □

Hence, the optimal projection vectors $\mathbf{W}_{opt}$ can be obtained by directly solving the following generalized eigen-value problem.

$$\mathbf{S}_w^{-1}\mathbf{S}_b\mathbf{W}_{opt} = \Lambda\mathbf{W}_{opt} \tag{2}$$

where Λ is the diagonal matrix whose diagonal elements are eigenvalues of $\mathbf{S}_w^{-1}\mathbf{S}_b$.

Like 2DPCA [14], the nearest-neighborhood classifier method is adopted for classification. In conventional LDA based methods, the feature dimension for classification is fixed to $(C-1)$, where C is the number of classes. However, in 2DFDA, the optimal number of *Fisher feature vector*, d, is not fixed. Since the $\mathbf{S}_w$ is invertible, d can be at most equal to the image's height. However, the optimal d for classification is database-dependent, i.e., the optimal d is different for different databases. In our experiments, we will discuss the optimal dimensions for different databases.

3 Bilateral 2DFDA

The above section describes a method of extracting the optimal discriminant directions via a left-multiplying U2DFDA. What if the projection is a right-multiplying operation? That is,

$$\mathbf{Y} = \mathbf{X}\mathbf{W} \tag{3}$$

In fact, it is trivial to check that the right-multiplying U2DFDA can be converted into left-multiplying U2DFDA by transposing the image matrix. Therefore, the right-multiplying *Fisher feature matrix* $\mathbf{Y}_r = \mathbf{X}\mathbf{W}_r$ and $\mathbf{W}_r$ can be obtained using Eq.2, where $\mathbf{S}_b$=$\sum_{i=1}^{L} L_i(\overline{\mathbf{M}}_i - \overline{\mathbf{M}})^T\ (\overline{\mathbf{M}}_i - \overline{\mathbf{M}})$ and $\mathbf{S}_w$=$\sum_{i=1}^{L}\sum_{j=1}^{L_i}(\mathbf{X}_i^j - \overline{\mathbf{M}}_i)^T\ (\mathbf{X}_i^j - \overline{\mathbf{M}}_i)$.

Will the left- and right-multiplying U2DFDA achieve the same recognition rate or will they recognize the same batch of face images? Our experimental results show that sometimes they have the same recognition rate, but most of the time, their performance is different. The reason is that the calculations of $\mathbf{S}_b$ and $\mathbf{S}_w$ are different in the left- and right-multiplying U2DFDA. It can also be found that either in left-multiplying U2DFDA or right-multiplying U2DFDA, the calculations of $\mathbf{S}_b$ and $\mathbf{S}_w$ solely emphasize the dependency among the row or column vectors of the image matrix and neglects the other one. Therefore, it may lose some information which is helpful for discrimination. Considering this, a bilateral-projection scheme which is called Bilateral 2D Fisher Discriminant Analysis (B2DFDA) is proposed by combining $\mathbf{Y}_l = \mathbf{W}_l^T \mathbf{X}$ and $\mathbf{Y}_r = \mathbf{X}\mathbf{W}_r$, where $\mathbf{W}_l = [\mathbf{w}_1^l, \mathbf{w}_2^l, \cdots, \mathbf{w}_{d_l}^l]$, $\mathbf{W}_r = [\mathbf{w}_1^r, \mathbf{w}_2^r, \cdots, z\mathbf{w}_{d_r}^r]$ are the left- and right-multiplying optimal projection vectors respectively, d_l is the number of left-multiplying projection directions and it can be equal to the image's height at most. d_r is the number of right-multiplying projection directions and it can be equal to the image's width at most.

After performing the left- and right-multiplying U2DFDA, $\mathbf{Y}_l$ and $\mathbf{Y}_r$ are obtained for each image. They are combined together for recognition. The steps for recognition is as follows: firstly $\mathbf{Y}_l$ and $\mathbf{Y}_r$ are transformed into 1D vectors for each images, then PCA is applied onto these vectors. Finally, two shorter vectors can be obtained for each image and they are combined into one vector for classification.

4 Kernel-based 2D Fisher Discriminant Analysis

Without losing generality, we take the kernelization of LU2DFDA as an example. The kernelization of RU2DFDA is similar to that of LU2DFDA. For the sake of simplicity, it is assumed that all the mapped data are centered. Similar to KFDA, a nonlinear mapping without explicit form is performed. Different from KFDA, this mapping is performed on each row vector of all the training and test image matrices, i.e., let $\Phi : \mathbf{R}^t \to \mathbf{R}^f, f > t$, be a nonlinearly mapping on each row of the image, where t is the length of the rows of an image and f can be arbitrarily large. The dot product in the feature space of $\mathbf{R}^f$ can be conveniently calculated via a pre-defined kernel function, such as the commonly used Gaussian RBF kernel function. Let $\Phi(\mathbf{X}_i^j)$ be the j-th mapped image of the i-th class. $\Phi(\mathbf{X}_i^j(:,k))$ be the k-th column vector of it. $\Phi(\overline{\mathbf{M}})$ is the mapped mean of all the training samples. $\Phi(\overline{\mathbf{M}}_i)$ is the mapped mean of i-th class.

Theorem 2: The above defined kernelized 2DFDA on the images is essentially the KFDA performed locally on the columns of all the training image matrices.

Proof: The within-class covariance matrix, $\mathbf{S}_w^\Phi$ and the between-class covariance matrix, $\mathbf{S}_b^\Phi$, in $\mathbf{R}^f$ are:

$$\mathbf{S}_w^\Phi = \sum_{i=1}^{L} \sum_{j=1}^{L_i} (\Phi(\mathbf{X}_i^j) - \Phi(\overline{\mathbf{M}}_i))(\Phi(\mathbf{X}_i^j) - \Phi(\overline{\mathbf{M}}_i))^T \tag{4}$$

$$\mathbf{S}_b^\Phi = \sum_{i=1}^{L} L_i (\Phi(\overline{\mathbf{M}}_i) - \Phi(\overline{\mathbf{M}}))(\Phi(\overline{\mathbf{M}}_i) - \Phi(\overline{\mathbf{M}}))^T \tag{5}$$

To perform Fisher discriminant analysis in $\mathbf{R}^f$, it is equivalent to maximizing:

$$\mathbf{J}(\vec{\omega}) = \frac{\vec{\omega}^T \mathbf{S}_b^\Phi \vec{\omega}}{\vec{\omega}^T \mathbf{S}_w^\Phi \vec{\omega}} \tag{6}$$

Because any solution $\vec{\omega} \in \mathbf{R}^f$ must lie in the span of the columns of all the training samples, i.e., there exist coefficients $\vec{\alpha} = [\alpha_1, \alpha_2, ..., \alpha_{(n\sum_{i=1}^{L} L_i)}]^T$ such that,

$$\vec{\omega} = [[\Phi(\mathbf{X}_1^1(:,1)), \Phi(\mathbf{X}_1^1(:,2)), ..., \Phi(\mathbf{X}_1^1(:,n))], \\ ,..., [\Phi(\mathbf{X}_L^{L_i}(:,1)), \Phi(\mathbf{X}_L^{L_i}(:,2)), ..., \Phi(\mathbf{X}_L^{L_i}(:,n))]]\vec{\alpha} \tag{7}$$

Therefore, the projection of $\Phi(\overline{\mathbf{M}}_i(:,k))$, the k-th column of the i-th class mean, onto $\vec{\omega}$, i.e., $\vec{\omega}^T \Phi(\overline{\mathbf{M}}_i(:,k))$, can be written as:

$$\vec{\alpha}^T \begin{bmatrix} \Phi(\mathbf{X}_1^1(:,1))^T \\ \Phi(\mathbf{X}_1^1(:,2))^T \\ \cdots \\ \Phi(\mathbf{X}_L^{L_i}(:,n))^T \end{bmatrix} \frac{1}{L_i} \sum_{j=1}^{L_i} \Phi(\overline{\mathbf{M}}_i^j)(:,k) = \vec{\alpha}^T \mathscr{M}_i^k \tag{8}$$

and the projection of $\Phi(\overline{\mathbf{M}}(:,k))$, the k-th column of the total class mean, onto $\vec{\omega}$, i.e., $\vec{\omega}^T \Phi(\overline{\mathbf{M}}(:,k))$ can be written as

$$\vec{\alpha}^T \begin{bmatrix} \Phi(\mathbf{X}_1^1(:,1))^T \\ \Phi(\mathbf{X}_1^1(:,2))^T \\ \cdots \\ \Phi(\mathbf{X}_L^{L_i}(:,n))^T \end{bmatrix} \frac{1}{\sum_{i=1}^{L} L_i} \sum_{i=1}^{L} \sum_{j=1}^{L_i} \Phi(\overline{\mathbf{M}}_i^j(:,k)) = \vec{\alpha}^T \mathscr{M}^k \tag{9}$$

Thus, the numerator of Eq.6, $\vec{\omega}^T \mathbf{S}_b^{\Phi} \vec{\omega}$, can be converted into:

$$\vec{\omega}^T (\sum_{i=1}^{L} L_i (\Phi(\overline{\mathbf{M}}_i) - \Phi(\overline{\mathbf{M}}))(\Phi(\overline{\mathbf{M}}_i) - \Phi(\overline{\mathbf{M}}))^T) \vec{\omega} = \vec{\omega}^T \mathbf{Q}\mathbf{Q}^T \vec{\omega} \tag{10}$$

where

$$\mathbf{Q} = [(\sqrt{L_1}(\Phi(\overline{\mathbf{M}}_1) - \Phi(\overline{\mathbf{M}}))), ...(\sqrt{L_L}(\Phi(\overline{\mathbf{M}}_L) - \Phi(\overline{\mathbf{M}})))] \tag{11}$$

Or it can be written in another form,

$$\vec{\omega}^T \mathbf{Q}\mathbf{Q}^T \vec{\omega} = \vec{\alpha}^T \mathbf{K}_b \vec{\alpha} \tag{12}$$

where $\mathbf{K}_b = \sum_{i=1}^{L} L_i (\mathscr{M}_i - \mathscr{M})(\mathscr{M}_i - \mathscr{M})^T$ and $\mathbf{K}_b$ is an $(n\sum_{i=1}^{L} L_i) \times (n\sum_{i=1}^{L} L_i)$ matrix, $\mathscr{M}_i = [\mathscr{M}_i^1, ..., \mathscr{M}_i^n]$ and $\mathscr{M}_i^j = [\Phi(\mathbf{X}_1^1(:,1)), \Phi(\mathbf{X}_1^1(:,2)), ..., \Phi(\mathbf{X}_L^{L_i}(:,n))]^T \Phi(\overline{\mathbf{M}}_i(:,j))$.

Similarly, the denominator of Eq.6,

$$\vec{\omega}^T \mathbf{S}_w^{\Phi} \vec{\omega} = \vec{\alpha}^T \mathbf{K}_w \vec{\alpha} \tag{13}$$

where $\mathbf{K}_w = \sum_{i=1}^{L} \sum_{j=1}^{L_i} (\mathscr{X}_i^j - \mathscr{M}_i)(\mathscr{X}_i^j - \mathscr{M}_i)^T$, $\mathscr{X}_i^j = [\mathscr{X}_i^j(:,1), ..., \mathscr{X}_i^j(:,n)]$, and $\mathscr{X}_i^j(:,k) = [\Phi(\mathbf{X}_1^1(:,1)), \Phi(\mathbf{X}_1^1(:,2)), ..., \Phi(\mathbf{X}_L^{L_i}(:,n))]^T \Phi(\mathbf{X}_i^j(:,k))$

Thus, the maximization of Eq.6 is converted into

$$\mathbf{J}(\vec{\alpha}) = \frac{\vec{\alpha}^T \mathbf{K}_b^{\Phi} \vec{\alpha}}{\vec{\alpha}^T \mathbf{K}_w^{\Phi} \vec{\alpha}} \tag{14}$$

Similar to FDA, this problem can be solved by finding the leading eigenvectors of $\mathbf{K}_w^{-1}\mathbf{K}_b$ if the $\mathbf{K}_w$ is not singular. Therefore, the K2DFDA performed on 2D image matrices is essentially the KFDA method performed on the columns of all the images if each

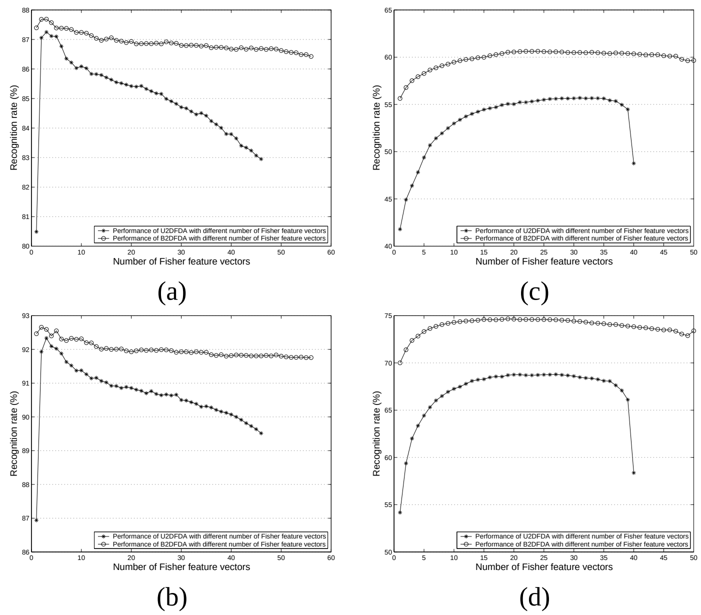

Figure 1: Comparison between U2DFDA and B2DFDA with different number of *Fisher feature vectors*. (a) and (b): Two and three training samples respectively for each subject on *ORL* database; (c) and (d): Two and three training samples respectively for each subject on YaleB.

column is viewed as a point in the vector space. Because the derivation above is based on the LU2DFDA, it is called KLU2DFDA. Similarly, the K2DFDA derived based on the RU2DFDA is called KRU2DFDA. We can also found that the KRU2DFDA is essentially the KFDA method performed on the rows of all the images if each row is viewed as a point in the vector space. □

However, because of the intrinsic shortcoming in constructing the $\mathbf{K}_w$, $\mathbf{K}_w$ is singular. To solve this problem, the $\mathbf{K}_w$ is replaced by $\mathbf{K}_w + \lambda\mathbf{I}$, where λ is a very small number and $\mathbf{I}$ is the identity matrix.

Similar to B2DFDA, a left-projection and a right-projection feature matrices can be obtained in K2DFDA. For KLU2DFDA, we project each column of the images to get the left-projection discriminant feature matrix for each image. For KRU2DFDA, we project each row of the images to get the right-projection discriminant feature matrix for each matrix. Therefore, a Kernel based B2DFDA (KB2DFDA) can be used further by combining the KLU2DFDA and KRU2DFDA. In classification, the two feature matrices are combined together for recognition via the nearest-neighborhood classifier as in B2DFDA.

5 Experimental Results and Discussions

The proposed B2DFDA and K2DFDA methods are applied to the face recognition and are evaluated on three well-known face databases: *ORL*, *UMIST* and *YaleB* (Yale Face

Table 1: Performance comparison on *ORL* database

	1	2	3	4	5
KPCA	69.5	82.5	88.8	92.1	94.2
LDA		75.8	87.0	90.1	91.7
KLDA		85.5	92.2	95.6	97.5
2DPCA	72.5	84.5	89.9	93.1	95
K2DPCA	74.5	86.9	92.0	94.6	96.2
N-LDA		74.3	82.9	87.05	88.7
D-LDA		80.6	85.6	89.5	91.7
KDDA		85.0	88.6	92.8	96.0
U2DFDA		87.4	92.5	95.1	96.3
B2DFDA		87.7	92.7	95.2	96.8
KB2DFDA		89.9	95.0	97.2	98.5

Database B) face databases. *ORL* face database contains images from 40 individuals, each providing 10 different images. *UMIST* face database consists of 564 images of 20 people with large pose variations. In our experiment, 360 images with 18 samples for each subject are used to ensure that face appearance changes from profile to frontal orientation with a step of 5° separation (labelled from #1 to #18). Yale Face Database B contains 5760 images of 10 subjects each seen under 576 viewing conditions (9 poses x 64 illumination conditions). In our experiment, altogether 640 images for 10 subjects are used (64 illumination conditions under the same frontal pose). All the images are grayscale. The images in *ORL* and *UMIST* databases are normalized to a resolution of 56×46 pixels. The images of the *YaleB* database are normalized to be the size of 50×40. *ORL* database is employed to check whether the proposed methods have good generalization ability under the circumstances that the pose, expression, and face scale variations exist concurrently. The *UMIST* and *YaleB* face databases are used to examine the performance when face orientation and illumination vary significantly respectively.

5.1 U2DFDA vs. B2DFDA

Without losing generality, the right-multiplying mode is used as an example for U2DFDA. The maximum size of the *Fisher feature matrix* is 56×46 for *ORL* database, i.e., containing at most 46 56-dimensional *Fisher feature vectors*; for Yale face database B, the maximum size of the *Fisher feature matrix* is 50×40, i.e., containing at most 40 50-dimensional *Fisher feature vectors*. We change the number of *Fisher feature vectors* from 1 to 46 for *ORL* database and from 1 to 40 for Yale face database B to see the effect on performance. We focus on testing the performance of 2DFDA when there are only few training samples for each subject, say, only 2 and 3 training samples for each subject. Fig.1 (a) and (b) show the performance of U2DFDA on *ORL* database. The optimal number of the *Fisher feature vectors* in both trials is 3. Fig.1 (c) and (d) show the performance of U2DFDA on Yale face database B. The optimal number of the *Fisher feature vectors* in the two trials is 31 and 27 respectively. From the experiment results, it can be seen that the optimal number of *Fisher feature vectors* for classification in U2DFDA is different on different database. Even on the same database, the optimal number will vary when the

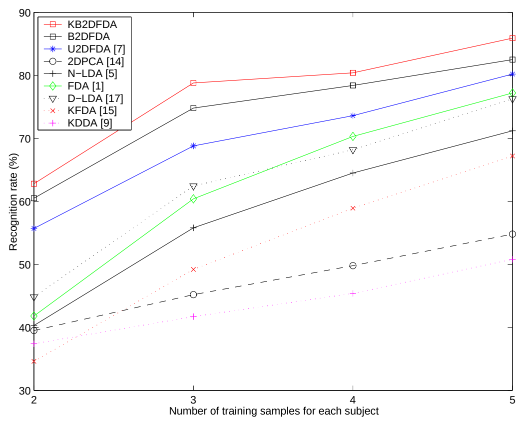

Figure 2: Performance comparison on *YaleB*

number of training samples for each subject is different. By fixing the optimal number of the *Fisher feature vectors* of the right-multiplying U2DFDA, we change the number of the *Fisher feature vectors* of the left-multiplying U2DFDA (from 1 to 56 for *ORL* database and from 1 to 50 for Yale face database B) and apply B2DFDA. Fig.1 (a) and (b) show the comparison of B2DFDA and U2DFDA on *ORL* database while Fig.1 (c) and (d) show the comparison results on Yale face database B. From these experiments, it can be found that B2DFDA can achieve higher recognition rate than U2DFDA, e.g., with an increase of up to 5 percentage on *YaleB*. We also notice that the improvement of B2DFDA over U2DFDA on *YaleB* is larger than that on *ORL*.

5.2 Recognition performance comparison on *ORL*, *YaleB* and *UMIST*

To test the recognition performance with different training numbers on *ORL* and *YaleB*, k ($2 \leq k \leq 5$) images of each subject are randomly selected for training and the remaining (p-k, for *ORL*, $p = 10$; for *YaleB*, $p = 64$) images of each subject for testing. 50 times of random selections are performed for each k. The final recognition rate is the average of all. The performance of B2DFDA and K2DFDA compared with that of the state-of-the-art methods is listed in the Table 1 and Fig.2.

Two experiments, with small number of training samples (2 and 3), are conducted on *UMIST* database. When the number of training samples for each individual is 2, we select {#5,#14} face images of each subject for training, the remaining for test. When the number of training samples is 3 for each subject, six groups are selected for training, i.e., 1{#1,#7,#13}, 2{#2,#8,#14}, 3{#3,#9,#15}, 4{#4,#10,#16}, 5{#5,#11,#17} and 6{#6,#12,#18}, the remaining images corresponding to each group are used to test. The performance of the B2DFDA and K2DFDA is compared with that of the state-of-the-art methods in the Table 2. In Fisherface [1], the size of PCA subspace is constrained to $(N-C)$, the classification dimension is set to be $(C-1)$, where N is the total number of the training samples, C is the number of the classes. In D-LDA [17] and N-LDA [5], the

Table 2: Performance comparison on $UMIST$ database

	#5, #14	#1, #7, #13	#2, #8, #14	#3, #9, #15	#4, #10, #16	#5, #11, #17	#6, #12, #18
KPCA	80.9	86.0	87.0	91.0	92.0	89.3	87.3
LDA	77.5	90.0	91.3	95.0	96.3	94.3	91.7
KLDA	92.5	94.7	96.7	98.3	99.0	98.0	97.3
2DPCA	90.3	91.0	93.0	95.0	95.0	93.7	92.3
KDDA	87.8	94.0	96.0	95.7	97.3	95.7	95.7
K2DPCA	92.7	94.0	94.3	95.7	97.0	95.7	94.0
U2DFDA	89.3	93.7	96.7	96.7	96.7	95.0	94.0
B2DFDA	91.3	96.3	96.7	97.3	97.0	95.7	96.0
KB2DFDA	93.8	97.5	97.6	98.4	98.6	96.9	97.8

classification dimensions are both set to be $(C-1)$. In 2DPCA and U2DFDA, we try to find the optimal numbers of *Eigen feature vector* and *Fisher feature vector* which give the best classification. In B2DFDA, the numbers of the right- and left-multiplying *Fisher feature vector* are set to be equal, both being the optimal number in U2DFDA. In KLDA, the dimension for classification is reserved to be $(C-1)$.

The Gaussian RBF kernel is adopted in the K2DPCA [6] and K2DFDA in all the experiments, the optimal results are achieved when the width, δ, of the kernel is around 2.718. Through experiments, we find that B2DFDA is better than U2DFDA, K2DFDA achieves the best recognition performance in all the experiments. We also find that B2DFDA is superior to the 2DPCA, K2DPCA, FDA/LDA, N-LDA, D-LDA and KDDA. B2DFDA is even comparable to KLDA. On *ORL*, 2DPCA is better than KPCA, LDA, N-LDA and D-LDA. However, on *YaleB*, the performance of 2DPCA is inferior to LDA-based algorithms. This verifies what we have analyzed previously that 2DPCA is sensitive to illumination variations.

6 Conclusions

A framework of G2DFDA is proposed to extend the original 2DFDA in three ways: firstly, the essence of 2DFDA is clarified. Secondly, an asynchronously B2DFDA scheme is introduced so that both the discriminative information encoded in rows and columns is extracted. Thirdly, a K2DFDA scheme is proposed to remedy the shortage of 2DFDA in exploring the higher-order statistics among the input rows/columns. Extensively experimental results shows that this generalization enhances the recognition performance compared with the current subspace methods.

References

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

[2] H. Cevikalp, M. Neamtu, M. Wilkes, and A. Barkana. Discriminative common vectors for face recognition. *IEEE Trans. on PAMI*, 2005.

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

[4] P. Howland and H. Park. Generalizing discriminant analysis using the generalized singular value decomposition. *IEEE Trans. on PAMI*, 2004.

[5] R. Huang, Q.S. Liu, H.Q. Lu, and S.D. Ma. Solving the small sample size problem of lda. In *Proceedings of ICPR*, 2002.

[6] H. Kong, X. Li, L. Wang, E.K. Teoh, J.G. Wang, and R. Venkateswarlu. Generalized 2d principal component analysis. In *Proceedings of IJCNN*, 2005.

[7] H. Kong, E.K. Teoh, J.G. Wang, and R. Venkateswarlu. Two dimensional fisher discriminant analysis: Forget about small sample size problem. In *Proceedings of ICASSP*, 2005.

[8] S.Z. Li, J. Yan, X.W. Hou, Z.Y. Li, and H.J. Zhang. Learning low dimensional invariant signature of 3-d object under varying view and illumination from 2-d appearances. In *IEEE ICCV*, 2001.

[9] J. Lu, K.N. Plataniotis, and A.N. Venetsanopoulos. Face recognition using kernel direct discriminant analysis algorithms. *IEEE Trans. on Neural Networks*, 2003.

[10] J. Lu, K.N. Plataniotis, and A.N. Venetsanopoulos. Regularized discriminant analysis for the small sample size problem in face recognition. *Pattern Recognition Letter*, 2003.

[11] S. Raudys and R.P.W. Duin. On expected classification error of the fisher linear classifier with pseudo-inverse covarianve matrix. *Pattern Recognition Letter*, 1998.

[12] B. Scholkopf, A. Smola, and K.R. Muller. Nonlinear component analysis as a kernel eigenvalue problem. *Neural Computation*, 1998.

[13] D.L. Swets and J. Weng. Using discriminant eigenfeatures for image retrieval. *IEEE Trans. on PAMI*, 1996.

[14] J. Yang, D. Zhang, A.F. Frangi, and J. Yang. Two-dimensional pca: A new approach to appearance-based face representation and recognition. *IEEE Trans. on PAMI*, 2004.

[15] M.H. Yang. Kernel eigenface vs. kernel fisherface: Face recognition using kernel methods. In *IEEE International Conf. on FGR*, 2002.

[16] J. Ye, R. Janardan, C.H. Park, and H. Park. An optimization criterion for generalized discriminant ananlysis on undersampled problems. *IEEE Trans. on PAMI*, 2004.

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