Discriminant Analysis Based on Kernelized Decision Boundary for Face Recognition

Baochang Zhang[1], Xilin Chen[1,2], and Wen Gao[1,2]

[1] Computer School, Harbin Institute of Technology, China
{Bczhang}@jdl.ac.cn
[2] ICT-ISVISION Joint R&D Lab for face recognition, ICT, CAS, China
{Bczhang,Xlchen,Wgao}@jdl.ac.cn

Abstract. A novel nonlinear discriminant analysis method, Kernelized Decision Boundary Analysis (KDBA), is proposed in our paper, whose Decision Boundary feature vectors are the normal vector of the optimal Decision Boundary in terms of the Structure Risk Minimization principle. We also use a simple method to prove a property of Support Vector Machine (SVM) algorithm, which is combined with the optimal Decision Boundary Feature matrix to make our method consistent with the Kernel Fisher method(KFD). Moreover, KDBA is easily used in its applications, and the traditional Decision Boundary Analysis implementations are computationally expensive and sensitive to the size of the problem. Text classification problem is first used to testify the effectiveness of KDBA. Then experiments on the large-scale face database, the CAS-PEAL database, have illustrated its excellent performance compared with some popular face recognition methods such as Eigenface, Fisherface, and KFD.

Keywords: Face Recognition, Kernel Fisher, Support Vector Machine

1 Introduction

Feature extraction has long been a hot topic in the pattern recognition field. Principle Component Analysis (PCA) and Fisher Linear Discriminant Analysis (FDA) are two classical techniques for the linear feature extraction. In many applications, both methods have been proven to be very powerful. PCA is designed to capture the variance in a dataset in terms of the principle components, and FDA is a well-known discriminant tool, which maximizes the so-called Fisher criterion $J(w) = tr[(wS_w w^T)^{-1}(wS_b w^T)]$, where S_w is the within-class scatter matrix, S_b is the between-class scatter matrix. Since the FDA is mainly based on a single class center, i.e.. the mean sample of the class, the feature vector calculated by which is not reliable if mean vectors can not reflect the distribution of the data set. Moreover, PCA and FDA are inadequate to describe the complex nonlinear variations in the training dataset. In recent years, the kernelized feature extraction methods have been paid much attention, such as Kernel Principal Component Analysis (KPCA)[1] and Kernel Fisher Discriminant analysis (KFD) [1,2,3], which are well-known nonlinear extensions to PCA and FDA respectively. However, the KFD cannot be easily used in real applications. The reason is that the projection directions of KFD often lie in the span of all the samples [4], therefore, the dimension of the feature often becomes very large, when the input space is mapped to a feature space through a kernel function. As a result, the scatter matrices become singular, which is the so-called "Small Sam-

T. Kanade, A. Jain, and N.K. Ratha (Eds.): AVBPA 2005, LNCS 3546, pp. 966–976, 2005.

ple Size problem" (SSS). Similar to [5], KFD simply adds a perturbation to within-class scatter matrix. Of course, it has the same stability problem as that in [5], because eigenvectors are sensitive to the small perturbation, moreover, the influence of which is not yet understood.

In this paper, we propose a new algorithm for feature extraction based on the proposed Kernelized Decision Boundary Analysis(KDBA). The algorithm tries to extract the necessary feature vectors to achieve the same classification accuracy as in the original space. The decision boundary theory was first proposed by Fukunaga et al [6, 7] and further developed by Lee [8] et al. However, the existing implementations of calculating the Decision Boundary Feature Matrix are computationally expensive and sensitive to the sample size. Moreover, the decision boundary method is a linear method, and fails to find the nonlinear structure from the training set. We proposed the KDBA method, which is a non-linear extension to the traditional one. Furthermore, the proposed method is easily implemented and strongly related to the theory of Structure Risk Minimization (SRM).

The rest of the paper is organized as following. In Section 2, we briefly introduce the Kernel Fisher analysis method. In Section 3, we define the Kernelized Decision Boundary Feature Matrix (KDBFM), which is based on the optimal Decision Boundary vector in terms of SRM. In section 4, we proposed the algorithm of KDBA, which is easily realized in its application. In section 5, we will give some experiment results on the Text classification problem and face recognition in the large-scale CAS-PEAL face database [14,18]. In the last section, we will make some conclusions about the proposed method.

2 Kernel Fisher Discriminant Analysis

We first describe, in this paper, the Kernel Fisher analysis method, which is a well-known extension to FDA. Moreover, many definitions will be used in our paper later, such as the kernel within-class and between-class scatter matrices. It is also the baseline algorithm in our paper, and we will make some comparative experiments with the proposed method.

The idea of Kernel FDA is to yield a nonlinear discriminant analysis in a higher dimensional space. The input data is first projected into an implicit feature space F by the nonlinear mapping $\Phi : x \in R^N -> f \in F$, and then seek to find a nonlinear transformation, which can maximize the between-class scatter and minimize the within-class scatter in F [1-4]. In its implementation, Φ is implicit and we will just compute the inner product of two vectors in F by using a kernel function:

$$k(x, y) = (\Phi(x) \cdot \Phi(y)). \tag{1}$$

We define between-class scatter matrix $\mathbf{S}_b$ and within-class scatter matrix $\mathbf{S}_w$ in F as following:

$$\mathbf{S}_b = \sum_{i=1}^{C} p(\varpi_i)(u_i - u)(u_i - u)^T , \tag{2}$$

$$\mathbf{S}_W = \sum_{i=1}^{C} p(\varpi_i) E\{((\Phi(x_i) - u_i)(\Phi(x_i) - u_i)^T) \mid \varpi_i\}, \tag{3}$$

$u_i = \frac{1}{n_i}\sum_{j=1}^{n_i}\phi(x_{ij})$ denotes the sample mean of the class i, and u is the mean of all training images in F, $p(\varpi_i)$ is the prior probability. To perform FDA in a higher dimensional space F, it is equal to maximize Eq.4.

$$J(w) = \frac{\mathbf{w}^T\mathbf{S}_b\mathbf{w}}{\mathbf{w}^T\mathbf{S}_w\mathbf{w}} = \frac{tr(\mathbf{S}_b)}{tr(\mathbf{S}_w)} . \quad (4)$$

Because any solution $\mathbf{w} \in F$ should lie in the span of all the samples in F [9,10], there exists:

$$\mathbf{w} = \sum_{i=1}^{n}\alpha_i\phi(x_i), \boldsymbol{\alpha}_i, i = 1,2...n. \quad (5)$$

Then we will get the following Maximizing Criterion:

$$J(\boldsymbol{\alpha}) = \frac{\boldsymbol{\alpha}^T\mathbf{K}_b\boldsymbol{\alpha}}{\boldsymbol{\alpha}^T\mathbf{K}_w\boldsymbol{\alpha}} , \quad (6)$$

where $\mathbf{K}_w$ and $\mathbf{K}_b$ are defined as following:

$$\mathbf{K}_w = \sum_{i=1}^{C}p(\varpi_i)E(\eta_j - m_i)(\eta_j - m_i)^T , \quad (7)$$

$$\mathbf{K}_b = \sum_{i=1}^{C}p(\varpi_i)(m_i - \bar{m})(m_i - \bar{m})^T , \quad (8)$$

where $\eta_j = (k(x_1, x_j), k(x_2, x_j),..., k(x_n, x_j))^T$,

$m_i = \left(\frac{1}{n_i}\sum_{j=1}^{n_i}k(x_1,x_j), \frac{1}{n_i}\sum_{j=1}^{n_i}k(x_2,x_j),.., \frac{1}{n_i}\sum_{j=1}^{n_i}k(x_n,x_j)\right)$, and $\bar{m}$ is the mean of all η_j.

Similar to FDA [5], this problem can be solved by finding the leading eigenvectors of $\mathbf{K}_w^{-1}\mathbf{K}_b$ used by Liu [9] and Baudat (GDA) [3], which is the Generalized Kernel Fisher Discriminant (GKFD) criterion. In our paper, we use the technique of the pseudo inverse of the within-class scatter matrix, and then perform PCA on $\mathbf{K}_w^{-1}\mathbf{K}_b$ to get the transformation matrix $\boldsymbol{\alpha}$. The projection of a data point x onto $\mathbf{w}$ in F is given by:

$$v = (\mathbf{w} . \Phi(x)) = \sum_{i=1}^{n}\boldsymbol{\alpha}_i k(x_i, x) . \quad (9)$$

3 Decision Boundary Feature Matrix

In this part, we will first briefly review some basic properties of the discriminant subspaces, for a detailed description and proofs, we can refer to[8].

Property 1. If a vector V is orthogonal to the vector normal to decision boundary at every point on decision boundary, V contains no information useful in discriminating classes, i.e., vectors discriminantly redundant. If a vector is normal to the decision boundary at least one point on the decision boundary, the vector contains information useful in discriminating classes, i.e., the vector is discriminantly informative. And we can also refer to Fig.1.

Property 2. The Decision Boundary Feature Matrix (DBFM): let $N(x)$ be the unit vector normal to the decision boundary at a point x on the decision boundary for a given pattern classification problem, $p(x)$ is the data density. Then the DBFM is defined as following

$$\sum_{DBFM} = \int_S N(x)N^t(x)p(x)dx \cdot \tag{10}$$

Property 3. The eigenvectors of the Decision Boundary Feature Matrix of a pattern recognition problem corresponding to non-zero eigenvectors are the necessary feature vectors to achieve the same classification accuracy as in the original space for the pattern recognition problem.

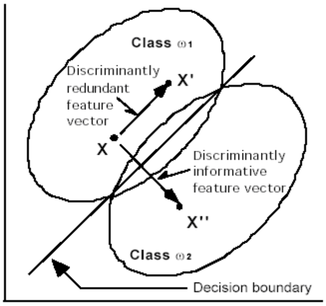

Fig. 1. X' is discriminantly redundant feature vector, X'' is discriminantly informative feature vector

3.1 The Property of SVM

Support Vector Machines (SVM) is a state-of-the-art pattern recognition technique, whose foundations stem from the statistical learning theory. However, the scope of SVM is beyond pattern recognition because they can also handle another two learning problems, i.e., regression estimation and density estimation. SVM is a general algorithm based on guaranteed risk bounds of statistical learning, the so-called structural risk minimization principle. And we can refer to the tutorials [11] about the SVM. The success of SVM in face recognition [12, 13] as a recognizer provides us with further motivations to utilize SVM to enhance the performance of our system. However, we did not construct SVM classifier, and just used it to find the support vectors shown in Fig.2.

Here, we will use a simple way to prove the property of the support vectors based within-class scatter matrix for two-class problem, which shows that the SVM is strongly related to the Kernel Fisher analysis method.

In a higher dimensional space, x_1, x_2 are represented as $\Phi(x_1), \Phi(x_2)$. The SVM aims to optimize the following objective function [11]:

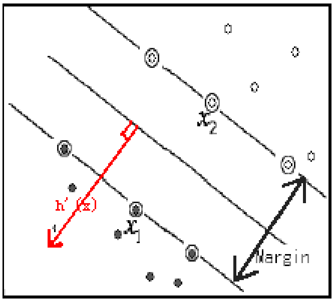

Fig. 2. Support Vectors are circled such as x_1, x_2

$$\underset{w}{\text{Min}} \ \frac{1}{2} w^T w \tag{11}$$

$$\text{Subject to: } y_i (w^T \Phi(x_i) + b) - 1 \geq 0 . \tag{12}$$

The decision hyperplane function is linear style in a higher dimensional space as following:

$$h(\Phi(x)) = w^T \Phi(x_i) + b \cdot \tag{13}$$

If $\Phi(x_i)$ is the support vector, we can know that [9]:

$$y_i (w^T \Phi(x_i) + b) - 1 = 0, y_i \in \{-1,1\} . \tag{14}$$

Thus, for $\Phi(x_1), \Phi(x_2)$ are the support vectors, we have:

$$\begin{gathered} w^T \Phi(x_1) + b = 1, \Phi(x_1) \in S_1, \\ S_1 = \{\Phi(x_i) \mid y_i = 1, w^T \Phi(x_i) + b = 1\} \end{gathered} , \tag{15}$$

$$\begin{gathered} w^T \Phi(x_2) + b = -1, \Phi(x_2) \in S_2, \\ S_2 = \{\Phi(x_i) \mid y_i = -1, w^T \Phi(x_i) + b = -1\} \end{gathered} . \tag{16}$$

The elements of S_1 and S_2 are support vectors, and it is easy for us to prove the following equations:

$$\begin{gathered} w^T \ \overline{\Phi(x_1)} = 1 - b, \\ \overline{\Phi(x_1)} = \frac{1}{n_1} \sum \Phi(x), \Phi(x) \in S_1 \end{gathered} , \tag{17}$$

$$\begin{gathered} w^T \ \overline{\Phi(x_2)} = -1 - b, \\ \overline{\Phi(x_2)} = \frac{1}{n_2} \sum \Phi(x), \Phi(x) \in S_2 \end{gathered} , \tag{18}$$

n_1 is the size of the S_1, n_2 is the size of the S_2. For two-class problem, the within-class scatter matrix is defined as following:

$$\begin{aligned} \mathbf{S}'_w = & \frac{n_1}{n_2 + n_1} \sum_{i=1, \Phi(x_i) \in S_1}^{n_1} (\Phi(x_i) - \overline{\Phi(x_1)})(\Phi(x_i) - \overline{\Phi(x_1)})^T + \\ & \frac{n_2}{n_2 + n_1} \sum_{i=1, \Phi(x_i) \in S_2}^{n_2} (\Phi(x_i) - \overline{\Phi(x_2)})(\Phi(x_i) - \overline{\Phi(x_2)})^T \end{aligned} . \tag{19}$$

Therefore, we can know that:

$$\mathbf{w}^T \mathbf{S}'_w \mathbf{w} = 0 , \tag{20}$$

where $\mathbf{S}_w^{'}$ is the within-class scatter matrix calculated by using the support vectors. $h_{\Phi(x)}^{'}$ is the normal vector of the decision hyperplane in a higher dimensional space, which is the optimal decision boundary in terms of the SRM. We called w the Kernelized Decision Bounday Feature vector calculated as following:

$$w = h_{\Phi(x)}^{'} = \sum_{i}^{n} \alpha_i y_i \Phi(x_i), \quad (21)$$

α_i is calculated in the SVM algorithm, and $\Phi(x_i)$ is the support vector.

3.2 Kernelized Decision Boundary Feature Matrix

In this part, we will define the KDBFM for the multi-class problem. The face Samples of S_i are from the class C_i , and the dataset $S_i^{'}$ includes samples from $C_l, l \neq i$. We will use SVM algorithm to find the Decision Boundary Feature vectors by using Eq.21, and first divide the dataset $S_i^{'}$ into k, $1 <= k <= C-1$ groups, represented by $S_{ij}^{'}, j = 1,...,k$. Now for any pair of data sets, such as $(S_i, S_{ij}^{'}), j = 1,...,k$, we use the two-class SVM algorithm to find the Decision Boundary Feature vector $\mathbf{w}_{ij}$ as in Eq.21. Therefore, we define the KDBFM as following:

$$\sum\nolimits_{KDBFM} = \frac{1}{kC} \sum_{i=1}^{C} \sum_{j=1}^{k} w_{ij} w_{ij}^T . \quad (22)$$

From the Eq.20, we know that support vector has the excellent property, and we will redefine the within-class scatter matrix based on the support vector set calculated in SVM algorithm for a pair of data sets $(S_i, S_{ij}^{'})$ as following:

$$\mathbf{S}_w^{'}(i,j) = \sum_{i=1}^{c} p(\varpi_i) \sum_{m=1, C_m \in SV_i^j, |C_m|>1, x_m^{'} \in C_m}^{c} p(\varpi_m | \varpi_i) E((\Phi(x_m^{'}) - u_m^{'})(\Phi(x_m^{'}) - u_m^{'})^T) . \quad (23)$$

In the case that only samples of C_i are included in the positive set, for two-class SVM algorithm, we will explain the parameters used in Eq.23 as following. $SV_i^j, j = 1,...,k$ Includes all the support vectors after performing SVM on the pair of data sets $(S_i, S_{ij}^{'})$, which is divided into two sub-sets, one of which is SV_{i1}^j , whose elements are the support vectors belonging to the class C_i, and the other is SV_{i2}^j including all other support vectors. $u_i^{'}$ denotes the sample mean of the set SV_{i1}^j , and $p(\varpi_i)$ is the prior probability. $u_{i0}^{'}$ denotes the sample mean of SV_{i2}^j, whose samples come from different classes. The number of classes in SV_{i2}^j is n_i, and the center of each class is represented as the mean vector $u_{ih}^{'}, h = 1,..,n_i$ (if only one sample for one class is contained in SV_{i2}^{j}, then the sample is the class center). And now SV_i^j can be represented by a multi-center vector $(u_i^{'}, u_{i0}^{'}, u_{i1}^{'}, ..., u_{in_i}^{'})$, so we can know $u_m^{'} \in \{u_i^{'}, u_{i0}^{'}, u_{i1}^{'}, ..., u_{in_i}^{'}\}$. We will calculate the within-class scatter **sub-matrix** for the class, if more than one samples of which are contained in SV_i^j ($|C_m| > 1$). In our paper, the kernel function is polynomial style used in SVM and the proposed method, $k(x, y) = (\frac{x \cdot y}{|x| \cdot |y|} + 1)^r$, r is a constant integer.

Here, we will define the whole within-class scatter matrix based on all the support vector sets as following:

$$\mathbf{S}_w = \sum_{i=1}^{c} p(\varpi_i) \sum_{j=1}^{k} \frac{1}{k} \mathbf{S}_w^{'}(i, j) \quad . \tag{24}$$

To be concluded, in this part, we propose a method to construct KDBFM and the new kernelized within-class scatter matrix based on the support vector set, which is represented by a multi-center vector, $(u_i^{'}, u_{j0}^{'}, u_{j1}^{'}, ..., u_{jn_i}^{'})$.

4 Kernelized Decision Boundary Analysis

From the above discussion, we know that the eigenvectors of $\sum_{KDBFM}$ corresponding to non-zero eigenvalues are the necessary feature vectors. Moreover, we also hope that Decision Boundary Feature vector satisfies the Eq.20. Now, we show that our method is closely related to the Kernel Fisher method. If the $\sum_{KDBFM}$ is thought of as the between-class scatter matrix, which means that the trace of which is maximized by choosing the nonzero eigenvalues, and the proposed method will be consistent with the Fisher criterion, since the Eq.20 means minimizing the trace of the within-class scatter matrix. In the kernelized version of our method, $\mathbf{W}$ is also defined as Eq.5, and $\mathbf{W} \sum_{KDBFM} \mathbf{W}^T = \boldsymbol{\alpha} \mathbf{K}_b \boldsymbol{\alpha}^T$. $\mathbf{K}_w$ is the kernel within-class scatter matrix corresponding to the new within-class scatter matrix defined in Eq.24. Now, we will use a simple method to implement the KDBA.

4.1 The Algorithm of KDBA

The KDBA improves the generalization capability by decomposing its procedure into a simultaneous diagonalization of two matrices. We can refer to Nullspace Linear Discriminant Analysis (NLDA) method about the discriminant procedure [10]. The simultaneous diagonalization is stepwisely equivalent to two operations, and we first whiten $\mathbf{K}_t = \mathbf{K}_b + \mathbf{K}_w$ as following:

$$\mathbf{K}_t \boldsymbol{\Xi} = \boldsymbol{\Xi} \boldsymbol{\Gamma} \text{ and } \boldsymbol{\Xi}^T \boldsymbol{\Xi} = \mathbf{I}, \tag{25}$$

$$\boldsymbol{\Gamma}^{-1/2} \boldsymbol{\Xi}^T \mathbf{K}_t \boldsymbol{\Xi} \boldsymbol{\Gamma}^{-1/2} = \mathbf{I}, \tag{26}$$

where $\mathbf{\Xi}, \mathbf{\Gamma}$ are the eigenvector and the diagonal eigenvalue matrices of $\mathbf{K}_t$. We can get the eigenvectors matrix $\mathbf{\Xi}'$, whose eigenvalues are bigger than zero (Corresponding diagonal eigenvalue matrix is $\mathbf{\Gamma}'^{-1/2}$). The new within-class scatter matrix is computed by using the following method:

$$\mathbf{\Gamma}'^{-1/2}\mathbf{\Xi}'^{T}\mathbf{K}_w\mathbf{\Xi}'\mathbf{\Gamma}'^{-1/2} = \mathbf{\Xi}_w . \tag{27}$$

Diagonalizing now the new within-class scatter matrix $\mathbf{\Xi}_w$.

$$\mathbf{\Xi}_w\boldsymbol{\theta} = \boldsymbol{\theta}\boldsymbol{\gamma} \text{ and } \boldsymbol{\theta}^T\boldsymbol{\theta} = \mathbf{I} , \tag{28}$$

where $\boldsymbol{\theta}, \boldsymbol{\gamma}$ are the eigenvector and the diagonal eigenvalue matrices of $\mathbf{\Xi}_w$ in an increasing order. We remove the Eigenvectors, whose eigenvalues are far from zero, and the remained Eigenvectors construct the transformation matrix $\boldsymbol{\theta}'$.

The overall transformation matrix is now defined as following.

$$\boldsymbol{\alpha}' = \mathbf{\Xi}'\mathbf{\Gamma}'^{-1/2}\boldsymbol{\theta}' . \tag{29}$$

We use $\mathbf{w}'$ as the transform matrix, v is the extracted feature calculated by using Eq.30

$$v = \mathbf{w}'\Phi(x) = \sum_{i=1}^{n} \boldsymbol{\alpha}'_i k(x_i, x) . \tag{30}$$

4.2 Similarity Measure for KDBA

If v_1, v_2 are the feature vectors corresponding to two face images x_1, x_2, which are calculated by using the Eq.30, then the similarity rule is based on the cross correlation between the corresponding extracted feature vectors as following:

$$d(x_1, x_2) = \frac{v_1 . v_2}{\| v_1 \| . \| v_2 \|} . \tag{31}$$

The first experiment is tested on the Text classification problem. The other experiments are performed on the large CAS-PEAL database, and the comparative performance is carried out against the Eigenface, Fisherface and GKFD.

5 Experiment

5.1 Text Classification

We first make an experiment on the two-class Text classification problem. The dataset is the Example2(Details of the dataset can be referred to http://svmlight.joachims .org/), the training database of which contains 5 positive examples and 5 negative examples, and the test database contains 600 samples. In the case of k=1, KDBA achieves 84.5% accuracy rate, and the SVM-Light package is 84.3%.

Fig.3 shows the decision value calculated by SVM-Light, and the first feature found by KDBA for all 600 test examples.

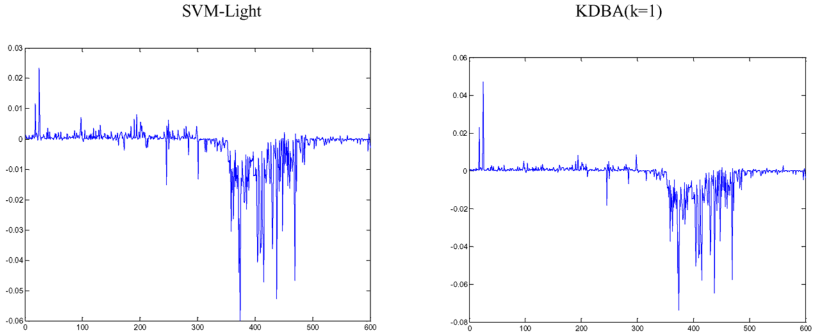

Fig. 3. Experiment on the Text classification problem

5.2 CAS-PEAL Face Database

In our experiments, the face image is cropped to size of 64X64 and overlapped with a mask to eliminate the background and hair. For all images concerned in the experiments, no preprocessing is exploited. To speed up the system, we first make PCA on the face images, and the lower dimensional vector in the PCA space is used in our experiments to capture the most expressive features of the original data. The CAS-PEAL face database was constructed under the sponsors of National Hi-Tech Program and ISVISION [14]. The goals to create the CAS-PEAL face database include: providing the worldwide researchers of FR community a large-scale face database for training and evaluating their algorithms; facilitating the development of FR by providing large-scale face images with different sources of variations, especially Pose, Expression, Accessories, and Lighting (PEAL); advancing the state-of-the-art face recognition technologies aiming at practical applications especially for the oriental. Currently, the CAS-PEAL face database contains 99,594 images of 1040 individuals (595 males and 445 females) with varying Pose, Expression, Accessory, and Lighting (PEAL). Gallery set contains one image for each person. One sample person was shown in Fig.4, and the size of the face image is 360X480. In this experiment, only one face image for each person was used as Gallery database. Details of the face database are shown at http://jdl.ac.cn [14].

Fig. 4. Sample of Face Images in CAS-PEAL database

Table 1. Experiment Result on CAS-PEAL database (Accurate rate)

	Eigenface	Fisherface*	GKFD	KDBA(k=1)	KDBA(k=5)	KDBA(k=10)
Accessory	37.1	61	58.7	62.4	63.1	64
Aging	50	72.7	77.3	84.8	87.9	87.9
Distance	74.2	93.5	94.9	96	96	96
Expression	53.7	71.3	78.2	78.2	78.5	79.9
Background	80.5	94.4	91.7	94.1	94.3	94.4

* Fisherface method refers to[15]

From above experiments, the KDBA method has achieved better performance than other popular face recognition schemes. The Kernelized Decision Boundary Feature vector can be thought of as the optimal direction in terms of SRM theory. We also know that, from the experiments, GKFD is just a little better than the Fisherface method, since the Fisherface method is based on the Enhanced Fisher Model[15], which can reserve the useful discriminant information by performing PCA on the within-class scatter matrix, and then modify between-class scatter matrix to get the whole transformation matrix. We also can know that the bigger k is used here, and a litter better performance of the face recognition system will be achieved. However, the complexity will also be increased accordingly, so we often choose a small value for k .

6 Conclusion and Future Work

We have proposed a novel discriminant analysis method named by KDBA. The contributions of our method include: (1) the proposed nonlinear discriminant approach, a kernelized extension to the Decision Boundary Analysis, is easily implemented and suitable to the large sample size problem by dividing the training database into several sub-datasets. (2) The KDBFM is constructed based on the normal vector of the optimal Decision Boundary (Decision Hyperplane of SVM) in terms of the SRM, and the PCs of which is the intrinsic discriminant subspace. (3) We also utilize a simple method to prove the property of the SVM, which is combined with KDBFM to make our method consistent with the Fisher Criterion. The feasibility of the new method has also been successfully tested on Text classification problem, and the face recognition task using data sets from the CAS-PEAL database, which is a very large one. The effectiveness of the method is shown in terms of accurate rate against some popular face recognition schemes, such as Eigenface, Fisherface, GKFD, and so on.

Gabor wavelet feature has been combined with some discriminant methods and successfully used in the face recognition problem [16, 17]. Therefore, we try to make full use of Gabor wavelet representation of face images before using KDBA to get the transformation matrix.

Acknowledgement

This research is partially sponsored by Natural Science Foundation of China under contract No.60332010, "100 Talents Program" of CAS, ShangHai Municipal Sciences and Technology Committee (No. 03DZ15013), and ISVISION Technologies Co., Ltd. Specially, thanks for Dr.Shiguang shan’s advice.

References

1. B. Scholkopf, A. Smola, K.R. Muller, “Nonlinear component analysis as a kernel eigenvalue problem,” Neural Computation, vol.10, pp.1299-1319, 1998.
2. S. Mika, G. Ratsch, J. Weston, B. Scholkopf and K.R. Muller, "Fisher discriminant analysis with kernels, " IEEE International Workshop on Neural Networks for Signal Processing, pp.41-48, 1999.

3. G. Baudat, F. Anouar, "Generalized discriminant analysis using a kernel approach, " Neural Computation, vol.12, no.10, pp.2385-2404, 2000.
4. Jian yang, Alejandro, "A new Kernel fisher discriminant algorithm with application to face recognition," Letters of Neurocomputinting, 2003.
5. Z. hong and J. Yang, "Optimal discriminant plane for a small number of samples and design method of classifier on the plane," Pattern recognition, vol.24, no.4, pp.317-324, 1991.
6. K. Fukunaga, J. M. Mantock, "nonparametric Discriminant Analysis," PAMI, Vol5(6), pp.671-678, 1983
7. Rik Fransens, Jan De Prins, "SVM-based Nonparametric Discriminant Analysis, An application to Face Detection," ICCV2003.
8. C.Lee, D.D. Langrebe, "Feature Extraction Based on Decision Boundaries," PAMI, Vol.15(4). Pp.388-400, 1993.
9. Qingshan Liu, Rui Huang, "Face Recognition Using Kernel Based Fisher Discriminant Analysis, " The 5th International Conference on Automatic Face and Gesture Recognition, pp.187-191, 2002.
10. Wei Liu, Yunhong Wang, "Null space based Kernel Fisher Discriminant analysis for face recognition," The 6th International Conference on Automatic Face and Gesture Recognition, 2004.
11. C.J.C. Burges, "A tutorial on support vector machines for pattern recognition," Knowledge Discovery and Data Mining, vol.2, pp.121-167, 1998.
12. G. Guo, S.Z. Li, and C. Kapluk. "Face recognition by support vector machines," Image and Vision Computing, vol.19, pp.631--638, 2001.
13. A. Tefas, C. Kotropoulos, and L. Pitas, "Using Support Vector Machines to Enhance the Perfornance of Elastic Graph Matching for Frontal Face Authentication, " IEEE Trans. on PAMI, Vol. 23, No. 7, pp.735-746, 2001.
14. Wen Gao, Bo Cao, Shiguang Shan, "The CAS-PEAL Large-Scale Face Database and Evaluation Protocols," Technical Report No. JDL_TR_04_FR_001, JDL, CAS, 2004.
15. Chenjun liu, Harry Wechsler, "Enhanced Fisher Linear Discriminant Models for face Recognition," 14th International Conference on Pattern Recognition, vol.2, pp.1368-1372, 1998.
16. Chengjun Liu and Harry Wechsler, "Gabor Feature Based Classification Using the Enhanced Fisher Linear Discriminant Model for Face Recognition, " IEEE Trans. Image Processing vol.11 no.4, pp.467-476, 2002.
17. Baochang Zhang, Wen Gao, Shiguang Shan, Yang Peng, "Discriminant Gaborfaces and Support Vector Machines Classifier for Face Recognition, " Asian Conference on Computer Vision, pp.37-42, 2004.
18. Stan Z. Li and Anil K. Jain (Eds). "Handbook of Face Recognition," Springer, to be published in 2004.