A DIRECT METHOD TO SOLVE THE BIASED DISCRIMINANT ANALYSIS IN KERNEL FEATURE SPACE FOR CONTENT BASED IMAGE RETRIEVAL

Dacheng Tao and Xiaoou Tang

{dctao2, xtang}@ie.cuhk.edu.hk

Department of Information Engineering, The Chinese University of Hong Kong, Shatin, Hong Kong

ABSTRACT

In recent years, relevance feedback has been widely used to improve the performance of content-based image retrieval. How to select a subset of features from a large-scale feature pool and to construct a suitable dissimilarity measure are key steps in a relevance feedback system. Biased discriminant analysis has been proposed to select features during relevance feedback iterations. However, to solve the BDA, we often encounter the matrix singular problem. In this paper, we propose a kernel-based discriminant analysis, which can overcome the matrix singular problem. The new method is shown to outperform the traditional kernel BDA and constrained support vector machine based relevance feedback algorithms.

1. INTRODUCTION

With the explosive increase of the amount of image records on the Internet and the rapid advance of computer technology, retrieving images from a large-scale image database based on visual features become one of the most active research fields [1-3] in multimedia information processing. In recent years, relevance feedback (RF) [4-6] has been shown to be an efficient technique to improve the performance of content-based image retrieval (CBIR) system.

Kernel biased discriminant analysis (KBDA) [7] has been used for RF mainly because it handles the positive and negative feedbacks separately. However, KBDA often suffers the matrix singular problem (MSP). To avoid MSP, the regularization method [8] adds small quantities to the diagonal of the scatter matrices. This apparently is not an optimal solution and sometimes it may lead to an ill-posed problem, which limits the performance of RF.

Recently, direct linear discriminant analysis (DLDA) [9] was proposed to solve MSP in face recognition. DLDA discards the null space of between-class scatter matrix, which does not contain much discriminant information. Then the discriminant vectors are the within-class scatter matrix's eigenvectors with smallest eigenvalues. The successes of the kernel-machine [10] based pattern classification algorithms have motivated us to generalize the idea of DLDA to BDA in the kernel feature space. We first project all the training samples from the input feature space to kernel feature space, and then the null-space of "the negative scatter with respect to positive centroid" matrix is removed. At last, the discriminant vectors are extracted as "the positive within class scatter" matrix's eigenvectors with the smallest eigenvalues.

2. REVIEW OF KBDA AND DLDA

The proposed algorithm combines the merits of KBDA and DLDA. We briefly review them first.

2.1. Kernel Biased Discriminant Analysis

KBDA [7] tries to find the subspace to discriminate the positive and negative samples in the kernel space. It is spanned by a set of vectors $\mathbf{W}=\{w_k\}_{k=1}^{m}$, which maximize the ratio between the positive covariance matrix $\mathbf{S}_x^{\phi}$ and the biased matrix $\mathbf{S}_y^{\phi}$. $\mathbf{W}$ can be obtained by solving the following general eigenvalue problem:

$$\mathbf{W} = \arg\max_{\mathbf{W}} \frac{\left\|\mathbf{W}^T \mathbf{S}_y^{\phi} \mathbf{W}\right\|}{\left\|\mathbf{W}^T \mathbf{S}_x^{\phi} \mathbf{W}\right\|}. \tag{1}$$

$\mathbf{S}_x^{\phi}$ and $\mathbf{S}_y^{\phi}$ are defined as:

$$\begin{cases} \mathbf{S}_x^{\phi} = \sum_{i=1}^{Nx}\left(\varphi(\mathbf{x}_i)-\bar{\varphi}(\mathbf{x})\right)\left(\varphi(\mathbf{x}_i)-\bar{\varphi}(\mathbf{x})\right)^T = \Phi_x \Phi_x^T \\ \mathbf{S}_y^{\phi} = \sum_{i=1}^{Ny}\left(\varphi(\mathbf{y}_i)-\bar{\varphi}(\mathbf{x})\right)\left(\varphi(\mathbf{y}_i)-\bar{\varphi}(\mathbf{x})\right)^T = \Phi_y \Phi_y^T \end{cases} \tag{2}$$

where $\bar{\varphi}(\mathbf{x})=\frac{1}{N_x}\sum_{i=1}^{N_x}\varphi(\mathbf{x}_i)$ is the centroid of the positive samples in kernel space, $\mathbf{x}_i$ and $\mathbf{y}_i$ are the positive and negative feedback samples respectively, N_x and N_y are the number of positive and negative feedback samples respectively, and φ is the kernel mapping function [10].

2.2. Direct Linear Discriminant Analysis

Linear discriminant analysis (LDA) [8] tries to find the subspace $\mathbf{W}$, which maximizes the ratio between the between-class scatter matrix $\mathbf{S}_b$ and the within-class scatter matrix $\mathbf{S}_w$,

$$\mathbf{W} = \arg\max_{\mathbf{W}} \frac{\left\| \mathbf{W}^{T}\mathbf{S}_{b}\mathbf{W} \right\|}{\left\| \mathbf{W}^{T}\mathbf{S}_{w}\mathbf{W} \right\|}. \tag{3}$$

Let the training set contains c individual classes and each class C_i has N_i samples. Then $\mathbf{S}_w$ and $\mathbf{S}_b$ are defined as,

$$\begin{cases} \mathbf{S}_b = \frac{1}{N}\sum_{i=1}^{c} N_i (\mathbf{m}_i - \mathbf{m})(\mathbf{m}_i - \mathbf{m})^T \\ \mathbf{S}_w = \frac{1}{N}\sum_{i=1}^{c}\sum_{j=1}^{N_i} (\mathbf{x}_j^i - \mathbf{m}_i)(\mathbf{x}_j^i - \mathbf{m}_i)^T, \mathbf{x}_j^i \in C^i, \end{cases} \tag{4}$$

where $\mathbf{m} = \frac{1}{N}\sum_{j=1}^{N}\mathbf{x}_j$ is the mean vector of the whole training set, $\mathbf{m}_i = \frac{1}{N_i}\sum_{j=1}^{N_i}\mathbf{x}_j^i$ is the mean vector for the individual class C_i, and $\mathbf{x}_j^i$ is the sample belonging to class C_i. $\mathbf{W}$ can be computed from the eigenvectors of $\mathbf{S}_w^{-1}\mathbf{S}_b$.

To avoid MSP in LDA, DLDA and its kernel version were proposed in [9] and [11]. It accepts high-dimensional data as the input, and optimizes LDA directly, without any dimension reduction step, so it takes all the information within and outside of the null space of $\mathbf{S}_w$. In DLDA, $\mathbf{S}_b$ is first diagonalized, and the null space of $\mathbf{S}_b$ is removed,

$$\mathbf{Y}^T\mathbf{S}_b\mathbf{Y} = \mathbf{D}_b > 0 \tag{5}$$

where $\mathbf{Y}$ are eigenvectors and $\mathbf{D}_b$ are the corresponding non-zero eigenvalues of $\mathbf{S}_b$. $\mathbf{S}_w$ is transformed to

$$\mathbf{K}_w = \mathbf{D}_b^{-1/2}\mathbf{Y}^T\mathbf{S}_w\mathbf{Y}\mathbf{D}_b^{-1/2} \tag{6}$$

$\mathbf{K}_w$ is diagonalized by eigenanalysis,

$$\mathbf{U}^T\mathbf{K}_w\mathbf{U} = \mathbf{D}_w. \tag{7}$$

The LDA transformation matrix for classification is then defined as,

$$\mathbf{W} = \mathbf{Y}\mathbf{D}_b^{-1/2}\mathbf{U}\mathbf{D}_w^{-1/2}. \tag{8}$$

3. KERNEL DIRECT BDA

Before we derive the kernel direct biased discriminant analysis (KDBDA), we first introduce the kernel matrix $\mathbf{K}$, because all the derived formulas are related to $\mathbf{K}$,

$$\mathbf{K} = \begin{bmatrix} \mathbf{K}_{xx} & \mathbf{K}_{xy} \\ \mathbf{K}_{yx} & \mathbf{K}_{yy} \end{bmatrix} \tag{9}$$

where $\mathbf{K}_{xx} = \left[k(\mathbf{x}_i, \mathbf{x}_j)\right]_{\substack{1\le i\le Nx \\ 1\le j\le Nx}}$, $\mathbf{K}_{xy} = \mathbf{K}_{yx}^T = \left[k(\mathbf{x}_i, \mathbf{y}_j)\right]_{\substack{1\le i\le Nx \\ 1\le j\le Ny}}$, $\mathbf{K}_{yy} = \left[k(\mathbf{y}_i, \mathbf{y}_j)\right]_{\substack{1\le i\le Ny \\ 1\le j\le Ny}}$, and $k(.)$ is the kernel function.

Just like DLDA, we begin the KDBDA with the analysis of $\mathbf{S}_y^{\phi}$. Because the dimension of Φ_y can be infinite, it is impossible to calculate $\mathbf{S}_y^{\phi}$ and perform eigenanalysis to $\mathbf{S}_y^{\phi}$ directly. However, it can be avoid according to:

$$\begin{aligned} &\Phi_y^T\Phi_y\mathbf{e}_i = \lambda_i\mathbf{e}_i \Rightarrow \Phi_y\Phi_y^T\left(\Phi_y\mathbf{e}_i\right) = \lambda_i\left(\Phi_y\mathbf{e}_i\right) \\ &\Phi_y\Phi_y^T\mathbf{u}_i = \lambda_i\mathbf{u}_i \Rightarrow \mathbf{u}_i = \Phi_y\mathbf{e}_i \\ &\therefore \mathbf{U} = \Phi_y\mathbf{E}, \end{aligned} \tag{10}$$

where $\mathbf{U}$ is the none-zero subspace (not normalized) of $\Phi_y\Phi_y^T$. The dimension of $\Phi_y^T\Phi_y$ is N_y, and $\Phi_y^T\Phi_y$ is:

$$\Phi_y^T\Phi_y = \left[\phi^T(\mathbf{y}_i)\phi(\mathbf{y}_j) - \phi^T(\mathbf{y}_i)\bar{\phi}(\mathbf{x}) - \bar{\phi}^T(\mathbf{x})\phi(\mathbf{y}_j) + \bar{\phi}^T(\mathbf{x})\bar{\phi}(\mathbf{x})\right]_{\substack{1\le i\le Ny \\ 1\le j\le Ny}}$$

Then we need to calculate $\phi^T(\mathbf{y}_i)\phi(\mathbf{y}_j)$, $\phi^T(\mathbf{y}_i)\bar{\phi}(\mathbf{x})$, $\bar{\phi}^T(\mathbf{x})\phi(\mathbf{y}_j)$, and $\bar{\phi}^T(\mathbf{x})\bar{\phi}(\mathbf{x})$.

$$\bar{\phi}^T(\mathbf{x})\bar{\phi}(\mathbf{x}) = \left(\frac{1}{N_x}\sum_{m=1}^{N_x}\phi(\mathbf{x}_m)\right)^T\left(\frac{1}{N_x}\sum_{i=n}^{N_x}\phi(\mathbf{x}_n)\right) = \frac{1}{N_x^2}\mathbf{1}_{Nx,1}^T\mathbf{K}_{xx}\mathbf{1}_{Nx,1}$$

$$\bar{\phi}^T(\mathbf{x})\phi(\mathbf{y}_j) = \left(\frac{1}{N_x}\sum_{m=1}^{N_x}\phi(\mathbf{x}_m)\right)^T\phi(\mathbf{y}_j) = \frac{1}{N_x}\sum_{m=1}^{N_x}k(\mathbf{x}_m, \mathbf{y}_j)$$

$$\phi^T(\mathbf{y}_i)\bar{\phi}(\mathbf{x}) = \phi^T(\mathbf{y}_i)\left(\frac{1}{N_x}\sum_{m=1}^{N_x}\phi(\mathbf{x}_m)\right) = \frac{1}{N_x}\sum_{m=1}^{N_x}k(\mathbf{y}_i, \mathbf{x}_m)$$

where $\mathbf{1}_{Nx,1}$ is a column vector with all elements equal to one.

So we can express $\Phi_y^T\Phi_y$ in terms of the elements of $\mathbf{K}$ as:

$$\Phi_y^T\Phi_y = \mathbf{K}_{yy} - \frac{1}{N_x}\mathbf{K}_{yx}\mathbf{1}_{Nx,Ny} - \frac{1}{N_x}\mathbf{1}_{Ny,Nx}\mathbf{K}_{xy} + \frac{\alpha}{N_x^2}\mathbf{1}_{Ny,Ny} \tag{11}$$

where $\alpha = \mathbf{1}_{Nx,1}^T\mathbf{K}_{xx}\mathbf{1}_{Nx,1}$, and $\mathbf{1}_{Nx,Nx}$ is a matrix with all elements equal to one.

The prime subspace $\mathbf{E}$ of $\Phi_y^T\Phi_y$ is preserved, and that is $\mathbf{E}^T\Phi_y^T\Phi_y\mathbf{E} = \mathbf{D}_y \ne 0$. According to (10), we can obtain the diagonalized none-zero subspace $\mathbf{W} = \Phi_y\mathbf{E}\mathbf{D}_y^{-1}$ of $\Phi_y\Phi_y^T$, i.e. $\mathbf{W}^T\mathbf{S}_y^{\phi}\mathbf{W} \ne 0$. To obtain the most discriminant directions, the "positive with-in class scatter" matrix is projected into the none-zero space:

$$\mathbf{W}^T\mathbf{S}_x^{\phi}\mathbf{W} = \mathbf{D}_y^{-1}\mathbf{E}^T\Phi_y^T\mathbf{S}_x^{\phi}\Phi_y\mathbf{E}\mathbf{D}_y^{-1}. \tag{12}$$

So $\Phi_y^T\mathbf{S}_x^{\phi}\Phi_y$ should be derived first:

$$\Phi_y^T\mathbf{S}_x^{\phi}\Phi_y = \Phi_y^T\Phi_x\Phi_x^T\Phi_y = \left(\Phi_x^T\Phi_y\right)^T\left(\Phi_x^T\Phi_y\right). \tag{13}$$

According to (13), to calculate $\Phi_y^T\mathbf{S}_x^{\phi}\Phi_y$, we only need to calculate $\Phi_x^T\Phi_y$:

$$\Phi_x^T\Phi_y = \left[\phi^T(\mathbf{x}_i)\phi(\mathbf{y}_j) - \phi^T(\mathbf{x}_i)\bar{\phi}(\mathbf{x}) - \bar{\phi}^T(\mathbf{x})\phi(\mathbf{y}_j) + \bar{\phi}^T(\mathbf{x})\bar{\phi}(\mathbf{x})\right]_{\substack{1\le i\le Nx \\ 1\le j\le Ny}}$$

Through the previous analysis of $\Phi_y^T\Phi_y$, we only need to calculate $\phi^T(\mathbf{x}_i)\bar{\phi}(\mathbf{x})$ as:

$$\phi^T(\mathbf{x}_i)\bar{\phi}(\mathbf{x}) = \phi^T(\mathbf{x}_i)\left(\frac{1}{N_x}\sum_{m=1}^{N_x}\phi(\mathbf{x}_m)\right) = \frac{1}{N_x}\sum_{m=1}^{N_x}k(\mathbf{x}_i, \mathbf{x}_m),$$

and then we can obtain the $\Phi_x^T\Phi_y$ as:

$$\Phi_x^T\Phi_y = \mathbf{K}_{xy} - \frac{1}{N_x}\mathbf{K}_{xx}\mathbf{1}_{Nx,Ny} - \frac{1}{N_x}\mathbf{1}_{Nx,Nx}\mathbf{K}_{xy} + \frac{\alpha}{N_x^2}\mathbf{1}_{Nx,Ny}.$$

According to (13), $\Phi_y^T\mathbf{S}_x^{\phi}\Phi_y$ can be expressed as:

$$\Phi_y^T \mathbf{S}_x^{\phi} \Phi_y = \mathbf{A} - \frac{1}{N_x}\mathbf{B} + \frac{1}{N_x^2}\mathbf{C} - \frac{\alpha}{N_x^3}\mathbf{D}, \tag{14}$$

where

$$\mathbf{A} = \mathbf{K}_{yx}\mathbf{K}_{xy} + \frac{\alpha}{N_x^2}\left(\mathbf{K}_{yx}\mathbf{1}_{Nx,Ny} + \mathbf{1}_{Ny,Nx}\mathbf{K}_{xy}\right) + \frac{\alpha^2}{N_x^3}\mathbf{1}_{Ny,Ny},$$

$$\mathbf{B} = \mathbf{K}_{yx}\mathbf{K}_{xx}\mathbf{1}_{Nx,Ny} + \mathbf{K}_{yx}\mathbf{1}_{Nx,Nx}\mathbf{K}_{xy} + \mathbf{1}_{Ny,Nx}\mathbf{K}_{xx}\mathbf{K}_{xy} + \mathbf{K}_{yx}\mathbf{1}_{Nx,Nx}\mathbf{K}_{xy},$$

$$\mathbf{C} = \mathbf{1}_{Ny,Nx}\mathbf{K}_{xx}\mathbf{K}_{xx}\mathbf{1}_{Nx,Ny} + \mathbf{1}_{Ny,Nx}\mathbf{K}_{xx}\mathbf{1}_{Nx,Nx}\mathbf{K}_{xy} + \mathbf{K}_{yx}\mathbf{1}_{Nx,Nx}\mathbf{K}_{xx}\mathbf{1}_{Nx,Ny} + N_x\mathbf{K}_{yx}\mathbf{1}_{Nx,Nx}\mathbf{K}_{xy}, \text{ and}$$

$$\mathbf{D} = 2\cdot\mathbf{1}_{Ny,Nx}\mathbf{K}_{xx}\mathbf{1}_{Nx,Ny} + N_x\cdot\left(\mathbf{K}_{yx}\mathbf{1}_{Nx,Ny} + \mathbf{1}_{Ny,Nx}\mathbf{K}_{xy}\right).$$

Under the idea of DLDA, we do the eigenanalysis of $\widehat{\mathbf{S}}_x^{\phi} = \mathbf{D}_y^{-1}\mathbf{E}^T\Phi_y^T\mathbf{S}_x^{\phi}\Phi_y\mathbf{E}\mathbf{D}_y^{-1}$, and the eigenvectors $\mathbf{V}$ of $\widehat{\mathbf{S}}_x^{\phi}$ with the smallest eigenvalues $\mathbf{D}_x$ selected, i.e.

$$\mathbf{V}^T\widehat{\mathbf{S}}_x^{\phi}\mathbf{V} = \mathbf{D}_x. \tag{15}$$

So the kernel projection matrix is $\mathbf{H} = \mathbf{E}\mathbf{D}_y^{-1}\mathbf{V}\mathbf{D}_x^{-1/2}$.

Obviously, it is possible that some diagonal values in the matrix $\mathbf{D}_x$ is zero, which means that $\mathbf{D}_x^{-1/2}$ does not exist. However, we can avoid the zero eigenvalue problem based on a modified KBDA criterion, as similarly analyzed in [12]. The modified KBDA criterion is:

$$\mathbf{W} = \arg\max_{\mathbf{W}} \frac{\left\|\mathbf{W}^{\mathrm{T}}\mathbf{S}_y^{\phi}\mathbf{W}\right\|}{\left\|\mathbf{W}^{\mathrm{T}}\left(\mathbf{S}_x^{\phi} + \mathbf{S}_y^{\phi}\right)\mathbf{W}\right\|}. \tag{16}$$

It is clear that the modified criterion is equivalent to the original KBDA criterion according to the proof in [12]. Now the singular value problem can be avoided because $\left\|\mathbf{W}^{\mathrm{T}}\mathbf{S}_y^{\phi}\mathbf{W}\right\| = \mathbf{I}$.

With the optimal discriminant directions, which are drawn from the previous derivations, the projection of a new pattern $\mathbf{z}$ to $\mathbf{H}$ is given by:

$$\begin{aligned} f(\mathbf{z}) &= \left\{\left(\Phi_y\mathbf{D}_y^{-1}\mathbf{E}\right)\cdot\left(\mathbf{V}\mathbf{D}_x^{-1/2}\right)\right\}^T\phi(\mathbf{z}) = \mathbf{H}^T\Phi_y^T\phi(\mathbf{z}) \\ &= \mathbf{H}^T\left(\sum_{i=1}^{Ny}k(\mathbf{y}_i,\mathbf{z}) - \frac{1}{N_x}\sum_{j=1}^{Nx}k(\mathbf{x}_j,\mathbf{z})\right) \end{aligned} \tag{17}$$

So the steps of KDBDA can be summarized as follows:

a. Calculate the kernel matrix $\mathbf{K}$ according to (9).
b. Calculate $\Phi_y^T\Phi_y$ according to (11).
c. Extract the prime subspace of $\Phi_y^T\Phi_y$ by eigenanalysis. Then $\mathbf{E}$ is extracted to satisfy $\mathbf{E}^T\Phi_y^T\Phi_y\mathbf{E} = \mathbf{D}_y \neq 0$.
d. Calculate $\Phi_y^T\mathbf{S}_x^{\phi}\Phi_y$ according to (14).
e. With the modified KBDA criterion, select eigenvectors $\mathbf{V}$ of $\widehat{\mathbf{S}}_x^{\phi} = \mathbf{D}_y^{-1}\mathbf{E}^T\Phi_y^T\mathbf{S}_x^{\phi}\Phi_y\mathbf{E}\mathbf{D}_y^{-1}$ with the smallest eigen-values $\mathbf{D}_x$ by eigenanalysis.
f. Calculate the kernel projection matrix $\mathbf{H} = \mathbf{E}\mathbf{D}_y^{-1}\mathbf{V}\mathbf{D}_x^{-1/2}$.
g. For a given pattern, the KDBDA transformation is:

$$f(\mathbf{z}) = \mathbf{H}^T\left(\sum_{i=1}^{Ny}k(\mathbf{y}_i,\mathbf{z}) - \frac{1}{N_x}\sum_{j=1}^{Nx}k(\mathbf{x}_j,\mathbf{z})\right).$$

4. EXPERIMENTS

4.1. The Image Retrieval System

To evaluate the performance of the proposed algorithm, we have designed the QueryGo image retrieval system. The user interface is shown in Fig 1.

Figure 1. User interface of the proposed system.

In QueryGo, three main features, color, texture, and shape, are extracted and used to represent the corresponding image. For color, we select the color histogram [13] in HSV color space to represent the color information of an image. Here, the color histogram is quantized into 256 levels. Hue, Saturation and Value are quantized into 8, 8, and 4 bins respectively. Texture is extracted from Y component in YCrCb space by pyramidal wavelet transform (PWT) with Haar wavelet. The mean value and standard deviation are calculated in the sub-bands of each decomposed level. The feature length is $2\times4\times3$. Edge histogram [14] is also calculated on Y component in YCrCb color space. Edges are grouped into four categories, which are horizontal, 45 diagonal, vertical, and 135 diagonal. From the edge histogram, we can get a four-dimension shape feature for image retrieval. Each feature is sensitive to a particular property of the content of an image. We combine the color, texture, and shape features into a feature vector, and then we normalize each feature into a normal distribution.

With the QueryGo system and the proposed RF algorithm, the retrieval procedure is: 1. User inputs a query image; 2. The visual feature of the query is extracted by QueryGo; 3. According to the Euclidean distance, all images in the database is ascending sorted based on dissimilarity; 4. User choose positive or negative feedback; 5. The kernel projection matrix is calculated based on the steps a-f of the KDBDA algorithm; 6. Query and all images in the database is projected in the kernel space based on step g of the KDBDA algorithm; 7. Go to step 3, until the user confirms a satisfied result.

4.2 Evaluation of the Experimental Results

We evaluate the performance of the proposed algorithm according to the accuracy, i.e. the ratio of the number of retrieved relevant images to the top N retrieved images. We did the statistical experiments on a large-scale image database, which includes 17,800 Corel Images [3] with 90 concepts (relabeled by ourselves).

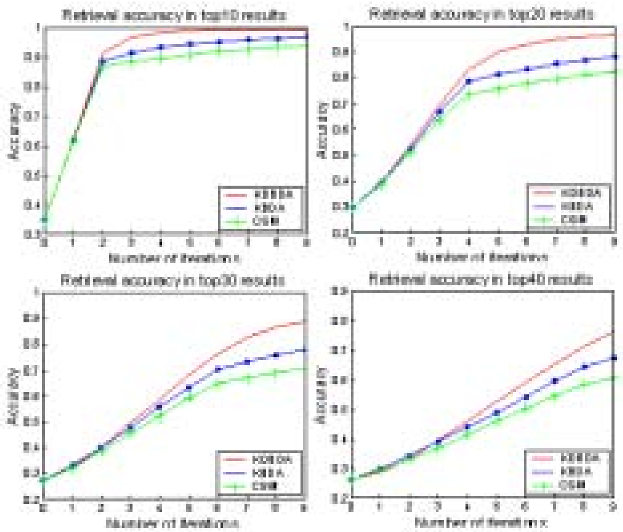

Figure 2. Top-left is the top 10 accuracy, top-right is the top 20 accuracy, bottom-left is the top 30 accuracy, and bottom-right is the top 40 accuracy. The x-coordinate is the i^{th} relevance feedback iteration, and the y-coordinate is the accuracy of the top n images.

In these experiments, we compared the new algorithm (KDBDA) with two state-of-the-art algorithms, the constrained support vector machine (CSM) [15] and KBDA [7]. The computer automatically did the feedback experiments with 300 queries. For each iteration, the system marked the first 5 incorrect (correct) retrieved images from the top 48 matches as irrelevant (relevant) examples. In the kernel based algorithms, we chose the Gaussian kernel $K(\mathbf{x},\mathbf{y}) = e^{-\rho\|\mathbf{x}-\mathbf{y}\|^2}$, because it shows the best performance. The kernel parameters are chosen from a series of values, which shows the best performance for each of the algorithms. For CSM, the best value is $\rho = 1$. For KDBDA and KBDA, the best performance is found at $\rho = 1/10$.

Figure 2 shows the performance of KDBDA, KBDA and CSM. The results show that our algorithm KDBDA outperforms KBDA and CSM consistently. In addition, the computational costs of the three methods are similar in our experiments.

5. CONCLUSION

In this paper, we proposed a straightforward method to solve the matrix singular problem of the modified biased discriminant analysis in the kernel feature space. The new algorithm removes the null space of “the negative scatter with respect to positive centroid” matrix, and then the eigenvectors of the “the positive with-in class scatter” matrix corresponding to the smallest eigenvalues are extracted as the most discriminant directions in the kernel space. From a large number of evaluation experiments, we can draw the conclusion that Kernel Direct Biased Discriminant Analysis outperforms both the traditional Kernel Biased Discriminant Analysis and the Constrained Support Vector Machine.

6. ACKNOWLEDGMENTS

The work described in this paper was fully supported by a AoE in IT grant from the Research Grants Council of the Hong Kong SAR.

7. REFERENCES

[1] A. Smeulders, M. Worring, S. Santini, A. Gupta, and R. Jain, “Content-based image retrieval at the end of the early years”, IEEE Trans. on PAMI, vol. 22, no. 12, pp. 1349 -- 1380, Dec. 2000.
[2] W. Ma and H. Zhang, “Content-based image indexing and retrieval”, in Handbook of Multimedia Computing, CRC Press, 1998.
[3] J. Wang, J. Li, and G. Wiederhold, "SIMPLIcity: semantics-sensitive integrated matching for picture libraries", IEEE Trans. on PAMI, vol. 23 no. 9, pp. 947 –963, Sept. 2001.
[4] Y. Rui, T. Huang, and S. Mehrotra, “Content-based image retrieval with relevance feedback in MARS”. IEEE Int. Conf. on IP, pp.815-818, Oct. 1997.
[5] Y. Ishikawa, R. Subramanya, and C. Faloutsos, “Mindreader: Querying databases through multiple examples”. In 24th Int. Conf. on VLDB 1998, pp. 433-438, 1998.
[6] Y. Rui, T. Huang, M. Ortega, and S. Mehrotra. “Relevance feedback: A power tool in interactive content-based image retrieval”. IEEE Tran. on CSVT, vol.8 no.5 pp. 644-655, Sept. 1998.
[7] X. Zhou and T. Huang, “small sample learning during multimedia retrieval using biasmap”. IEEE CVPR , pp. 11-17, Dec. 2001.
[8] K. Fukunaga. Introduction to statistical pattern recognition (2nd). Academic Press, Boston 1990.
[9] H. Yu and J. Yang, “A direct LDA algorithm for high-dimensional data with application to face recognition”, IJPR, vol. 34, pp 2067-2070, 2001.
[10] K. Muller, S. Mika, G. Ratsch, K. Tsuda, and B. Scholkopf, “An Introduction to Kernel-Based Learning Algorithms”, IEEE. Trans on NN, vol. 12, no. 2, Mar. 2001.
[11] J. Lu, K. Plataniotis, and A. Venetsanopoulos,"Face recognition using kernel direct discriminant analysis algorithms", IEEE Trans. on NN, vol. 14, no. 1, pp. 117-126, Jan. 2003
[12] K. Liu, Y. Cheng, J, Yang, and X. Liu, An efficient algorithm for foley-sammon optimal set of discriminant vectors by algebraic method, IJPR, vol. 6, pp. 817-829, 1992.
[13] M.J. Swain and D.H. Ballard. “Color indexing”. IJCV, pp. 11--32, 1991.
[14] B. Manjunath, J. Ohm, V. Vasudevan, and A. Yamada, “Color and texture descriptors”, IEEE Trans. on , CSVT, vol. 11, pp. 703-715, Jun. 2001.
[15] G. Guo, A. Jain, W. Ma, and H. Zhang, "learning similarity measure for natural image retrieval with relevance feedback". IEEE Trans. on NN., vol. 12, no. 4, pp.811-820, July 2002.