

# A New Kernel Direct Discriminant Analysis (KDDA) Algorithm for Face Recognition

WU Xiao-Jun<sup>1)2)3)</sup> Josef Kittler<sup>2)</sup> YANG Jing-Yu<sup>4)</sup> Kieron Messer<sup>2)</sup>

Wang Shi-Tong<sup>4)</sup>

1) (Department of Computer Science, Jiangsu University of Science and Technology, Zhenjiang, China, 212003)

2) (CVSSP, Department of Electrical Engineering, University of Surrey, GU2 7XH, Surrey, UK)

3) (Robotics Laboratory, Shenyang Institute of Automation, Chinese Academy of Sciences, Shenyang, China, 110015)

4) (School of Information, Nanjing University of Science & Technology, Nanjing, China, 210094)

## Abstract

We propose a new kernel direct discriminant analysis (KDDA) algorithm in this paper. First, a recently advocated direct linear discriminant analysis (DLDA) algorithm is overviewed. Then the new KDDA algorithm is developed which can be considered as a kernel version of the DLDA algorithm. The design of the minimum distance classifier in the new kernel subspace is then discussed. The results of experiments on two well-known facial databases show the effectiveness of the proposed method in face recognition. The results of experiments also confirm that DLDA can be viewed as a special case of the proposed KDDA algorithm.

## 1. Introduction

Feature extraction is the most fundamental problem in pattern recognition. For any pattern recognition application, extracting efficient features is always a key to solving the problem [1-2]. Among the many representations proposed in the literature algebraic features are very popular as they represent intrinsic attributes of an image. Turk and Pentland used eigenfaces as the bases for constructing features for human face recognition [3]. Hong and Yang proposed an algebraic feature method in which the components of a singular value vector were used as the image features [4]-[5]. Although eigenfaces and singular value vector have good properties for representing images, they are not quite suitable for recognizing images. Based on this viewpoint, Cheng presented an efficient approach to human face recognition based on projective images [6]. Also Foley-Sammon transform (FST) has been considered as an excellent method of dimensionality reduction in terms of discriminant information content [7]-[13]. Various methods of determining FST under different conditions have been developed for human facial image recognition [14]-[20]. Liu proposed a class separability criterion leading to a generalized optimal set of discriminant vectors. A unified approach to finding the vectors of the generalized optimal set was introduced [13]-[14]. Guo proposed an iterative algorithm to find the generalized optimal set of discriminant vectors [15]. Wu proposed an approximate analytical algorithm for determining the generalized optimal set of discriminant vectors [16]. An accurate analytical solution

to the problem is proposed in [17]. Chen developed an LDA-based face recognition system which can cope with the small sample size (SSS) problem [18]. However, it has been shown that the discarded null space may contain significant information which may be of benefit for the classification [19]-[21]. The so called direct linear discriminant analysis (DLDA) algorithm was advocated in [19]. Recently, an improved DLDA algorithm was suggested in [20].

Although the idea of the new DLDA [20] is very important, it is still a linear method. It is impossible to solve the face recognition problem using a linear method because of the complex nature of human faces. Since much of the important information may be contained in higher order relationships among the image pixels of a face pattern, the non-linear discriminant analysis, i.e. the kernel discriminant analysis has been an attractive topic in the field of pattern recognition over the last five years [21]-[27]. Inspired by the idea of kernel direct discriminant analysis (KDDA) introduced in [21], this paper develops a new KDDA algorithm which is a kernel version of the DLDA algorithm in [20].

The rest of the paper is organized as follows. The new DLDA algorithm is overviewed in Section 2. The linear case of direct discriminant analysis is then generalized to the nonlinear case. The proposed KDDA algorithm is developed in Section 3. Experimental results are presented in Section 4 and conclusions are drawn in Section 5.

## 2. An overview of the direct linear discriminant analysis (DLDA) algorithm

We assume to have available a set of  $N$  training face images  $z_i$   $i = 1, \dots, N$ . Each image is defined as a vector of length  $M (= I_w \times I_h)$ , i.e.  $z_i \in R^M$ , where  $M$  is the face image size and  $R^M$  denotes an  $M$ -dimensional real space. It is assumed that each image belongs to one of the  $C$  classes denoted by  $\omega_i$   $i = 1, \dots, C$ . The between-class scatter matrix  $S_b$  and the within-class scatter matrix  $S_w$  and the population mean vector  $\mu$  can be obtained from the training samples. The new DLDA algorithm developed in [20] draws on a variant of discriminant analysis criterion and exploits the strength of the DLDA algorithm.

The concrete steps of the new DLDA algorithm can be summarized as follows [20]:

**Input:** A set of training face images  $\{z_i, i = 1, 2, \dots, N\}$ , each of which is represented as a  $M$ -dimensional vector.

**Output:** A low-dimensional representation  $z^*$  of  $z$  with enhanced discriminatory power obtained by transformation  $z^* = A_0 z$ .

**New DLDA Algorithm [20]:**

**Step 1.** Determine the set of eigenvectors of  $\Phi_b^T \Phi_b$  associated with the  $m \leq C - 1$  non-zero eigenvalues:  $F_m = [f_1, \dots, f_m]$ , where  $\Phi_b = [\sqrt{N_1}(\mu_1 - \mu), \dots, \sqrt{N_C}(\mu_C - \mu)]$ , and  $N_i, \mu_i$  are

the number of samples and mean vector of the  $i$ -th class respectively.

**Step 2.** Calculate the first  $m$  most significant eigenvectors and the corresponding eigenvalues of  $S_b$  by

$$Y_0 = \Phi_b F_m \text{ and } \Delta_b = \text{diag}[\tau_1, \tau_2, \dots, \tau_m].$$

**Step 3.** Let  $Z_0 = Y_0 \Delta_b^{-\frac{1}{2}}$ , calculate the eigenvectors  $U_0$  and their corresponding eigenvalues  $\Delta_w$  of  $Z_0^T S_w Z_0$

**Step 4.** Let  $P_0 = Z_0 U_0, Q_0 = P_0 \Delta_w^{-\frac{1}{2}}$ , calculate the first  $m'$  most significant eigenvectors  $V_0$  and their corresponding eigenvalues  $\Delta_t$  of  $Q_0^T S_t Q_0$ , i.e.  $V_0^T Q_0^T S_t Q_0 V_0 = \Delta_t$

**Step 5.** Let  $A_0 = V_0^T Q_0^T = (Q_0 V_0)^T$ , the low dimensional transformed vector  $z^*$  for each testing sample  $z$  is

$$z^* = A_0 z$$

### 3. Kernel version of the DLDA algorithm

#### 3.1 Dimension reduction based on the between-class scatter matrix in the kernel space

A nonlinear mapping is associated with the kernel discriminant analysis. Let  $\phi : z \in R^M \rightarrow \phi(z) \in H$  be a nonlinear mapping from the input space to a high-dimensional feature space  $H$ , where different classes of faces are linearly separable. Denoting the total mean of the all the images and the mean vector of the  $i$ -th class in the kernel space respectively as

$$\bar{\phi} = \frac{1}{N} \sum_{j=1}^N \phi(z_j)$$

$$\bar{\phi}_i = \frac{1}{N_i} \sum_{j=1}^{N_i} \phi(z_j)$$

the between-class scatter matrix  $C_b$  in the kernel space can be rewritten as follows,

$$C_b = \sum_{i=1}^c \left( \sqrt{\frac{N_i}{N}} (\bar{\phi}_i - \bar{\phi}) \right) \left( \sqrt{\frac{N_i}{N}} (\bar{\phi}_i - \bar{\phi}) \right)^T = \sum_{i=1}^c \tilde{\phi}_i \tilde{\phi}_i^T = P_b P_b^T$$

where  $\tilde{\phi}_i = \sqrt{\frac{N_i}{N}} (\bar{\phi}_i - \bar{\phi})$ , and  $P_b = [\tilde{\phi}_1, \dots, \tilde{\phi}_c]$

Unfortunately, it is intractable to solve the eigenvalue problem of  $C_b$  for the large dimensionality of kernel space. However, similarly to the general D-LDA framework developed in [19], we can solve the eigenvalue problem indirectly by finding the eigenvalue solution to  $P_b^T P_b$  first. Computing  $P_b^T P_b$  requires a dot product evaluation in the kernel space which can be done by utilizing the kernel approach [21].

Accordingly, for two arbitrary classes  $\omega_l$  and  $\omega_h$ , we define an  $N_l \times N_h$  dot product matrix  $K_{lh}$  as:  $K_{lh} = (k_{ij})_{\substack{i=1,\dots,N_l \\ j=1,\dots,N_h}}$ , where  $k_{ij} = k(z_{l_i}, z_{h_j}) = \phi_{l_i} \cdot \phi_{h_j}$

which for all the classes  $\{\omega_i, i = 1, 2, \dots, C\}$ , constitutes an  $N \times N$  kernel matrix  $K$ ,

$$K = (K_{lh})_{\substack{l=1,\dots,C \\ h=1,\dots,C}}$$

The matrix allows us to express  $P_b^T P_b$  as follows [21]:

$$P_b^T P_b = \frac{1}{N} B \left( A_{NC}^T K A_{NC} - \frac{1}{N} (A_{NC}^T K I_{NC}) - \frac{1}{N} (I_{NC}^T K A_{NC}) + \frac{1}{N^2} (I_{NC}^T K I_{NC}) \right) B$$

where  $B = \text{diag}[\sqrt{N_1}, \dots, \sqrt{N_C}]$ ,  $I_{NC}$  is an  $N \times C$  matrix with all the terms equal to one,

$A_{NC} = \text{diag}[a_{N_1}, \dots, a_{N_C}]$  is an  $N \times C$  block diagonal matrix, and  $a_{N_i}$  is an  $N_i \times 1$  vector

with all terms equal to  $\frac{1}{N_i}$ .

Let  $\lambda_i$  and  $e_i$  be the  $i$ -th eigenvalue and its corresponding eigenvector of  $P_b^T P_b$   $i = 1, 2, \dots, C$ ,

sorted in the decreasing order of magnitude. The first  $m$  eigenvectors,  $Y = [y_1, y_2, \dots, y_m] = P_b E_m$ ,

whose corresponding eigenvalues are greater than 0, are used, where  $E_m = [e_1, e_2, \dots, e_m]$ .

### 3.2 The New Algorithm of Kernel Direct Discriminant Analysis

Let  $D_b = \text{diag}[\lambda_1, \dots, \lambda_m]$ , further let  $Z = Y D_b^{-\frac{1}{2}}$  denote a projection matrix. Transforming the within-class scatter matrix  $C_w$  into the subspace spanned by  $Z$ , we have

$$\tilde{C}_w = Z^T C_w Z$$

Note that  $\tilde{C}_w$  is a tractable matrix with size  $m \times m$ . Let  $D_w$  be a diagonal matrix of eigenvalues

and further let  $U$  be a full matrix of eigenvevectors of matrix  $\tilde{C}_w$ . Clearly,

$$\left(UD_w^{-\frac{1}{2}}\right)^T \tilde{C}_w \left(UD_w^{-\frac{1}{2}}\right) = I$$

or

$$\left(UD_w^{-\frac{1}{2}}\right)^T \left(P_b^T E_m D_b^{-\frac{1}{2}}\right)^T C_w \left(P_b^T E_m D_b^{-\frac{1}{2}}\right) \left(UD_w^{-\frac{1}{2}}\right) = I$$

where  $I$  is an identity matrix.

Let  $P = ZU = P_b E_m D_b^{-\frac{1}{2}} U$  and  $Q = PD_w^{-\frac{1}{2}}$ . Projecting the within-class scatter matrix  $C_w$  and population scatter matrix  $C_t$  into the subspace spanned by  $Q$ , we have  $Q^T C_w Q = I$  and

$$\begin{aligned} \tilde{C}_t &= Q^T C_t Q = Q^T C_b Q + Q^T C_w Q \\ &= \left(UD_w^{-\frac{1}{2}}\right)^T \left(E_m D_b^{-\frac{1}{2}}\right)^T P_b^T C_b P_b \left(E_m D_b^{-\frac{1}{2}}\right) \left(UD_w^{-\frac{1}{2}}\right) + I \\ &= \left(UD_w^{-\frac{1}{2}}\right)^T \left(E_m D_b^{-\frac{1}{2}}\right)^T \left(P_b^T P_b\right) \left(P_b^T P_b\right) \left(E_m D_b^{-\frac{1}{2}}\right) \left(UD_w^{-\frac{1}{2}}\right) + I \end{aligned}$$

Therefore, we can obtain a kernel representation of  $\tilde{C}_t$  using the kernel representation of  $P_b^T P_b$ .

In order to maximize the Fisher criterion function  $\frac{|\Phi^T \tilde{C}_t \Phi|}{|\Phi^T \tilde{C}_w \Phi|}$  [28]-[29], we need only select the

eigenvectors of  $\tilde{C}_t$  corresponding to the largest eigenvalues. We denote the selected eigenvectors as

$V = [v_1, v_2, \dots, v_{m'}]$ , and the corresponding diagonal matrix of eigenvalues as  $D_t$ . Note that  $D_t$  is an  $m' \times m'$  diagonal matrix. The optimal kernel discriminant feature extractor can be derived through  $A = V^T Q^T$ .

According to the above analysis, the steps of the new KDDA algorithm can be summarized as follows:

**Input:** The same as that of DLDA algorithm.

**Output:** A formal representation of the kernel feature space with enhanced discriminatory power

**KDDA Algorithm:**

**Step 1.** Calculate the kernel matrix  $K$ .

**Step 2.** Calculate the kernel representation of  $P_b^T P_b$  and compute the eigenvectors of  $P_b^T P_b$  with non-zero eigenvalues:

$$E_m = [e_1, \dots, e_m], \text{ where } m \leq c-1 \text{ and } P_b \text{ relates to } C_b = P_b P_b^T$$

**Step 3.** Calculate the first  $m$  most significant eigenvectors of  $S_b$  by  $Y = P_b E_m$  and denote their corresponding eigenvalues as  $D_b = \text{diag}[\lambda_1, \lambda_2, \dots, \lambda_m]$

**Step 4.** Let  $Z = YD_b^{-\frac{1}{2}}$ . Calculate the kernel representation of  $Z^T C_w Z$  and the eigenvectors  $U$  and their corresponding eigenvalues  $D_w$  of  $Z^T C_w Z$

**Step 5.** Let  $P = ZU, Q = PD_w^{-\frac{1}{2}}$ . Calculate the first  $m'$  most significant eigenvectors  $V$  of  $\tilde{C}_t = Q^T C_t Q$ . The kernel feature space is then given as  $A = V^T Q^T$ .

### 3.3 The Design of a Classifier in the Kernel Feature Space

It should be noted that the kernel space cannot be determined explicitly as we do not have a representation of  $P_b$  on which  $A$  is dependent. However, this problem can be solved at the stage of classifier design using once again the kernel technique.

We illustrate the design process using the minimum distance classifier as an example.

Let us project all the samples mapped by the kernel function into the kernel feature subspace spanned by  $A$ . i.e.

$$x_i = A\phi(z_i) = V^T \left( UD_w^{-\frac{1}{2}} \right)^T \left( E_m D_b^{-\frac{1}{2}} \right)^T P_b^T \phi(z_i)$$

where  $P_b^T \phi(z_i) = \frac{1}{\sqrt{N}} B \left( A_{NC}^T \gamma(\phi(z_i)) - \frac{1}{N} \mathbf{I}_{NC}^T \gamma(\phi(z_i)) \right)$

and  $\gamma(\phi(z_i)) = [\phi_{11}^T \phi(z_i), \phi_{12}^T \phi(z_i), \dots, \phi_{CN_C}^T \phi(z_i)]^T$  is an  $N \times 1$  kernel vector.

Thus,  $x_i = X\gamma(\phi(z_i))$ , where  $X = \frac{1}{\sqrt{N}} \left( E_m D_b^{-\frac{1}{2}} \cdot UD_w^{-\frac{1}{2}} \right)^T \left( B \left( A_{NC}^T - \frac{1}{N} \mathbf{I}_{NC}^T \right) \right)$

The population mean vector  $m$  is then

$$m = \frac{1}{N} \sum_{i=1}^N x_i$$

The projected samples, after centralization, can be represented as

$$\tilde{x}_i = x_i - m$$

The mean vector of  $j$ -th class can be determined by

$$m_j = \frac{1}{N_j} \sum_{k=1}^{N_j} \tilde{x}_k \quad \tilde{x}_k \in \omega_j$$

For any test sample  $z$ , centralizing after projecting  $\phi(z)$  into  $A$ , we have

$$y = A\phi(z) - m$$

The minimum distance classifier then becomes

$$z \in \omega_i \quad \text{iff} \quad \|y - m_i\| = \min_{j=1,2,\dots,C} \|y - m_j\|$$

## 4. Experimental Results

### 4.1 Experiments on ORL

The Olivetti Research Lab (ORL) face database (<http://www.cam-orl.co.uk/facedatabase.html>) can be used freely for academic research. The Cambridge ORL database contains 40 distinct subjects, each subject having ten different images, taken at different times. The images are subject to varying lighting, changing facial expressions (open/closed eyes, smiling/nonsmiling), and facial details (glasses/no glasses). All the images are taken against a dark homogeneous background and the persons are in upright, frontal position (with a tolerance for some side movement). Four images per subject are chosen as training samples, the remaining six images are treated as test samples.

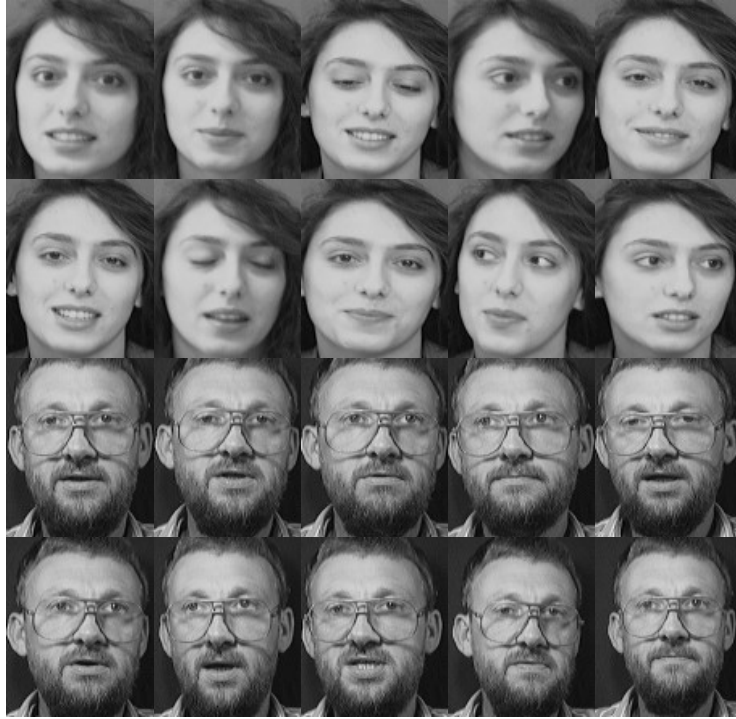


Fig.1 Part of the images of ORL

The experiments are made on ORL images of different sizes. Table 1 shows the performance comparison of the present algorithm and the algorithm in [21] both employing the polynomial kernel function  $k(x_i, x_j) = (a(x_i \cdot x_j) + b)^d$  when  $a = b = 1$ ,  $d = 2$ . In contrast table 2 compares the two algorithms in the case of a linear kernel function with  $a = b = d = 1$ . The experimental results show that the two algorithms exhibit a similar performance on ORL which is a comparatively small database. It is interesting to note that the results of the new algorithm on ORL are exactly the same as those reported in [20] when a linear kernel function is adopted.

Table 1 The performance comparison of two algorithms

| Size of image   | Ref [21] (%) | New algorithm (%) |
|-----------------|--------------|-------------------|
| $112 \times 92$ | 90.8         | 90.8              |
| $56 \times 46$  | 91.3         | 91.3              |
| $28 \times 23$  | 90.4         | 90.4              |
| $14 \times 12$  | 90.4         | 90.4              |
| $7 \times 6$    | 77.9         | 77.9              |

Table 2 The performance comparison of the two algorithms based on a linear kernel function

| Size of image   | Ref [21] (%) | New algorithm (%) |
|-----------------|--------------|-------------------|
| $112 \times 92$ | 90.8         | 90.8              |
| $56 \times 46$  | 89.6         | 89.6              |
| $28 \times 23$  | 90.8         | 90.8              |
| $14 \times 12$  | 90.8         | 90.8              |
| $7 \times 6$    | 76.7         | 76.7              |

## 4.2 Experiments on XM2VTS

The XM2VTS database is a multi-modal database consisting of face images, video sequences and speech recordings taken of 295 subjects at one month intervals. Since the data acquisition was distributed over a long period of time, significant variability of appearance of clients, e.g. changes of hair style, facial hair, shape and presence or absence of glasses, is present in the recordings. The XM2VTS database contains 4 sessions with 2 shots at each session [1]. The facial images of size  $55 \times 51$  were extracted from the XM2VTS with reference to manually annotated eye coordinates. Figure 2 shows examples of such manually extracted images. One half of the images are used for training and another half for testing. Table 3 shows a performance comparison of the present algorithm and the algorithm in [21] both employing the polynomial kernel function with  $a=1e-8, b=1, \text{ and } d=2$ . Table 4 shows the performance comparison of the two algorithms based on the linear kernel function when  $a=1e-8, b=1, \text{ and } d=1$ . The experimental results show that the proposed algorithm outperforms the previous one in terms of recognition accuracy. It is also interesting to note that the results of the new algorithm on XM2VTS are very similar to those reported in [20] when linear kernel function is adopted.



Figure 2 Part of the automatic extracted images in XM2VTS

Table 3 The performance comparison of two algorithms on XM2VTS

| Size of image  | Ref [21] (%) | New algorithm (%) |
|----------------|--------------|-------------------|
| $55 \times 51$ | 87.6         | 92.5              |
| $28 \times 26$ | 81.3         | 91.4              |
| $19 \times 17$ | 77.5         | 90.3              |
| $9 \times 8$   | 60.9         | 71.7              |



Table 4 The performance comparison of two algorithms on XM2VTS for linear kernel function

| Size of image | Ref [21] (%) | New algorithm (%) |
|---------------|--------------|-------------------|
| 55 × 51       | 83.5         | 91.3              |
| 28 × 26       | 78.9         | 91.2              |
| 19 × 17       | 75.9         | 89.7              |
| 9 × 8         | 60.7         | 79.5              |

## 5. Conclusions

In this paper, we have developed a novel kernel direct discriminant analysis (KDDA) algorithm which is a nonlinear version of the recently proposed DLDA algorithm. The results of the face recognition experiments show that the performance the new KDDA algorithm is very similar to that of the DLDA algorithm provided a linear kernel function is adopted. This confirms that the new KDDA algorithm generalizes DLDA algorithm. The research issues for future work include: (1) a study of the effect of the choice of kernel function on classifier performance. (2) an investigation of the trade off between the type of kernel function, number of classes, number of training samples, and the dimensionality of the feature space. (3) applications to other problems including face verification.

## Acknowledgements

This work was supported in part by the following projects: EU Project Vampire, National Natural Science Foundation of P. R. China (Grant No.: 60072034), Robotics Laboratory, Shenyang Institute of Automation, Chinese Academy of Sciences foundation (Grant No.: RL200108), Natural Science research program of University of Jiangsu Province, P. R. China (Grant No.: 01KJB520002), Natural Science Foundation of Jiangsu Province (Grant No.:BK2002001), and Open Foundation of Image Processing and Image Communication Lab (Grant No.: KJS03038).

## References

- [1] J.Kitter, and M.S.Nixon, Audio-and Video-Based Biometric Person Authentication, 4<sup>th</sup> International Conference, AVBPA2003, Guildford, UK, June 2003
- [2] R.Chellappa, C. L. Wilson, S.Sirohey S. Human and machine recognition of faces: a survey. Proceedings of the IEEE, 83(5): 705-740, 1995.
- [3] M.Turk, A.Pentland, Eigenfaces for recognition. J. Cognitive Neuroscience, 3(1): 71-86, 1991.
- [4] Z.Q.Hong, Algebraic feature extraction of image for recognition. Pattern Recognition,24(3):211-219, 1991.
- [5] Z. Q. Hong, Jing-Yu Yang, Optimal discriminant plane for a small number of samples and design method of classifier on the plane. Pattern Recognition, 24(4): 317-314, 1991.
- [6] Y.Q.Cheng, Y. M. Zhuang, J.Y.Yang. Optimal Fisher discriminant analysis using the rank decomposition. Pattern Recognition, 25(1):101-111,1992.
- [7] D.H.Foley, J.W.Sammon. An optimal set of discriminant vectors. IEEE Trans. Computer, 24(3): 281-289, 1975
- [8] J.W.Sammon. An Optimal discriminant plane. IEEE Trans. Computer, 19(3): 826-829, 1970
- [9] J. Kittler, On the discriminant vector method of feature selection. IEEE Trans. Comput., 26(6): 604-606, 1977.

- [10] P.N.Belhumeur, J.P.Hespanha, and D.J.Kriegman. Eigenfaces vs. fisherfaces: Recognition using class specific linear projection. *IEEE Trans. on PAMI*, 19(7): 711-720,1997
- [11] W. Zhao, A. Krishnaswamy, R. Chellappa, D. Swets, and J. Weng, "Discriminant analysis of principal components for face recognition," *Face Recognition: From Theory to Applications*, Eds. H. Wechsler, P.J. Phillips, V. Bruce, F.F. Soulie and T.S. Huang, Springer-Verlag, pp. 73-85, 1998.
- [12] D. L.Swets and J. Weng. Using discriminant eigenfeatures for image retrieval. *IEEE Trans. PAMI*, 1996, 18(8): 831-836.
- [13] K.Liu, Y.Q.Cheng, J.Y.Yang . An efficient algorithm for Foley-Sammon optimal set of discriminant vectors by algebraic method. *International Journal of Pattern Recognition and Artificial intelligence*, 6(5): 817-829,1992
- [14] K.Liu, Y.Q.Cheng, J.Y.Yang. A generalized optimal set of discriminant vectors. *Pattern Recognition*, 25(1): 731-739, 1992
- [15] Yuefei Guo, Shijin Li, Jingyu Yang, et al. A generalized Foley-Sammon transform based on generalized fisher discriminant criterion and its application to face recognition. *Pattern Recognition Letters*, 2003, vol.24, Issues 1-3, pp147-158
- [16] Xiaojun Wu, Jingyu Yang, Shitong Wang, et al., A new algorithm for solving optimal discriminant vectors. *Journal of Computer Science and Technology*, vol.19, no.3, pp.324~330, 2002
- [17] Xiaojun Wu, Josef Kittler, Jingyu Yang, etc. An analytical algorithm for determining the generalized optimal set of discriminant vectors. *Pattern Recognition*, vol.37, no.9, pp.1949-1952
- [18] Li-Fen Chen, Hong-Yuan Mark Liao, Ming-Tat Ko, etc. A new LDA-based face recognition system which can solve the small sample size problem. *Pattern Recognition*, 2000, 33(9):1713-1726
- [19] Hua Yu, Jie Yang. A direct LDA algorithm for high-dimensional data-with application to face recognition. *Pattern Recognition*, vol.34, pp2067-2070, 2001
- [20] Xiaojun Wu, Josef Kittler, Jingyu Yang, etc. A new direct LDA (DLDA) algorithm for face recognition, *ICPR 2004*, Cambridge, UK, August, 2004
- [21] Juewei Lu, K.N.Platanotis, and A.N. Venetsanopoulos. Face recognition using kernel direct discriminant analysis algorithms, *IEEE Trans. NN*, vol.14, No.1, pp.117-126, 2003
- [22] B.Scholkopf, A.Smola, and K.R.Muller, Nonlinear component analysis as a kernel eigenvalue problem, *Neural Computation*, vol.10, pp.1299-1319, 1998
- [23] S.Mika, G.Ratsch, J.Weston, B.Scholkopf, and K.R.Muller, "Fisher discriminant analysis with kernels," in *Neural Networks for Signal Processing IX*, Y.-H. Hu, J.Larsen, E.Wilson, and S.Douglas, Eds. Piscataway, NJ: IEEE Press, 1999, pp.41-48
- [24] G.Baudat and F.Anouar, Generalized discriminant analysis using a kernel approach, *Neural Computation*, vol.12, pp.2385-2404, 2000
- [25] B.Scholkopf, A.Smola, *Learning with kernels*. Cambridge, MA: MIT Press, 2002.
- [26] Sebastian Mika, Gunnar Ratsch, and Jason Weston, et al. Constructing descriptive and discriminative nonlinear features: Rayleigh coefficients in kernel feature spaces, *IEEE Trans. PAMI*, Vol.25, No.5, 2003
- [27] Ming-Hsuan Yang, Narendra Ahuja and David Kriegman, "Face Recognition Using Kernel Eigenfaces", in *Proceedings of the 2000 IEEE International Conference on Image Processing (ICIP 2000)* **1**, pp. 37-40, Vancouver, Canada, September, 2000.
- [28] P.Devijver and J.Kittler. *Pattern recognition: A statistical approach*, Prentice-Hall, London, 1982
- [29] K.Fukunaga, *Introduction to Statistical Pattern recognition*, Academic Press, New York, 1990

This document was created with Win2PDF available at <http://www.daneprairie.com>.  
The unregistered version of Win2PDF is for evaluation or non-commercial use only.