A KERNEL MACHINE BASED APPROACH FOR MULTI-VIEW FACE RECOGNITION

Juwei Lu, K.N. Plataniotis, A.N. Venetsanopoulos

Multimedia Laboratory
Edward S. Rogers Sr. Department of Electrical and Computer Engineering
University of Toronto, Toronto, M5S 3G4, ONTARIO, CANADA

ABSTRACT

Techniques that can introduce low-dimensional feature representation with enhanced discriminatory power is of paramount importance in face recognition applications. It is well known that the distribution of face images, under a perceivable variation in viewpoint, illumination or facial expression, is highly nonlinear and complex. It is therefore, not surprising that linear techniques, such as those based on Principle Component Analysis (PCA) or Linear Discriminant Analysis (LDA) cannot provide reliable and robust solutions to those complex face recognition problems. In this paper, we propose a kernel machine based discriminant analysis method, which deals with the nonlinearity of the face patterns' distribution. The proposed method also effectively solves the "small sample size" (SSS) problem which exists in most face recognition tasks. The new algorithm has been tested, in terms of error rate performance, on the multi-view UMIST Face Database. Results indicate that the proposed methodology outperform other commonly used approaches, such as the Kernel-PCA (KPCA) and the Generalized Discriminant Analysis (GDA).

1. INTRODUCTION

Feature selection for face representation is one of central issues to face recognition (FR) systems. Among various solutions to the problem, the most successful are those based on statistical pattern recognition techniques such as PCA and LDA. These methods, including the well known methods of Eigenfaces [1] and Fisherfaces [2], generally operate directly on images or appearances of objects and avoid difficulties in 3D modeling.

When it comes to solving problems of pattern classification, research indicates that LDA based algorithms outperform PCA based approaches, since the former optimizes the low-dimensional representation of the objects with focus on the most discriminant feature extraction while the latter achieves simply object reconstruction [2, 3, 4]. However, many LDA based algorithms suffer from the so-called "small sample size" (SSS) problem which exists in high-dimensional pattern recognition tasks where the number of available samples is smaller than the dimensionality of the samples [2, 3, 4]. The traditional solution to the SSS problem is to utilize PCA concepts in construction with LDA (PCA+LDA) for example in Fisherface [2]. Recently, more effective solutions, called Direct LDA (D-LDA) methods, have been presented [3, 4]. Although successful in many cases, these linear methods fail to deliver good performance, when face patterns are subject to large variations in viewpoints, resulting in a highly non convex and complex distribution. The limited success of these methods should be attributed to their linear nature. As a result, it is reasonable to assume that a better solution to this inherent nonlinear problem could be achieved using nonlinear methods.

In this paper, motivated by the success of Support Vector Machines (SVM) [5], KPCA [6] and GDA [7] in pattern regression and classification tasks, we propose a new kernel machine based discriminant analysis algorithm for face recognition. The algorithm combines the strengths of the traditional D-LDA and kernel machine algorithms while at the same time overcomes many of their shortcomings and limitations. In the proposed here method, following the SVM paradigm, we first nonlinearly map the original input space to an implicit high-dimensional feature space, where the distribution of face patterns is hoped to be linearized and simplified. Then, a new variant of the D-LDA method is introduced to effectively solve the SSS problem and derive a set of optimal discriminant basis vectors in the feature space. The proposed variant of D-LDA avoids a problem resulting from the wage of the zero eigenvalues of the within-class scatter matrix as possible divisors in [4]. Therefore, the proposed algorithm can be seen as an enhanced kernel D-LDA method (hereafter KDDA).

2. METHODS

The problem to be solved can be stated as follows: A set of L training face images $\{\mathbf{z}_i\}_{i=1}^{L}$, each of which is defined as a vector of length $N(= I_w \times I_h)$, i.e. $\mathbf{z}_i \in \mathbb{R}^N$, where $I_w \times I_h$ is the face image size and $\mathbb{R}^N$ denotes a N-dimensional real space, is available. Each image belongs to one of C classes $\{\mathbf{Z}_i\}_{i=1}^{C}$, and each class $\mathbf{Z}_i$ contains C_i images, resulting to $L = \sum_{i=1}^{C} C_i$. Our objective is to find a transformation φ, based on optimization of certain separability criteria, to produce a mapping $\mathbf{y}_i = \varphi(\mathbf{z}_i)$, which leads to an enhanced separability of different face objects.

2.1. Direct LDA (D-LDA)

Let $\mathbf{S}_{BTW}$ and $\mathbf{S}_{WTH}$ denote the between- and within-class scatter matrices of the training set $\{\mathbf{z}_i\}_{i=1}^{L}$. The LDA technique finds a set of optimal discriminant basis vectors, denoted by $\Psi = [\psi_1, \ldots, \psi_M]$, in such a way that the ratio

The authors would like to thank Dr. Daniel Graham and Dr. Nigel Allinson for providing the UMIST face database.

of $\mathbf{S}_{BTW}$ and $\mathbf{S}_{WTH}$ is maximized. The maximization is equivalent to solve the following eigenvalue problem,

$$\Psi = \arg\max_{\Psi} \frac{\left|(\Psi^T \mathbf{S}_{BTW} \Psi)\right|}{\left|(\Psi^T \mathbf{S}_{WTH} \Psi)\right|} \quad (1)$$

In most cases, the number of training samples L is much smaller than the dimensionality N leading to a degenerated $\mathbf{S}_{WTH}$, which produced the SSS problem. Traditional methods including the GDA and Fisherfaces solve the SSS problem just by removing null space from $\mathbf{S}_{WTH}$. Nevertheless, it has been shown that the null space may contain significant discriminant information [3, 4].

Recently, Yang et al. [4] and Chen et al. [3] proposed direct LDA (D-LDA) algorithms respectively to solve the SSS problem without losing significant discriminant information. The basic idea behind the algorithms is that the null space of $\mathbf{S}_{WTH}$ may contain significant discriminant information if the projection of $\mathbf{S}_{BTW}$ is not zero in that direction, and that no significant information will be lost if the null space of $\mathbf{S}_{BTW}$ is discarded. Assuming, for example, that $\mathcal{A}$ and $\mathcal{B}$ represent the null space of $\mathbf{S}_{BTW}$ and $\mathbf{S}_{WTH}$ respectively, the complement space of $\mathcal{A}$ and $\mathcal{B}$ can be written as $\mathcal{A}' = \mathbb{R}^N - \mathcal{A}$ and $\mathcal{B}' = \mathbb{R}^N - \mathcal{B}$. Then, the optimal discriminant subspace sought by the D-LDA algorithm is the intersection space $(\mathcal{A}' \cap \mathcal{B})$.

The difference between Chen's and Yang's methods is that Yang's method firstly diagonalizes $\mathbf{S}_{BTW}$ to find $\mathcal{A}'$, while Chen's method firstly diagonalizes $\mathbf{S}_{WTH}$ to find $\mathcal{B}$. Although there is no significant difference between the two, it may be intractable to calculate $\mathcal{B}$ when the size of $\mathbf{S}_{WTH}$ is large, which is the case in most FR tasks.

2.2. Kernel Direct Discriminant Analysis (KDDA).

Let $\phi : \mathbf{z} \in \mathbb{R}^N \rightarrow \phi(\mathbf{z}) \in \mathbb{F}$ be a nonlinear mapping from the input space to a high-dimensional feature space $\mathbb{F}$ with linear properties. The KDDA attempts to implement an improved D-LDA in the feature space.

2.2.1. Eigen-analysis of $\mathbf{S}_{BTW}$ in the Feature Space

We start by solving the eigenvalue problem of $\mathbf{S}_{BTW}$, which can be expressed as: $\mathbf{S}_{BTW} = \sum_{i=1}^{C} \tilde{\bar{\phi}}_i \tilde{\bar{\phi}}_i^T = \Phi_b \Phi_b^T$, where $\Phi_b = \left[\tilde{\bar{\phi}}_1 \cdots \tilde{\bar{\phi}}_c\right]$, $\tilde{\bar{\phi}}_i = \sqrt{\frac{C_i}{L}}\left(\bar{\phi}_i - \bar{\phi}\right)$, $\bar{\phi}_i = \frac{1}{C_i}\sum_{j=1}^{C_i} \phi(\mathbf{z}_{ij})$ is the mean of class $\mathbf{Z}_i$ and $\bar{\phi} = \frac{1}{L}\sum_{n=1}^{L} \phi(\mathbf{z}_n)$ is the mean of the ensemble. Since the dimensionality of the feature space $\mathbb{F}$, denoted as N', could be arbitrarily large or possibly infinite, it is intractable to directly compute the eigenvectors of the $(N' \times N')$ matrix $\mathbf{S}_{BTW}$. Fortunately, the first m $(\leq C-1)$ most significant eigenvectors of $\mathbf{S}_{BTW}$, which correspond to non-zero eigenvalues, can be indirectly derived from the eigenvectors of $\Phi_b^T \Phi_b$ (with size $C \times C$).

Computing $\Phi_b^T \Phi_b$, requires dot product evaluation in $\mathbb{F}$. This can be done in a manner similar to the one used in SVM by utilizing kernel machine methods. For any $\phi(\mathbf{z}_i), \phi(\mathbf{z}_j) \in \mathbb{F}$, we assume that there exists a kernel function $k(\cdot)$ such that $k(\mathbf{z}_i, \mathbf{z}_j) = \phi(\mathbf{z}_i) \cdot \phi(\mathbf{z}_j)$. The introduction of the kernel function allows us to avoid the explicit evaluation of the mapping. Typical kernel functions that can be used for the task, include polynomial function, radial basis function (RBF) and multi-layer perceptrons [5].

Using the kernel function, for two arbitrary classes $\mathbf{Z}_l$ and $\mathbf{Z}_h$, a $C_l \times C_h$ dot product matrix K_{lh} can be defined as: $K_{lh} = (k_{ij})_{\substack{i=1,\cdots,C_l \\ j=1,\cdots,C_h}}$, where $k_{ij} = k(\mathbf{z}_{li}, \mathbf{z}_{hj})$. For all of classes $\{\mathbf{Z}_i\}_{i=1}^{C}$, we can define a $L \times L$ kernel matrix $\mathbf{K} = (K_{lh})_{\substack{l=1,\cdots,C \\ h=1,\cdots,C}}$, which allows us to express $\Phi_b^T \Phi_b$ as:

$$\begin{aligned}\Phi_b^T \Phi_b = \; & \tfrac{1}{L}\mathbf{B} \cdot (\mathbf{A}_{LC}^T \cdot \mathbf{K} \cdot \mathbf{A}_{LC} - \tfrac{1}{L}(\mathbf{A}_{LC}^T \cdot \mathbf{K} \cdot \mathbf{1}_{LC}) - \\ & \tfrac{1}{L}(\mathbf{1}_{LC}^T \cdot \mathbf{K} \cdot \mathbf{A}_{LC}) + \tfrac{1}{L^2}(\mathbf{1}_{LC}^T \cdot \mathbf{K} \cdot \mathbf{1}_{LC})) \cdot \mathbf{B}\end{aligned}$$

where $\mathbf{B} = \mathbf{diag}\left[\sqrt{C_1} \; \cdots \; \sqrt{C_c}\right]$, $\mathbf{1}_{LC}$ is a $L \times C$ matrix with terms all equal to: one, $\mathbf{A}_{LC} = \mathbf{diag}\left[\mathbf{a}_{c_1} \; \cdots \; \mathbf{a}_{c_c}\right]$ is a $L \times C$ block diagonal matrix, and $\mathbf{a}_{c_i}$ is a $C_i \times 1$ vector with all terms equal to: $1/C_i$.

Let λ_i and $\mathbf{e}_i$ be the i-th eigenvalue and its corresponding eigenvector of $\Phi_b^T \Phi_b$, where $i = 1 \cdots C$, sorted in decreasing order of eigenvalues. Thus, $\mathbf{v}_i = \Phi_b \mathbf{e}_i$ is the eigenvector of $\mathbf{S}_{BTW}$. To remove the null space of $\mathbf{S}_{BTW}$, we only use its first m $(\leq C-1)$ eigenvectors whose corresponding $\lambda_i > 0$, denoting them as $\mathbf{V} = [\mathbf{v}_1 \; \cdots \; \mathbf{v}_m] = \Phi_b \mathbf{E}_m$, and it is not difficult to see that $\mathbf{V}^T \mathbf{S}_{BTW} \mathbf{V} = \Lambda_b$, where $\mathbf{E}_m = [\mathbf{e}_1 \ldots \mathbf{e}_m]$ and $\Lambda_b = \mathbf{diag}[\lambda_1^2 \; \cdots \; \lambda_m^2]$, a $m \times m$ diagonal matrix.

2.2.2. Eigen-analysis of $\mathbf{S}_{WTH}$ in the Feature Space

Let $\phi_{ij} = \phi(\mathbf{z}_{ij})$, $\mathbf{S}_{WTH}$ can be expressed as: $\mathbf{S}_{WTH} = \frac{1}{L}\sum_{i=1}^{C}\sum_{j=1}^{C_i}(\phi_{ij} - \bar{\phi}_i)(\phi_{ij} - \bar{\phi}_i)^T$.

Let $\mathbf{U} = \mathbf{V}\Lambda_b^{-1/2}$. Projecting $\mathbf{S}_{BTW}$ and $\mathbf{S}_{WTH}$ into the subspace spanned by $\mathbf{U}$, we have $\mathbf{U}^T \mathbf{S}_{BTW} \mathbf{U} = \mathbf{I}$ and $\mathbf{U}^T \mathbf{S}_{WTH} \mathbf{U} = (\mathbf{E}_m \Lambda_b^{-1/2})^T (\Phi_b^T \mathbf{S}_{WTH} \Phi_b)(\mathbf{E}_m \Lambda_b^{-1/2})$. Using the kernel matrix $\mathbf{K}$, we can obtain a close form expression of $\Phi_b^T \mathbf{S}_{WTH} \Phi_b = \frac{1}{L}(\mathbf{J}1 - \mathbf{J}2)$, where

$$\begin{aligned}\mathbf{J}1 = \; & \tfrac{1}{L}\mathbf{B} \cdot (\mathbf{A}_{LC}^T \cdot \mathbf{K} \cdot \mathbf{K} \cdot \mathbf{A}_{LC} - \tfrac{1}{L}(A_{Nc}^T \cdot \mathbf{K} \cdot \mathbf{K} \cdot \mathbf{1}_{LC}) - \\ & \tfrac{1}{L}(\mathbf{1}_{LC}^T \cdot \mathbf{K} \cdot \mathbf{K} \cdot \mathbf{A}_{LC}) + \tfrac{1}{L^2}(\mathbf{1}_{LC}^T \cdot \mathbf{K} \cdot \mathbf{K} \cdot \mathbf{1}_{LC})) \cdot \mathbf{B}\end{aligned}$$

$$\begin{aligned}\mathbf{J}2 = \; & \tfrac{1}{L}\mathbf{B}(\mathbf{A}_{LC}^T \mathbf{K} \cdot \mathbf{W} \cdot \mathbf{K}\mathbf{A}_{LC} - \tfrac{1}{L}(\mathbf{A}_{LC}^T \mathbf{K} \cdot \mathbf{W} \cdot \mathbf{K}\mathbf{1}_{LC}) - \\ & \tfrac{1}{L}(\mathbf{1}_{LC}^T \mathbf{K} \cdot \mathbf{W} \cdot \mathbf{K}\mathbf{A}_{LC}) + \tfrac{1}{L^2}(\mathbf{1}_{LC}^T \mathbf{K} \cdot \mathbf{W} \cdot \mathbf{K}\mathbf{1}_{LC}))\mathbf{B}\end{aligned}$$

where $\mathbf{W} = \mathbf{diag}\left[\mathbf{w}_1 \; \cdots \; \mathbf{w}_c\right]$ is a $L \times L$ block diagonal matrix, and $\mathbf{w}_i$ is a $C_i \times C_i$ matrix with terms all equal to: $\frac{1}{C_i}$.

Then, diagonalize $\mathbf{U}^T \mathbf{S}_{WTH} \mathbf{U}$ which is a tractable matrix with size $m \times m$, and let $\mathbf{p}_i$ be the i-th eigenvector of $\mathbf{U}^T \mathbf{S}_{WTH} \mathbf{U}$, where $i = 1 \cdots m$, sorted in increasing order of $\mathbf{p}_i$'s corresponding eigenvalue λ_i'. In the set of ordered eigenvectors, those that correspond to the smallest eigenvalues should be considered the most discriminative features, which maximize the ratio in Eq.1. We can optionally discard those eigenvectors with the largest eigenvalues, and denote the $M(\leq m)$ selected eigenvectors as $\mathbf{P} = [\mathbf{p}_1 \; \cdots \; \mathbf{p}_M]$. Let $\mathbf{Q} = \mathbf{UP}$, and thus $\mathbf{Q}^T \mathbf{S}_{WTH} \mathbf{Q} = \Lambda_w$, with $\Lambda_w = \mathbf{diag}[\lambda_1' \; \cdots \; \lambda_M']$, a $M \times M$ diagonal matrix.

Based on the calculations presented above, a set of optimal discriminant feature vectors can be derived through $\Gamma = \mathbf{Q}\Lambda_w^{-1/2}$. The features form a low-dimensional subspace in $\mathbb{F}$, where the ratio in Eq.1 is maximized. Similar to the D-LDA framework, the subspace obtained contains the

intersection space ($\mathcal{A}' \cap \mathcal{B}$) shown in section 2.1. However, it is possible that there exist eigenvalues with $\lambda'_i = 0$ in Λ_w. To alleviate the problem, any value below ϵ is promoted to ϵ (a very small value) in [4]. Obviously, performance heavily depends on the proper choice of the value ϵ, which is done in a heuristic manner [4].

To avoid the problems related to the heuristic evaluation of the parameter ϵ, we propose to use a modified Fisher's criterion instead of the conventional definition used in Eq.1 when $\mathbf{U}^T\mathbf{S}_{WTH}\mathbf{U}$ is singular. The new criterion can be expressed as:

$$\Psi = \arg\max_{\Psi} \frac{\left|(\Psi^T \mathbf{S}_{BTW} \Psi)\right|}{\left|(\Psi^T \mathbf{S}_{BTW} \Psi) + (\Psi^T \mathbf{S}_{WTH} \Psi)\right|} \qquad (2)$$

The modified Fisher's criterion of Eq.2 has been proved to be equivalent to the conventional one (Eq.1) in [8].

The expression $\mathbf{U}^T(\mathbf{S}_{BTW} + \mathbf{S}_{WTH})\mathbf{U}$ which is used in Eq.2 instead of the $\mathbf{U}^T\mathbf{S}_{WTH}\mathbf{U}$ can be shown to be non-singular by the following lemma (its proof is not difficult).

Lemma 1 Suppose $\mathbf{D}$ is a real matrix of size $\mathcal{N} \times \mathcal{N}$, and can be represented by $\mathbf{D} = \Phi\Phi^T$ where Φ is a real matrix of size $\mathcal{N} \times \mathcal{M}$. Then, $(\mathbf{I}+\mathbf{D})$ is positive definite, i.e. $\mathbf{I}+\mathbf{D} > 0$, where $\mathbf{I}$ is a $\mathcal{N} \times \mathcal{N}$ identity matrix.

Similar to $\mathbf{S}_{BTW}$, $\mathbf{S}_{WTH}$ can be expressed as $\mathbf{S}_{WTH} = \Phi_w\Phi_w^T$, resulting in $\mathbf{U}^T\mathbf{S}_{WTH}\mathbf{U} = (\mathbf{U}^T\Phi_w)(\mathbf{U}^T\Phi_w)^T$. Since $\mathbf{U}^T\mathbf{S}_{WTH}\mathbf{U}$ satisfies the conditions on $\mathbf{D}$ discussed in Lemma 1 and $\mathbf{U}^T\mathbf{S}_{BTW}\mathbf{U} = \mathbf{I}$, it can be easily seen that $\mathbf{U}^T(\mathbf{S}_{BTW} + \mathbf{S}_{WTH})\mathbf{U}$ is positive definite, resulting in a non-singular matrix: $\mathbf{Q}^T(\mathbf{S}_{BTW} + \mathbf{S}_{WTH})\mathbf{Q} = \Lambda_w$.

2.2.3. Dimensionality Reduction and Feature Extraction

For any input pattern $\mathbf{z}$, its projection into the subspace spanned by Γ, can be calculated by $\mathbf{y} = \Gamma^T\phi(\mathbf{z})$.

Let $\gamma(\phi(\mathbf{z})) = [\,\phi_{11}^T\phi(\mathbf{z}) \quad \phi_{12}^T\phi(\mathbf{z}) \quad \cdots \quad \phi_{cc_c}^T\phi(\mathbf{z})\,]^T$ be a $L \times 1$ kernel vector. After rearrangement, we can obtain $\mathbf{y} = \Theta \cdot \gamma(\phi(\mathbf{z}))$, where

$$\Theta = \frac{1}{\sqrt{L}} \left(\mathbf{E}_m \cdot \Lambda_b^{-1/2} \cdot \mathbf{P} \cdot \Lambda_w^{-1/2}\right)^T \left(\mathbf{B} \cdot \left(\mathbf{A}_{LC}^T - \frac{1}{L}\mathbf{1}_{LC}^T\right)\right)$$

is a $M \times L$ matrix which can be calculated offline.

Thus, through $\mathbf{y} = \Theta \cdot \gamma(\phi(\mathbf{z}))$, a low-dimensional representation $\mathbf{y}$ on $\mathbf{z}$ with enhanced discriminant power has been introduced.

3. EXPERIMENTAL RESULTS

Two sets of experiments are included in the paper to demonstrate the effectiveness of the proposed KDDA algorithm. To show the high nonlinearity and complexity of the face patterns' distribution, the multi-view UMIST face database, consisting of 575 gray-scale images of 20 people, is used in the experiments. Fig.1 summarizes sample images of a typical subset in the database.

Fig. 1. Some samples from the UMIST database.

3.1. Distribution of Multi-view Face Patterns

The first experiment aims to provide insights on how the KDDA linearizes and simplifies the face pattern distribution. For the sake of simplicity in visualization, we only use a subset of the database, which contains 170 images of 5 randomly selected subjects (classes). Four types of feature bases are generalized from the subset by utilizing the PCA, KPCA, D-LDA and KDDA algorithms respectively. For each image, its projections in the first two most significant feature bases of each subspace are visualized in Fig.2-3.

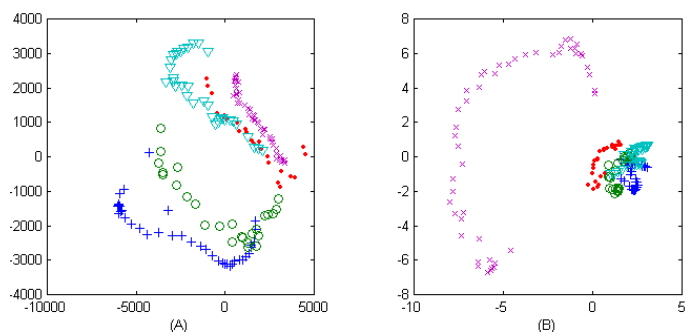

Fig. 2. Distribution of 170 samples of 5 subjects in **A**: PCA-based subspace and **B**: KPCA-based subspace.

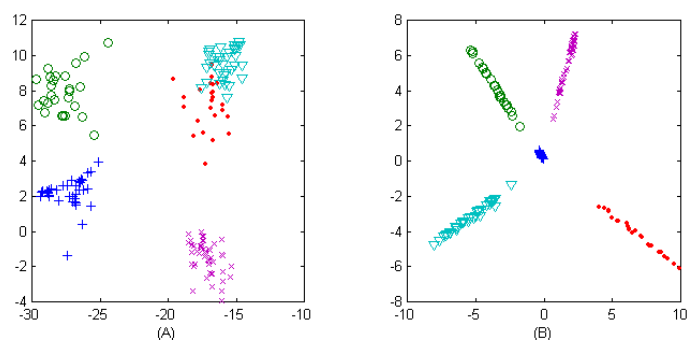

Fig. 3. Distribution of 170 samples of 5 subjects in **A**: D-LDA-based subspace and **B**: KDDA-based subspace.

In Fig.2, the visualized projections are the first two most significant principal components extracted by PCA

and KPCA, and they provide a low-dimensional representation for the samples, which can be used to capture the structure of data. Thus we can roughly learn the original distribution of the samples from Fig.2:**A**, which is non convex and complex as we expected. In Fig.2:**B**, KPCA generalizes PCA to its nonlinear counterpart using a RBF kernel: $k(\mathbf{z}_1, \mathbf{z}_2) = exp(-||\mathbf{z}_1 - \mathbf{z}_2||^2/\sigma^2)$ with $\sigma^2 = 5\mathbf{e}6$. It is hard to find any useful improvement for the purpose of pattern classification from Fig.2:**B**. It can be therefore concluded that the low-dimensional representation obtained by PCA like techniques are not necessarily useful for classification.

Fig.3 depicts the first two most discriminant features extracted by the D-LDA and KDDA respectively. Obviously, these features outperform, in terms of discriminant power, those obtained using the PCA technique. However, subject to limitation of linearity, some classes are still non-separable in the D-LDA-based subspace as shown in Fig.3:**A**. In contrast to this, we can see the linearization property of the KDDA-based subspace, depicted in Fig.3:**B**, where all of classes are well linearly separable when a RBF kernel with $\sigma^2 = 5\mathbf{e}6$ is used.

3.2. Comparison with KPCA and GDA

The second experiment compares the KDDA to the KPCA and the GDA in terms of the error rates. The FR procedure is completed in two stages: (1) Feature extraction. The overall database is randomly partitioned into two subsets: a training set and a test set. The training set is composed of 120 images: 6 images per person are randomly chosen. The remaining 455 images are used to form the test set. There is no overlapping between the two. Then, both of them are projected into the feature spaces derived by the KPCA, D-LDA and KDDA methods. (2) Classification. This is implemented by feeding feature vectors obtained in step-1 into a nearest neighbor classifier.

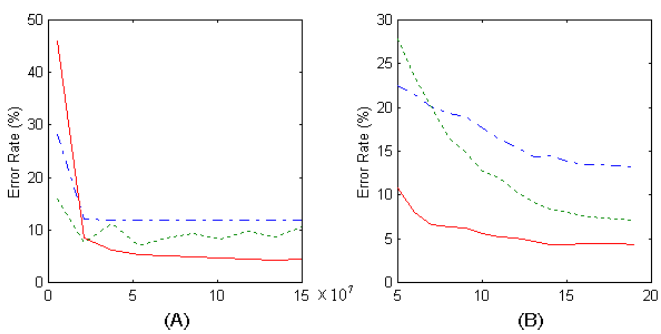

Fig. 4. Comparison of error rates, where '-.': KPCA, ':': GDA and '-': KDDA. **A**: error rates as functions of σ^2. **B**: error rate as functions of M.

The obtained error rates of the three methods when a typical RBF kernel is used, is depicted in Fig.4. Fig.4:**A** shows the error rates as functions of σ^2 within the range from $0.5\mathbf{e}7$ to $1.5\mathbf{e}8$, when the optimal number of feature vectors: ($M = M_{opt}$) is used. The optimal number can be found by searching the number of used feature vectors that results in the minimal summation of the error rates over the variation range of σ^2. In Fig.4:**A**, $M_{opt} = 99$ is found for KPCA, $M_{opt} = 19$ for GDA and $M_{opt} = 19$ for KDDA. Fig.4:**B** shows the error rates as functions of M within the range from 5 to 19, when optimal $\sigma^2 = \sigma^2_{opt}$ is used. Similar to M_{opt}, σ^2_{opt} means the variance that results in the minimal summation of the error rates over the variation range of M. $\sigma^2_{opt} = 1.5\mathbf{e}8$ is found for KPCA, $\sigma^2_{opt} = 5.3333\mathbf{e}7$ for GDA and $\sigma^2_{opt} = 1.3389\mathbf{e}7$ for KDDA in Fig.4:**B**.

Let α_M and β_M be the error rates of KDDA and any one of other two methods respectively. From Fig.4:**B**, we can obtain an interesting quantity comparison: the average percentages of the error rates of KDDA over those of other methods by $\sum_{M=5}^{19}(\alpha_M/\beta_M)$. The results show that the average error rate of KDDA is only about 33.669% and 47.866% of those of KPCA and GDA respectively.

4. CONCLUSION

A new method for face recognition has been introduced in this paper. The proposed method combines kernel-based methodologies with discriminant analysis techniques. The kernel function is utilized to map the original face patterns to a high-dimensional feature space, where the highly complex distribution of face patterns is linearized and simplified, so that linear discriminant techniques can be used for feature extraction. The SSS problem caused by high dimensionality of mapped patterns, is addressed by an improved D-LDA technique. Experimental results indicate that the performance of the KDDA algorithm is overall superior to that obtained by the KPCA or GDA approaches.

5. REFERENCES

[1] Matthew A. Turk and Alex P. 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] Li-Fen Chen, Hong-Yuan Mark Liao, Ming-Tat Ko, Ja-Chen Lin, and Gwo-Jong Yu, ""A new LDA-based face recognition system which can solve the small sample size problem"," Pattern Recognition, vol. 33, pp. 1713–1726, 2000.

[4] Hua Yu and Jie Yang, ""A direct lda algorithm for high-dimensional data with application to face recognition"," Pattern Recognition, vol. 34, pp. 2067–2070, 2001.

[5] V. Vapnik, The Nature of Statistical Learning Theory, Springer-Verlag, 1995.

[6] Schölkopf B., Smola A., and Müller K. R., ""Nonlinear component analysis as a kernel eigenvalue problem"," Neural Computation, vol. 10, pp. 1299–1319, 1999.

[7] Baudat G. and Anouar F., ""Generalized discriminant analysis using a kernel approach"," Neural Computation, vol. 12, pp. 2385–2404, 2000.

[8] K. Liu, Y.Q. Cheng, J.Y. Yang, and X. Liu, ""An efficient algorithm for foley-sammon optimal set of discriminant vectors by algebraic method"," Int. J. Pattern Recog. Artif. Intell., vol. 6, pp. 817–829, 1992.