Kernel Uncorrelated Discriminant Analysis for Radar Target Recognition

Ling Wang, Liefeng Bo, and Licheng Jiao

Institute of Intelligent Information Processing
710071, Xidian University, Xi'an, China
{wliiip, blf0218}@163.com

Abstract. Kernel fisher discriminant analysis (KFDA) has received extensive study in recent years as a dimensionality reduction technique. KFDA always encounters an intrinsic singularity of scatter matrices in the feature space, namely 'small sample size' (SSS) problem. Several novel methods have been proposed to cope with this problem. In this paper, kernel uncorrelated discriminant analysis (KUDA) is proposed, which not only can bear on the SSS problem but also extract uncorrelated features, a desirable property for many applications. And then, we have conducted a comparative study on the application of KUDA and other variants of KFDA in radar target recognition problem. The experimental results indicate the effectiveness of KUDA and illustrate the utility of KFDA on the problem.

1 Introduction

Radar target recognition is a difficulty of task in pattern recognition due to the complex movement of radar target, including transformation and rotation. Particularly for military application, the target is so incooperative that the samples data is much insufficient and noisy. A very simple and rapid approach for recognizing radar target is through the use of radar range profiles which are essentially one-dimension radar images. Due to the high dimensionality of range profiles, it is necessary to perform feature extraction at first to reduce the dimensionality and then perform classification for recognition.

Linear discriminant analysis (LDA), also called fisher discriminant analysis is a widely-used statistical method for feature extraction and dimension reduction, which has been successfully applied in many problems such as face recognition. Because of the nature of linearity, LDA is inadequate to describe the complexity in real world problems. The nonlinearly clustered structure is not easily captured by LDA. In recent years, a category of nonlinear algorithms using the so-called kernel trick have aroused considerable interest in the fields of pattern recognition and machine learning [1]. Generalization of LDA for solving nonlinear problems based on kernel trick has become an active research area. A group of kernel-based fisher discriminant analysis (KFDA) algorithms has been proposed [2]. Extensive empirical comparisons have shown that KFDA works as well as other kernel based classifiers. However, because

of the implicit high-dimensional nonlinear mapping, the so-called "small sample size" (SSS) problem is very common in the feature space.

Several techniques that might alleviate this problem have been proposed. Mika et al. used the regularization technique to make the inner product matrix nonsingular [3]. But his method was developed to handle binary classification only. Following that, Baudat and Anouar developed a GDA for multiple classification [4]. Yang et al. performed LDA in KPCA feature space to deal with the problem [5]. Recently, Park et al. proposed a kernel based disciminant analysis based on the generalized singular value decomposition called KDA/GSVD, which works regardless of the nonsingularity of the scatter matrices in either the input space or feature space [6].

For feature extraction, the uncorrelated attributes with minimum redundancy are highly desirable. Jin et al. proposed uncorrelated LDA (ULDA) for extracting feature vectors with uncorrelated attributes [7]. However, the proposed method has two limitations, i.e. the expensive computation of the d generalized eigenvalue problems, where d is number of optimal discriminant vectors by ULDA, and the non-applicability to the SSS problem as the classical LDA. To overcome these limitations, Ye et al. presented an efficient algorithm to compute the optimal discriminant vectors of ULDA and at the same time addressed the SSS problem of ULDA [8]. In [9], the optimization criteria of classical LDA was extended to solve the SSS problem, and the solutions to the proposed criterion form a family of algorithms to which ULDA and a novel algorithm, namely orthogonal LDA (OLDA) belong.

In this paper, we present the nonlinear extension of ULDA based on kernel trick, called KUDA, which can work regardless of the SSS problem. We also investigate the application of KUDA and some KFDA variants in radar target recognition problem. Through the experiments, we not only demonstrate that KUDA is an effective nonlinear dimension reduction approach, but also conclude that all the KFDA variants achieve higher classification accuracy on radar target recognition problem compared with classical LDA. Another surprisingly observation is that a special kernel function, Cauchy kernel, has a remarkable performance on the problem.

2   Related Work on Kernel Fisher Discriminant Analysis

Classical fisher discriminant analysis aims to find the optimal transformation, which maximizes the between-class scatter matrix while minimizing the within-class scatter matrix simultaneously. Thus, the cluster structure of the original high-dimensional space is preserved in the reduced-dimensional space. But this method fails for a nonlinear problem. There have been extensive researches in nonlinear discriminant analysis using kernel function, called by a joined name kernel fisher discriminant analysis (KFDA). Due to the nonlinear map by a kernel function, the dimension of the feature space often becomes much larger than that of the original data space, and as a result, the scatter matrices become singular, which is referred to as "small sample size" (SSS) problem. In the following, we will review some recent proposed KFDA algorithms, all of which attempt to deal with the SSS problem in the feature space.

KPCA plus LDA. PCA plus LDA, a two stage approach, is a popular technique for face recognition [5]. In Euclidean space, the theoretical foundation of why LDA can be performed in the PCA transformed space has been given in [10]. Since real-world problems are always turned into SSS problems by a nonlinear mapping, we can generalize the result directly to the data in a mapped feature space. At first stage, PCA is performed in the feature space. It is equivalent to performing KPCA in the input space. And then, in the KPCA transformed space, LDA is performed.

The biggest challenge in using KPCA plus LDA is that it is difficult to choose an optimal reduced dimension m. If m is chosen large, the eigenvalue problem in the discriminant stage will be expensive and unstable because of the high dimensionality. If too small, it may not provide sufficient discriminant information.

GDA. Generalized discriminant analysis (GDA) is proposed for multiclass classification. As such for LDA, the purpose of GDA method is to maximize the between class scatter matrix while minimizing the within class scatter matrix in the feature space. In order to cope with the singularity of scatter matrices in the feature space, the eigenvectors decomposition of the kernel matrix is employed, and the singularity is avoided by removing some small eigenvalues. As KPCA plus LDA, it is difficult to determine the magnitude of eigenvalue that should be removed.

KDA/GSVD. A recent work on overcoming SSS problem in LDA lies in the use of Generalized Singular Value Decomposition (GSVD), named LDA/GSVD [11]. The method avoids inversing the within-class scatter matrix, so it computes the solution exactly without losing any information. Recently, Park presented the nonlinear extension of LDA based on kernel functions and the GSVD, named KDA/GSVD. The GSVD is employed to solve the generalized eigenvalue problem which is formulated in the feature space defined by a nonlinear mapping through kernel functions. The adventage of KDA/GSVD is that it can be applied regardless of singularity of the scatter matrics both in the original space and in the feature space. The detailed derivation can be found in [6].

3   Kernel Uncorrelated Discriminant Analysis

Uncorrelated linear discriminant analysis (ULDA) [7] was proposed for feature extraction. The feature vectors transformed by ULDA were shown to be statistically uncorrelated, which is a desirable property for many applications. ULDA aims to find the optimal discriminant vectors that are S_t-orthogonal (Two vectors x and y are S_t-orthogonal, if $x^T S_t y = 0$). In this section, we present a nonlinear extension of ULDA based on kernel functions, and solve it using the technique of simultaneous diagonalization of the three scatter matrices [9].

Let n denotes the dimension of the original sample space, and r is the number of classes. And let $X = \{\mathbf{x}_1, \mathbf{x}_2, \cdots, \mathbf{x}_l\}$ be the training samples set, where $\mathbf{x}_i \in X \subset R^n$. For a given nonlinear mapping ϕ, the input data space R^n can be mapped into the

feature space $F : \phi : R^n \to F$. As a result, a sample in the original input space R^n is mapped into a potentially much higher dimensional feature vector: $\mathbf{x} \to \phi(\mathbf{x})$ in the feature space F . To avoid computing the dot products in a high-dimensional feature space, kernel trick is introduced to facilitate the computation. A kernel is defined by an inner product $k(\mathbf{x}_i, \mathbf{x}_j) = \left(\phi(\mathbf{x}_i) \bullet \phi(\mathbf{x}_j) \right)$.

Let $\mathbf{K} = \left[k(\mathbf{x}_i, \mathbf{x}_j) \right]_{(1 \le i \le l, 1 \le j \le l)}$ be the kernel matrix. Then, we can consider each column in $\mathbf{K}$ as a data point in the n –dimensional space. As in the LDA, we define between-class scatter matrix and within-class scatter and total scatter matrix in the feature space as below:

$$\mathbf{S}_b^F = \mathbf{K}_b \mathbf{K}_b^T , \quad \mathbf{S}_w^F = \mathbf{K}_w \mathbf{K}_w^T , \quad \mathbf{S}_t^F = \mathbf{K}_t \mathbf{K}_t^T , \tag{1}$$

where

$$\mathbf{K}_b = [b_{ij}]_{(1 \le i \le n, 1 \le j \le r)}, \quad b_{ij} = \sqrt{n_j} \left(\frac{1}{n_j} \sum_{s \in N_j} k(\mathbf{x}_i, \mathbf{x}_s) - \frac{1}{n} \sum_{s=1}^{n} k(\mathbf{x}_i, \mathbf{x}_s) \right)$$

$$\mathbf{K}_w = [w_{ij}]_{(1 \le i \le n, 1 \le j \le n)}, \quad w_{ij} = k(\mathbf{x}_i, \mathbf{x}_j) - \frac{1}{n_l} \sum_{s \in N_l} k(\mathbf{x}_i, \mathbf{x}_s), \ x_j \in \text{class } l . \tag{2}$$

$$\mathbf{K}_t = [t_{ij}]_{(1 \le i \le n, 1 \le j \le n)}, \quad t_{ij} = k(\mathbf{x}_i, \mathbf{x}_j) - \frac{1}{n} \sum_{s=1}^{n} k(\mathbf{x}_i, \mathbf{x}_s)$$

According to the definition of S_t -orthogonal discriminant vector, we can define a trace optimization problem in the feature space as follows:

$$\mathbf{G} = \arg \max_{\mathbf{G} \in \mathbb{R}^{p \times t} : \mathbf{G}^T \mathbf{S}_t^F \mathbf{G} = I_t} \left(trace(\mathbf{G}^T \mathbf{S}_w^F \mathbf{G})^{-1} \mathbf{G}^T \mathbf{S}_b^F \mathbf{G} \right). \tag{3}$$

Since $\mathbf{S}_t^F = \mathbf{S}_w^F + \mathbf{S}_b^F$, the problem above is equivalent to

$$\mathbf{G} = \arg \max_{\mathbf{G} \in \mathbb{R}^{p \times t} : \mathbf{G}^T \mathbf{S}_t^F \mathbf{G} = I_t} \left(trace(\mathbf{G}^T \mathbf{S}_t^F \mathbf{G})^{-1} \mathbf{G}^T \mathbf{S}_b^F \mathbf{G} \right). \tag{4}$$

Note that $\mathbf{S}_t^F$ and $\mathbf{S}_b^F$ are both singular. In order to solve the problem, a natural extension is that the inverse of a matrix is replaced by the pseudo-inverse [12]:

$$\mathbf{G} = \arg \max_{\mathbf{G} \in \mathbb{R}^{p \times t} : \mathbf{G}^T \mathbf{S}_t^F \mathbf{G} = I_t} \left(trace(\mathbf{G}^T \mathbf{S}_t^F \mathbf{G})^{+} \mathbf{G}^T \mathbf{S}_b^F \mathbf{G} \right). \tag{5}$$

The above optimization problem can be solved by diagonalizing the three scatter matrices $\mathbf{S}_b^F$, $\mathbf{S}_w^F$, and $\mathbf{S}_t^F$ simultaneously.

Let $\mathbf{K}_t = \mathbf{U} \Sigma \mathbf{V}^T$ be the SVD of $\mathbf{K}_t$, where $\mathbf{K}_t$ is defined in (2), $\mathbf{U}$ and $\mathbf{V}$ are orthogonal, $\Sigma = \begin{pmatrix} \Sigma_t & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix}$, $\Sigma_t \in \mathbb{R}^{t \times t}$ is diagonal, and $t = rank(\mathbf{S}_t^F)$. Then, we have

$$\mathbf{S}_t^F = \mathbf{K}_t\mathbf{K}_t^T = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^T\mathbf{V}\mathbf{\Sigma}^T\mathbf{U}^T = \mathbf{U}\mathbf{\Sigma}\mathbf{\Sigma}^T\mathbf{U}^T = \mathbf{U}\begin{pmatrix} \mathbf{\Sigma}_t^2 & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix}\mathbf{U}^T . \tag{6}$$

Let $\mathbf{U} = (\mathbf{U}_1, \mathbf{U}_2)$ be a partition of $\mathbf{U}$, such that $\mathbf{U}_1 \in \mathbb{R}^{n\times t}, \mathbf{U}_2 \in \mathbb{R}^{n\times(n-t)}$. (6) can be rewritten as

$$\begin{aligned}
\begin{pmatrix} \mathbf{\Sigma}_t^2 & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix} &= \mathbf{U}^T(\mathbf{S}_b^F + \mathbf{S}_w^F)\mathbf{U} \\
&= \begin{pmatrix} \mathbf{U}_1^T \\ \mathbf{U}_2^T \end{pmatrix}\mathbf{S}_b^F(\mathbf{U}_1, \mathbf{U}_2) + \begin{pmatrix} \mathbf{U}_1^T \\ \mathbf{U}_2^T \end{pmatrix}\mathbf{S}_w^F(\mathbf{U}_1, \mathbf{U}_2) \\
&= \begin{pmatrix} \mathbf{U}_1^T\mathbf{S}_b^F\mathbf{U}_1 & \mathbf{U}_1^T\mathbf{S}_b^F\mathbf{U}_2 \\ \mathbf{U}_2^T\mathbf{S}_b^F\mathbf{U}_1 & \mathbf{U}_2^T\mathbf{S}_b^F\mathbf{U}_2 \end{pmatrix} + \begin{pmatrix} \mathbf{U}_1^T\mathbf{S}_w^F\mathbf{U}_1 & \mathbf{U}_1^T\mathbf{S}_w^F\mathbf{U}_2 \\ \mathbf{U}_2^T\mathbf{S}_w^F\mathbf{U}_1 & \mathbf{U}_2^T\mathbf{S}_w^F\mathbf{U}_2 \end{pmatrix}
\end{aligned} \tag{7}$$

Since both $\mathbf{S}_b^F$ and $\mathbf{S}_w^F$ are positive semidefinite, we thus have

$$\mathbf{U}^T\mathbf{S}_b^F\mathbf{U} = \begin{pmatrix} \mathbf{U}_1^T\mathbf{S}_b^F\mathbf{U}_1 & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix}, \; \mathbf{U}^T\mathbf{S}_w^F\mathbf{U} = \begin{pmatrix} \mathbf{U}_1^T\mathbf{S}_w^F\mathbf{U}_1 & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix}. \tag{8}$$

According to (7) and (8), we can derive the following equation

$$\mathbf{I}_t = \mathbf{\Sigma}_t^{-1}\mathbf{U}_1^T\mathbf{S}_b^F\mathbf{U}_1\mathbf{\Sigma}_t^{-1} + \mathbf{\Sigma}_t^{-1}\mathbf{U}_1^T\mathbf{S}_w^F\mathbf{U}_1\mathbf{\Sigma}_t^{-1} . \tag{9}$$

Denote $\mathbf{B} = \mathbf{\Sigma}_t^{-1}\mathbf{U}_1^T\mathbf{K}_b$ and let $\mathbf{B} = \mathbf{P}\tilde{\mathbf{\Sigma}}\mathbf{Q}^T$ be the SVD of $\mathbf{B}$. Then, we get

$$\mathbf{\Sigma}_t^{-1}\mathbf{U}_1^T\mathbf{S}_b^F\mathbf{U}_1\mathbf{\Sigma}_t^{-1} = \mathbf{P}\tilde{\mathbf{\Sigma}}^2\mathbf{P}^T = \mathbf{P}\mathbf{\Sigma}_b\mathbf{P}^T , \tag{10}$$

where $\mathbf{\Sigma}_b \equiv \tilde{\mathbf{\Sigma}}^2 = diag(\lambda_1, \cdots, \lambda_t)$, $\lambda_1 \geq \cdots \geq \lambda_q > 0 = \lambda_{q+1} = \cdots = \lambda_t$, and $q = rank(\mathbf{S}_b^F)$. It follows from (9) that

$$\mathbf{P}^T\mathbf{\Sigma}_t^{-1}\mathbf{U}_1^T\mathbf{S}_w^F\mathbf{U}_1\mathbf{\Sigma}_t^{-1}\mathbf{P} = \mathbf{I}_t - \mathbf{\Sigma}_b \equiv \mathbf{\Sigma}_w . \tag{11}$$

According to (9), (10), and (11), $\mathbf{S}_b^F$, $\mathbf{S}_w^F$ and $\mathbf{S}_t^F$ can be diagonalized as

$$\mathbf{X}^T\mathbf{S}_b^F\mathbf{X} = \begin{pmatrix} \mathbf{\Sigma}_b & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix} \equiv \mathbf{D}_b, \; \mathbf{X}^T\mathbf{S}_w^F\mathbf{X} = \begin{pmatrix} \mathbf{\Sigma}_w & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix} \equiv \mathbf{D}_w, \; \mathbf{X}^T\mathbf{S}_t^F\mathbf{X} = \begin{pmatrix} \mathbf{I}_t & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix} \equiv \mathbf{D}_t \tag{12}$$

where $\mathbf{X} = \mathbf{U}\begin{pmatrix} \mathbf{\Sigma}_t^{-1}\mathbf{P} & \mathbf{0} \\ \mathbf{0} & \mathbf{I} \end{pmatrix}$.

Let $\mathbf{X}_q$ be the matrix consisting of the first q columns of $\mathbf{X}$, where $q = rank(\mathbf{S}_b^F)$. $\mathbf{G}^F = \mathbf{X}_q$ is the solution to the optimization problem (5) [9]. Consequently, kernel uncorrelated discriminant analysis (KUDA) algorithm can be described as the following.

Algorithm. KUDA

Given a data matrix $X = [\mathbf{x}_1, \cdots, \mathbf{x}_l] \in \mathbb{R}^{n \times l}$ with r classes and a kernel function k

1. Compute $\mathbf{K}_b \in \mathbb{R}^{n \times r}$, $\mathbf{K}_w \in \mathbb{R}^{n \times n}$, and $\mathbf{K}_t \in \mathbb{R}^{n \times n}$ as in (2);
2. Compute the reduced SVD of $\mathbf{K}_t$ as $\mathbf{K}_t = \mathbf{U}_1 \mathbf{\Sigma}_t \mathbf{V}_1^T$;
3. $\mathbf{B} = \mathbf{\Sigma}_t^{-1} \mathbf{U}_1^T \mathbf{K}_b$;
4. Compute SVD of $\mathbf{B}$ as $\mathbf{B} = \mathbf{P} \tilde{\mathbf{\Sigma}} \mathbf{Q}^T$; $q = rank(\mathbf{B})$;
5. $\mathbf{X} = \mathbf{U}_1 \mathbf{\Sigma}_t^{-1} \mathbf{P}$;
6. $\mathbf{G}^F = \mathbf{X}_q$;

---

4   Performance Comparison on Radar Target Recognition

Radar target recognition refers to the detection and recognition of target signatures using high-resolution range profiles, in our case, in inverse synthetic aperture radar. A radar image represents a spatial distribution of microwave reflectivity that is sufficient to characterize the illuminated target. Range resolution allows the sorting of reflected signals on the basis of range. When range-gating or time-delay sorting is used to inter-rogate the entire range extent of the target space, a one-dimensional image, called a range profile, will be generated.

Our task is to recognize the range profile of the three different plane models, i.e. J-6, J-7 and B-52, based on experimental data acquired in a microwave anechoic chamber. The dimensionality of the range profiles is 64. The full data set is split into 363 training samples and 721 test samples. Training samples consist of 104 1-dimensional images of J-6, 151 1-dimensional images of J-7 and 108 1-dimensional images of B-52. Test samples consist of 207 1-dimension images of J-6, 300 1-dimension images of J-7 and 214 1-dimension images of B-52.

Table 1. Number of misclassification of several classifiers

Method	Recognition Rate	Error Number
LDA	94.73	38
GDA	98.61	10
KPCA-LDA	99.69	2
KDA/GSVD	99.71	2
KUDA	99.86	1

A simple classifier, k-nearest neighbor (KNN), is employed to evaluate the quality of different dimension reduction algorithms. Leave-one-out error is used to find the best number of neighbor k. The experimental results for several methods using an optimal kernel function are summarized in Table 1. For KPCA-LDA, we find the

optimal value of the principle components on an interval. From Table 1, we can see that only one wrong recognition occur in KUDA, and only 2 in KDA/GSVD and KPCA-LDA. This indicates that these algorithms proposed to bear on SSS problem are superior to LDA, GDA, and have similar high performance on the radar target recognition problem.

After performing discriminant analysis, the dimensionality of range files is reduced to 2 because the class number is three. Therefore, these real world data can be visualized in Figure 2. From the projection image of low dimension, we can see that LDA is not good enough because of the intrinsic nonlinearity for the problem, and on the contrary, the variants of kernel based discriminant analysis preserve the information for classification well.
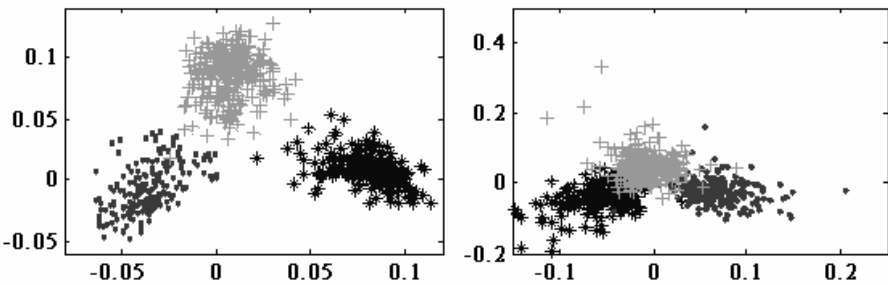

Fig. 2. 2-dimensional visualization of the radar range profiles with kernel (left) and without kernel (right)

Table 2. Performance of variants of kernel discriminant analysis with different kernels

Method	RBF	Coswave	Cauchy
GDA	98.61	98.20	98.61
KPCA-LDA	97.23	97.09	99.69
KDA/GSVD	97.45	97.05	99.71
KUDA	97.23	97.09	99.86

We also compare the performance of variants of kernel fisher discriminant analysis on three popular kernels, i.e. Gaussian RBF kernel $k(\mathbf{x}, \mathbf{y}) = \exp(\dfrac{-\|\mathbf{x}-\mathbf{y}\|^2}{2p^2})$, Coswave kernel

$k(\mathbf{x}, \mathbf{y}) = \dfrac{p}{p+\|\mathbf{x}-\mathbf{y}\|^2}$ and Cauchy kernel $k(\mathbf{x}, \mathbf{y}) = \cos(1.75 \times \dfrac{(\mathbf{x}-\mathbf{y})}{p})\exp(-\dfrac{\|\mathbf{x}-\mathbf{y}\|^2}{2p^2})$,

where $p \in R$. The results are summarized in Table 2. From the experimental results, we find unexpectedly that the Cauchy kernel has a predominant performance on the problem.

5 Conclusion

In this paper, we propose a new kernel fisher discriminant analysis, namely KUDA to deal with the SSS problem in the feature space. And then, we describe the application of KUDA and some other KFDA variants in radar target recognition problem. Experiment results have shown that KUDA and the KFDA variants developed for solving the SSS problem perform significantly better than the classical LDA. Furthermore, it is worth to mention that a specific kernel, i.e. Cauchy kernel, performs best on the problem. These observations are expected to be useful when we attempt to apply kernel discriminant analysis to other target recognition problems.

References

1. Bo, L.F., Wang, L., and Jiao, L.C.: Training support vector machines using greedy stage-wise algorithm. Lecture Notes in Computer Science (PAKDD'05) 3518 (2005) 632-638
2. Bo, L.F., Wang, L., and Jiao, L.C.: Feature scaling for kernel fisher discriminant analysis using leave-one-out cross validation. Neural Computation 18(4) (2006) 961-978
3. Mika, S., Ratsch, G., and Weston, J.: Fisher discriminant analysis with kernels. In Proceedings of the IEEE Workshop on Neural Networks for signal Processing (1999) 41-48
4. Baudat, G. and Anouar, F.: Generalized discriminant analysis using a kernel approach. Nerual Compuation 12(10) (2000) 2385-2404
5. Yang, M.H.: Kernel eigenfaces vs. kernel fisherfaces: face recognition using kernel methods. In Proceedings of Fifth IEEE International Conference Automatic Face and Gesture Recognition (2002) 215-220
6. Park, C.H. and Park, H: Nonlinear discriminant analysis using kernel functions and the generalized singular value decomposition, SIAM Journal on Matrix Analysis and Applications, to appear
7. Jin, Z., Yang, J.Y., Tang, Z.M., and Hu, Z.S.: A theorem on the uncorrelated optimal discriminant vectors. Pattern Recognition 34 (2001) 2041-2047
8. Ye, J.P, Janardan, R., Li, Q., and Park, H.: Feature extraction via generalized uncorrelated linear discriminant analysis. In Proceedings of the 21st International Conference on Machine Learning, Banff, Canada 2004
9. Ye, J.P.: Characterization of a family of algorithms for generalized discriminant analysis on undersampled problems. Journal of Machine Learning Research 6(4) (2005) 483-502
10. Yang, J., Yang, J.Y.: Why can LDA be performed in PCA transformed space? Pattern Recognition 36 (2003) 563-566
11. Howland, P. and Park, H.: Generalizing Discriminant Analysis Using the Generalized Singular Value Decomposition. IEEE Transactions on PAMI 26(8) (2004) 995-1006
12. Ye, J.P., Janardan, R., Park, C.H., and Park, H.: An optimization criterion for generalized discriminant analysis on undersampled problems. IEEE Transactions on PAMI 26(8) (2004) 982-994