A KERNEL MAXIMUM UNCERTAINTY DISCRIMINANT ANALYSIS AND ITS APPLICATION TO FACE RECOGNITION

Carlos Eduardo Thomaz

Department of Electrical Engineering, Centro Universitario da FEI, FEI, Sao Paulo, Brazil

Gilson Antonio Giraldi

Department of Computer Science, National Laboratory for Scientific Computing, LNCC, Rio de Janeiro, Brazil

Keywords: Non-linear discriminant analysis, Limited sample size problems, Face recognition.

Abstract: In this paper, we extend the Maximum uncertainty Linear Discriminant Analysis (MLDA), proposed recently for limited sample size problems, to its kernel version. The new Kernel Maximum uncertainty Discriminant Analysis (KMDA) is a two-stage method composed of Kernel Principal Component Analysis (KPCA) followed by the standard MLDA. In order to evaluate its effectiveness, experiments on face recognition using the well-known ORL and FERET face databases were carried out and compared with other existing kernel discriminant methods, such as Generalized Discriminant Analysis (GDA) and Regularized Kernel Discriminant Analysis (RKDA). The classification results indicate that KMDA performs as well as GDA and RKDA, with the advantage of being a straightforward stabilization approach for the within-class scatter matrix that uses higher-order features for further classification improvements.

1 INTRODUCTION

The primary purpose of Linear Discriminant Analysis (LDA) is to separate samples of distinct groups by maximizing their between-class separability while minimizing their within-class variability (Fukunaga, 1990; Devijver and Kittler, 1982).

However, in limited sample and high dimensional problems, such as face recognition, the within-class scatter matrix is either singular or mathematically unstable and the standard LDA cannot be used to perform the separating task. In the last years, a number of linear methods have been proposed to overcome this difficulty (Swets and Weng, 1996; Belhumeur et al., 1997; Chen et al., 2000; Yu and Yang, 2001; Yang and Yang, 2003; Thomaz et al., 2006), making LDA applicable to limited sample size problems that have been assumed to be linearly separable in the original space.

More recently, in order to make LDA applicable to non-linearly separable data as well, kernel-based methods have been applied. The main idea of kernel-based methods is to map the original input data to a feature space by a non-linear mapping where inner products in the feature space can be computed by a kernel function without knowing the non-linear mapping explicitly (Park and Park, 2005). Works in this area include the Kernel Principal Component Analysis (KPCA) (Scholkopf et al., 1998), Generalized Discriminant Analysis (GDA)(Baudat and Anouar, 2000), and Regularized Kernel Discriminant Analysis (RKDA) (Lu et al., 2003), among others. In the specific case of GDA, it has been demonstrated in (Yang et al., 2004) that GDA is in fact equivalent to the two-stage method composed of KPCA followed by the standard LDA.

In this paper, we extend the Maximum uncertainty Linear Discriminant Analysis (MLDA) approach (Thomaz et al., 2006), proposed recently for solving limited sample size problems in discriminant analysis, to its kernel or non-linear version. This non-linear version of MLDA, here called Kernel Maximum uncertainty Discriminant Analysis (KMDA), is a two-stage method composed of Kernel Principal Component Analysis (KPCA) followed by the standard MLDA. The effectiveness of KMDA is evaluated on face recognition through comparisons with KPCA, GDA and RKDA, using the well-known Olivetti-Oracle Research Lab (ORL) (Samaria and Harter, 1994) and FERET face databases (Phillips et al.,

1998). One advantage of the proposed method for face images is the possibility of improving classification performance by using more non-linear intermediate features than pixels in the images. In addition, KMDA does not require the selection of parameters, like RKDA, for within-class scatter matrix stabilization.

The paper is organized as follows. In section 2 we review briefly the LDA and MLDA approaches. Then, in section 3, we explain how we have extended the MLDA approach to its non-linear version using the mathematical result described in (Yang et al., 2004). The set up of the experiments carried out in this work as well as the classification results on face recognition are presented respectively in sections 4 and 5, comparing the KMDA recognition rates with KPCA, GDA and RKDA. In section 6, we analyze and discuss the non-linear classification results of KMDA with the MLDA ones published in (Thomaz et al., 2006). Finally, in section 7, we conclude the paper, summarizing its main contribution and indicating possible future work.

2 LINEAR DISCRIMINANT ANALYSIS (LDA)

Let the scatter matrices between-class $\mathbf{S}_b$ and within-class $\mathbf{S}_w$ be defined, respectively, as

$$\mathbf{S}_b = \sum_{i=1}^{g} N_i (\overline{\mathbf{x}}_i - \overline{\mathbf{x}})(\overline{\mathbf{x}}_i - \overline{\mathbf{x}})^T \qquad (1)$$

and

$$\mathbf{S}_w = \sum_{i=1}^{g} (N_i - 1)\mathbf{S}_i = \sum_{i=1}^{g} \sum_{j=1}^{N_i} (\mathbf{x}_{i,j} - \overline{\mathbf{x}}_i)(\mathbf{x}_{i,j} - \overline{\mathbf{x}}_i)^T, \qquad (2)$$

where $\mathbf{x}_{i,j}$ is the n-dimensional pattern (or sample) j from class i, N_i is the number of training patterns from class i, and g is the total number of classes or groups. The vector $\overline{\mathbf{x}}_i$ and matrix $\mathbf{S}_i$ are respectively the unbiased mean and sample covariance matrix of class i (Fukunaga, 1990). The grand mean vector $\overline{\mathbf{x}}$ is given by

$$\overline{\mathbf{x}} = \frac{1}{N} \sum_{i=1}^{g} N_i \overline{\mathbf{x}}_i = \frac{1}{N} \sum_{i=1}^{g} \sum_{j=1}^{N_i} \mathbf{x}_{i,j}, \qquad (3)$$

where N is the total number of samples, that is, $N = N_1 + N_2 + \ldots + N_g$. It is important to note that the within-class scatter matrix $\mathbf{S}_w$ defined in equation (2) is essentially the standard pooled covariance matrix $\mathbf{S}_p$ multiplied by the scalar $(N - g)$, where $\mathbf{S}_p$ can be written as

$$\mathbf{S}_p = \frac{1}{N - g} \sum_{i=1}^{g} (N_i - 1)\mathbf{S}_i$$
$$= \frac{(N_1 - 1)\mathbf{S}_1 + (N_2 - 1)\mathbf{S}_2 + \ldots + (N_g - 1)\mathbf{S}_g}{N - g}. \qquad (4)$$

The main objective of LDA is to find a projection matrix $\mathbf{W}_{lda}$ that maximizes the ratio of the determinant of the between-class scatter matrix to the determinant of the within-class scatter matrix (Fisher's criterium), that is,

$$\mathbf{W}_{lda} = \arg\max_{W} \frac{|\mathbf{W}^T \mathbf{S}_b \mathbf{W}|}{|\mathbf{W}^T \mathbf{S}_w \mathbf{W}|}. \qquad (5)$$

The Fisher's criterium described in equation (5) is maximized when the projection matrix $\mathbf{W}_{lda}$ is composed of the eigenvectors of $\mathbf{S}_w^{-1}\mathbf{S}_b$ with at most $(g - 1)$ nonzero corresponding eigenvalues (Fukunaga, 1990; Devijver and Kittler, 1982).

However, in limited sample and high dimensional problems, such as face recognition, $\mathbf{S}_w$ is either singular or mathematically unstable and the standard LDA cannot be used to perform the separating task. To avoid both critical issues, Thomaz et al. have calculated $\mathbf{W}_{lda}$ by using a maximum uncertainty LDA-based approach (MLDA) that considers the issue of stabilizing the $\mathbf{S}_w$ estimate with a multiple of the identity matrix (Thomaz et al., 2004; Thomaz et al., 2006).

The MLDA algorithm can be described as follows:

1. Find the Ψ eigenvectors and Λ eigenvalues of $\mathbf{S}_p$, where $\mathbf{S}_p = \frac{S_w}{N-g}$;

2. Calculate the $\mathbf{S}_p$ average eigenvalue $\overline{\lambda}$, that is,

$$\overline{\lambda} = \frac{1}{n} \sum_{j=1}^{n} \lambda_j = \frac{Tr(\mathbf{S}_p)}{n}; \qquad (6)$$

3. Form a new matrix of eigenvalues based on the following largest dispersion values

$$\Lambda^* = diag[max(\lambda_1, \overline{\lambda}), max(\lambda_2, \overline{\lambda}), \ldots, max(\lambda_n, \overline{\lambda})]; \qquad (7)$$

4. Form the modified within-class scatter matrix

$$\mathbf{S}_w^* = \mathbf{S}_p^*(N - g) = (\Psi \Lambda^* \Psi^T)(N - g). \qquad (8)$$

The MLDA method is constructed by replacing $\mathbf{S}_w$ with $\mathbf{S}_w^*$ in the Fisher's criterium formula described in equation (5).

3 A KERNEL MLDA (KMDA)

Since the non-linear mapping of the original space to a higher dimensional feature space would commonly lead to an ill-posed within-class scatter matrix, the aforementioned MLDA approach might be suitable for solving not only the singularity and instability issues of the linear Fisher methods, but also the Fisher discriminant analysis with kernels.

Let a non-linear function ϕ that maps the input sample space $\mathbb{R}^n$ into the feature space $\mathfrak{F}$, as follows:

$$\phi : \mathbf{x} \in \mathbb{R}^n \rightarrow \phi(\mathbf{x}) \in \mathfrak{F}. \quad (9)$$

The between-class and within-class scatter matrices in the feature space $\mathfrak{F}$ can be defined, respectively, as

$$\tilde{\mathbf{S}}_b = \sum_{i=1}^{g} N_i (\overline{\phi}_i - \overline{\phi})(\overline{\phi}_i - \overline{\phi})^T \quad (10)$$

and

$$\tilde{\mathbf{S}}_w = \sum_{i=1}^{g} \sum_{j=1}^{N_i} (\phi(\mathbf{x}_{i,j}) - \overline{\phi}_i)(\phi(\mathbf{x}_{i,j}) - \overline{\phi}_i)^T, \quad (11)$$

where $\overline{\phi}_i$ is the mean of the training samples of class i mapped into the feature space, that is,

$$\overline{\phi}_i = \frac{1}{N_i} \sum_{j=1}^{N_i} \phi(\mathbf{x}_{i,j}), \quad (12)$$

and $\overline{\phi}$ is the grand mean vector of all the training samples mapped into the feature space, that is,

$$\overline{\phi} = \frac{1}{N} \sum_{i=1}^{g} \sum_{j=1}^{N_i} \phi(\mathbf{x}_{i,j}). \quad (13)$$

It is important to note that when $\phi(\mathbf{x}) = \mathbf{x}$ equations (10) and (11) reduce to their corresponding linear versions described in equations (1) and (2), respectively.

Thus, according to the Fisher's criterium described in equation (5), the kernel LDA projection matrix $\tilde{\mathbf{W}}_{lda}$ can be determined by calculating the eigenvectors of $\tilde{\mathbf{S}}_w^{-1} \tilde{\mathbf{S}}_b$ with at most $(g-1)$ nonzero corresponding eigenvalues. These eigenvectors are then the optimal linear discriminant directions on the feature space, which represent non-linear discriminant directions in the input sample space.

To extend the MLDA approach to its non-linear (or kernel) version, we need essentially to replace $\tilde{\mathbf{S}}_w$ with $\tilde{\mathbf{S}}_w^*$, using the MLDA algorithm described in the previous section. One way to do this would be to perform the eigen-analyzes of $\tilde{\mathbf{S}}_b$ and $\tilde{\mathbf{S}}_w$ in the feature space, as proposed in (Lu et al., 2003). Alternatively, we could use the more intuitive mathematical result described in (Yang et al., 2004). According to

Yang et al. (Yang et al., 2004), the kernel LDA is in fact equivalent to the two-stage method composed of KPCA (Scholkopf et al., 1998) followed by the standard LDA.

Therefore, instead of solving the eigenvalue problem of $\tilde{\mathbf{S}}_b$ and $\tilde{\mathbf{S}}_w^*$ directly in the feature space, we perform firstly KPCA (Scholkopf et al., 1998) in the input sample space, changing the dimension of feature space to m, and next the standard MLDA to extract the linear discriminant features in the non-linear transformed space $\mathbb{R}^m$ given by the KPCA projection. The whole process is summarized in Figure 1.
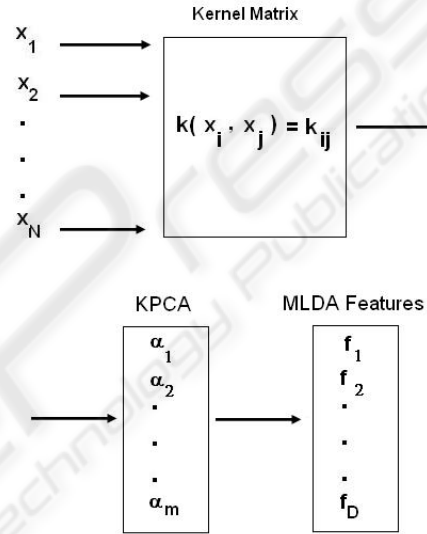

Figure 1: Pipeline of the KMDA method. Firstly, the kernel matrix is generated from input samples and KPCA is applied. Then, the standard MLDA is used to extract the D discriminant features in the space given by KPCA projection, where $D \leq (g-1)$.

The goal of KPCA is to diagonalize the covariance matrix $\tilde{\mathbf{S}}$ defined as (Zheng et al., 2005):

$$\tilde{\mathbf{S}} = \frac{1}{N} \sum_{i=1}^{g} \sum_{j=1}^{N_i} (\phi(\mathbf{x}_{i,j}) - \overline{\phi})(\phi(\mathbf{x}_{i,j}) - \overline{\phi})^T. \quad (14)$$

For simplicity, let us suppose that $\overline{\phi} = \mathbf{0}$. So, we must find the eigenvectors $\mathbf{v}$ and the corresponding eigenvalues $\lambda \geq 0$, solutions of the eigenequation:

$$\lambda \mathbf{v} = \tilde{\mathbf{S}} \mathbf{v}. \quad (15)$$

However, in kernel methods we do not know the function ϕ explicitly but a kernel k such that $k(\mathbf{x}, \mathbf{y}) = \phi(\mathbf{x})^T \cdot \phi(\mathbf{y})$. Thus, we must obtain a kernel version of expression (15). In fact, it can be shown that the eigenvectors $\mathbf{v}$ can be written as follows (Scholkopf et al., 1998):

$$\mathbf{v} = \sum_{i=1}^{g} \sum_{j=1}^{N_i} \alpha_{ij} \phi(\mathbf{x}_{i,j}) = \mathbf{F}(\mathbf{X})\alpha, \qquad (16)$$

where $\mathbf{F}(\mathbf{X}) = \left[\phi(\mathbf{x}_{1,1}) \phi(\mathbf{x}_{1,2}) \cdots \phi(\mathbf{x}_{g,N_g}) \right]$, and $\alpha = \left[\alpha_{11} \alpha_{12} \cdots \alpha_{1N_1} \cdots \alpha_{g1} \alpha_{g2} \cdots \alpha_{gN_g} \right]$. By substituting (16) into (15), we obtain the KPCA eigenvalue problem:

$$N\lambda\alpha = \mathbf{K}\alpha, \qquad (17)$$

where $\mathbf{K} = \left(k_{\omega,\gamma} = k(\mathbf{x}_{i,j}, \mathbf{x}_{s,t}) \right)$, with $\omega = j + N_{i-1} + N_{i-2} + ... + N_1$ and $\gamma = t + N_{s-1} + N_{s-2} + ... + N_1$, is a $N \times N$ matrix called kernel matrix.

The proposed strategy, that is KPCA+MLDA or simply KMDA, is particularly useful when solving limited sample and high-dimensional problems, because m is upper bounded by N, i.e. $m \leq (N-1)$. Since the MLDA approach deals with the singularity and instability of the within-class scatter matrix in such limited sample size situations, we have selected $m = (N-1)$ to reproduce the total variability of the samples in the feature space.

4 EXPERIMENTS

To evaluate the effectiveness of KMDA on face recognition, comparisons with KPCA (Scholkopf et al., 1998), GDA (Baudat and Anouar, 2000), and RKDA (Lu et al., 2003), were performed using the well-known Olivetti-Oracle Research Lab (ORL) (Samaria and Harter, 1994) and FERET (Phillips et al., 1998) face databases. Figure 2 shows some samples of these datasets.
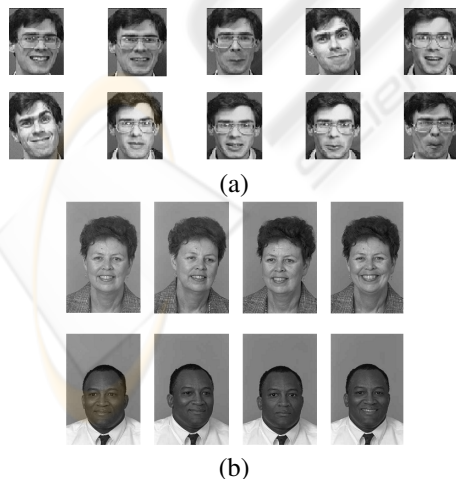

(a)

(b)

Figure 2: (a) A set of ten images of one subject from the ORL face database. (b) Sets of four images of two subjects from the FERET database.

We have implemented the KPCA, GDA and RKDA using the respective authors' Matlab codes available at the following website: http://www.kernel-machines.org/software.

For simplicity, an Euclidean distance classifier was used to perform classification in the non-linear feature space. Also, we have used only the well-known Gaussian kernel

$$k(\mathbf{x}_1, \mathbf{x}_2) = exp(\frac{-\| \mathbf{x}_1 - \mathbf{x}_2 \|^2}{\delta}) \qquad (18)$$

to compute indirectly the non-linear transformations, where the δ parameter range was taken to be $[0.001, 0.002, 0.004, 0.008, 0.01, ..., 1.0]$ times the dimension of the input sample space n for all the aforementioned algorithms tested.

To determine the regularization parameter η of the RKDA approach (Lu et al., 2003), experimental analyzes were carried out based on the best classification accuracy given the following parameter set $[0.001, 0.01, 0.1, 0.2, ..., 0.9, 0.99, 0.999, 1.0]$. The best results were obtained when $\eta = 0.001$ for both ORL and FERET experiments.

Each experiment was repeated 25 times using several features. Distinct training and test sets were randomly drawn, and the mean and standard deviation of the recognition rate were calculated. The classification of the ORL 40 subjects was computed using for each individual 5 images to train and 5 images to test. In the FERET database with 200 subjects, the training and test sets were respectively composed of 3 and 1 frontal images.

Analogously to the experiments carried out in (Thomaz et al., 2006), to represent a recognition problem where the within-class scatter matrix is singular, the ORL face images were resized to 32x32 pixels, that is, the total number of training observations was $N = 200$ and the dimensionality of the original images was $n = 1024$. The FERET images were resized to 16x16 pixels in order to pose an alternative pattern recognition problem where the within-class scatter matrix is non-singular but poorly estimated, i.e., $N = 600$ and $n = 256$.

5 RESULTS

Table 1 shows the maximum test average recognition rates with standard deviation (std) of the ORL and FERET datasets over the Gaussian kernel parameter δ, and the corresponding number of principal (F1) and discriminant (F2) features. The notation '—' in the rows of Table 1 indicates that the corresponding method has been calculated using either F1 or F2

features, but not both. In fact, KMDA is the only discriminant kernel method investigated in this work that is explicitly composed of a two-stage non-linear transformation.

As we should expect, all non-linear discriminant methods (GDA, RKDA and KMDA) led to higher classification results than KPCA.

Table 1: ORL and FERET classification results.

| Dataset | | Features | | |
Method	δ	F1	F2	% (std)
ORL				
KPCA	0.02*1024	160	—	93.0 (1.9)
GDA	0.08*1024	—	39	96.5 (1.3)
RKDA	1.00*1024	—	31	95.7 (1.3)
KMDA	1.00*1024	199	39	96.2 (1.6)
FERET				
KPCA	0.10*256	599	—	92.3 (1.3)
GDA	1.00*256	—	104	95.0 (1.0)
RKDA	0.20*256	—	159	97.8 (0.8)
KMDA	1.00*256	599	20	98.3 (0.9)

In the ORL experiments, the best classification result was reached by GDA (96.5%), followed by KMDA (96.2%) and RKDA (95.7%). Given the similarity of these recognition rates and their corresponding standard deviations, we cannot see clearly an overall best classification performance of any kernel discriminant method in these experiments. Since the ORL face database contains only 40 subjects to be separated, the discriminant features of the kernel Fisher-based methods were limited to 39 components. Although in such experiments, where $n(= 1024) > N(= 200)$, the intermediate KPCA transformation of KMDA allows the within-class and between-class scatter matrices to be calculable in computers with a standard memory size, KMDA requires a two-stage final transformation that uses more features than the GDA and RKDA one-stage methods.

One advantage of using a non-linear two-stage method such as KMDA in limited sample and high dimensional problems can be in the FERET results. In this case, the discriminant features of the kernel Fisher-based methods were limited to 199 components, because the FERET dataset contains only 200 subjects to be separated. However, since $N(= 600) > n(= 256)$, the classification performance of KMDA can be further improved by using more non-linear intermediate features ($N - 1 = 599$) than there are pixels in the 16x16 images. In this application, where the within-class scatter matrix was non-singular but poorly estimated, KMDA achieved the best classification accuracy (98.3%) using a higher-order KPCA transformation with 599 principal components followed by an MLDA transformation composed of only

20 discriminant components.

Another KMDA advantage, comparing specifically with the other regularized Fisher discriminant method RKDA, is the fact that KMDA is based on a straightforward stabilization approach for the within-class scatter matrix, avoiding the RKDA optimization for which the user has to select experimentally the best η regularization parameter.

6 DISCUSSION

We have used the same ORL and FERET face databases and carried out the same training and test experiments described in (Thomaz et al., 2006) for the standard MLDA. So, it is possible to compare the classification results of KMDA with the ones presented in (Thomaz et al., 2006) for MLDA.

Table 2 highlights the MLDA maximum test average recognition rates with standard deviation (std) of the ORL and FERET datasets over the corresponding number of principal (F1) and discriminant (F2) features, as published in (Thomaz et al., 2006), and also the KMDA results described previously in Table 1.

Table 2: MLDA versus KMDA classification results.

| Dataset | Features | | |
Method	F1	F2	% (std)
ORL			
MLDA	—	39	95.8 (1.6)
KMDA	199	39	96.2 (1.6)
FERET			
MLDA	—	10	95.4 (1.4)
KMDA	599	20	98.3 (0.9)

As can be seen, for the ORL dataset with face images resized to 32x32 pixels, there is no significant classification improvement in using KMDA rather than MLDA in these experiments, because the corresponding MLDA and KMDA recognition rates (and standard deviations) are very similar. In such small sample and high-dimensional problem, where the two-stage KMDA could not extract higher-order features because $N(= 200) < n(= 1024)$, MLDA seems the best choice because it is simpler and much faster to compute.

However, the superiority of KMDA compared to MLDA is clear in the FERET dataset with face images resized to 16x16 pixels. The KMDA classifier performed better than its linear version, achieving a higher maximum average classification accuracy with lower standard deviation. In these experiments, KMDA outperformed MLDA by seeking discriminant hyperplanes not in the 256-dimensional origi-

nal space, but in a much higher 599-dimensional feature space composed of non-linear transformations of the original pixels. In such limited sample and high-dimensional problem, where $N(= 600) > n(= 256)$, it seems that we can further improve the classification accuracy by exploring more features than is possible in the linear case.

7 CONCLUSIONS

In this work, we extended the MLDA approach to its non-linear version. This non-linear version of MLDA, here called KMDA, is a KPCA+MLDA two-stage method. To evaluate the KMDA effectiveness, experiments on face recognition using the well-known ORL and FERET face databases were carried out and compared with other existing kernel discriminant methods, such as GDA and RKDA. The classification results indicate that KMDA performs as well as GDA and RKDA, with the advantages of being a straightforward stabilization approach for the within-class scatter matrix that uses a pre-defined number of higher-order features whenever the number of training samples is larger than the original dimensionality of the input data.

As future work, we intend to directly regularize the eigen-analysis of the within-class scatter matrix in the feature space, without a KPCA intermediate step.

ACKNOWLEDGEMENTS

REFERENCES

Baudat, G. and Anouar, F. (2000). Generalized discriminant analysis using a kernel approach. *Neural Computation*, 12(10):2385–2404.

Belhumeur, P. N., Hespanha, J. P., and Kriegman, D. J. (1997). Eigenfaces vs. fisherfaces: Recognition using class specific linear projection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 19(7):711–720.

Chen, L., Liao, H., Ko, M., Lin, J., and Yu, G. (2000). A new lda-based face recognition system which can solve the small sample size problem. *Pattern Recognition*, 33(10):1713–1726.

Devijver, P. and Kittler, J. (1982). *Pattern Classification: A Statistical Approach*. Prentice-Hall.

Fukunaga, K. (1990). *Introduction to Statistical Pattern Recognition*. Morgan Kaufmann, San Francisco, 2nd edition.

Lu, J., Plataniotis, K. N., and Venetsanopoulos, A. N. (2003). Face recognition using kernel direct discriminant analysis algorithms. *IEEE Transactions on Neural Networks*, 14(1):117–126.

Park, C. H. and Park, H. (2005). Nonlinear discriminant analysis using kernel functions and the generalized singular value decomposition. *SIAM J. Matrix Anal. Appl.*, 27(1):87–102.

Phillips, P. J., Wechsler, H., Huang, J., and Rauss, P. (1998). The feret database and evaluation procedure for face recognition algorithms. *Image and Vision Computing*, 16:295–306.

Samaria, F. and Harter, A. (1994). Parameterisation of a stochastic model for human face identification. In *Proceedings of 2nd IEEE Workshop on Applications of Computer Vision*.

Scholkopf, B., Smola, A., and Muller, K.-R. (1998). Non-linear component analysis as a kernel eigenvalue problem. *Neural Computation*, 10(5):1299–1319.

Swets, D. L. and Weng, J. J. (1996). Using discriminant eigenfeatures for image retrieval. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 18(8):831–836.

Thomaz, C. E., Gillies, D. F., and Feitosa, R. Q. (2004). A new covariance estimate for bayesian classifiers in biometric recognition. *IEEE Transactions on Circuits and Systems for Video Technology, Special Issue on Image- and Video-Based Biometrics*, 14(2):214–223.

Thomaz, C. E., Kitani, E. C., and Gillies, D. F. (2006). A maximum uncertainty lda-based approach for limited sample size problems - with application to face recognition. *Journal of the Brazilian Computer Society*, 12(2):7–18.

Yang, J., Jin, Z., yu Yang, J., Zhang, D., and Frangi, A. F. (2004). Essence of kernel fisher discriminant: Kpca plus lda. *Pattern Recognition*, 37:2097–2100.

Yang, J. and Yang, J. (2003). Why can lda be performed in pca transformed space? *Pattern Recognition*, 36:563–566.

Yu, H. and Yang, J. (2001). A direct lda algorithm for high dimensional data - with application to face recognition. *Pattern Recognition*, 34:2067–2070.

Zheng, W., Zou, C., and Zhao, L. (2005). An improved algorithm for kernel principal component analysis. *Neural Process. Lett.*, 22(1):49–56.