Subspace KDA Algorithm for Non-linear Feature Extraction in Face Identification

Wen-Sheng Chen[1], Pong C. Yuen[2], and Jian-Huang Lai[3]

[1] College of Mathematics and Computational Science, Shenzhen University, China, 518060
chenws@szu.edu.cn
[2] Department of Computer Science, Hong Kong Baptist University, Hong Kong, China
pcyuen@comp.hkbu.edu.hk
[3] Department of Electronics & Communication Engineering, Sun Yat-Sen University, Guangzhou, China, 510275
stsljh@mail.sysu.edu.cn

Abstract. Kernel discriminant analysis (KDA) method is a promising approach for non-linear feature extraction in face identification tasks. However, as a linear algorithm to address nonlinear problem, Fisher discriminant analysis (FDA) approach will not give a satisfactory performance. Moreover, FDA usually suffers from small sample size (S3) problem. To overcome these two shortcomings in FDA method, Shannon wavelet kernel based subspace FDA (SKDA) algorithm is developed in this paper. Two public databases such as FERET and CMU PIE databases are selected for evaluation. Comparing with the existing kernel based FDA-based methods, the proposed method gives superior results.

Keywords: Face identification, Kernel discriminant analysis, Shannon wavelet.

1 Introduction

Over the past decade, Fisher discriminant analysis [1] method has been shown to be an effective approach in face identification tasks and its superior performance has been reported in many literatures [1]-[11] . FDA is theoretically sound and its objective is to find the most discriminant feature for pattern classification. However, there are two major limitations upon FDA approach. First, it is a linear method and is hard to solve nonlinear problem, while the second is the small sample size (S3) problem, which always occurs when the sample size is smaller than the dimensionality of feature vector.

KDA is a useful approach to deal with nonlinear problem. The basic idea of KDA is to apply a nonlinear mapping $\Phi : x \in R^d \rightarrow \Phi(x) \in F$ to the input data vector x in input space R^d and then to perform the FDA on the mapped

Y. Wang, Y. Cheung, and H. Liu (Eds.): CIS 2006, LNAI 4456, pp. 1106–1114, 2007.

higher dimension feature space F. The feature space F could be considered as a linearization space. By utilizing kernel trick, the inner products $\langle \Phi(x_i), \Phi(x_j) \rangle$ in F can be replaced with a Mercer kernel function $K(x_i, x_j)$, i.e. $K(x_i, x_j) = \langle \Phi(x_i), \Phi(x_j) \rangle = \Phi(x_i)^T \cdot \Phi(x_j)$, where x_i, x_j are input pattern column vectors in input space R^d. So the nonlinear mapping Φ can be performed implicitly in input space R^d. This paper exploits Shannon wavelet kernel method [9] to address the nonlinear problems such as pose and illumination variations in face identification, while subspace FDA (SFDA) algorithm [6] will be used to solve S3 problem.

Therefore, combining Shannon wavelet kernel with subspace FDA method, we design and develop a novel Shannon wavelet kernel-based subspace FDA algorithm (SKDA) in this paper. Two public databases such as FERET and CMU PIE databases are selected for evaluation. Comparing with the existing kernel based FDA-based methods, the proposed method gives the best performance.

The rest of this paper is organized as follows. Section 2 briefly describes the Shannon wavelet based Mercer kernel function. The proposed SKDA algorithm is developed and evaluated in section 3 and section 4 respectively. Finally, Section 5 draws the conclusions.

2 Mercer Kernel Function Based on Shannon Wavelet

This section briefly reviews on Shannon wavelet based Mercer kernel function. Details can be found in paper [9].

Assume

$$\cdots \subset V_{-2} \subset V_{-1} \subset V_0 \subset V_1 \subset V_2 \subset \cdots$$

is the multiresolution analysis (MRA) [12,13] generated by the Shannon sampling function $\phi(x) = \mathrm{sinc}(x) := \frac{\sin \pi x}{\pi x}$. Thus the scaling subspaces

$$V_j = \{ f \in L^2(R) | \mathrm{supp} \hat{f} \subset [-2^j \pi, 2^j \pi] \}$$

and the corresponding wavelet subspaces $\{W_j\}_{j \in Z}$, where $W_j \perp V_j$ and $W_j \oplus V_j = V_{j+1}$, can be generated by the Shannon wavelet:

$$\psi(x) := 2\mathrm{sinc}(2x) - \mathrm{sinc}(x), \tag{1}$$

whose Fourier transform is given by $\hat{\psi}(\xi) = \chi_{[-2\pi, -\pi] \cup [\pi, 2\pi]}(\xi)$, where $\chi(x)$ is an indicator function. Let $H(\xi)$ and $G(\xi)$ be the 2π-periodic functions respectively as:

$$H(\xi) = \begin{cases} 1, & \xi \in [-\frac{\pi}{2}, \frac{\pi}{2}) \\ 0, & \xi \in [-\pi, -\frac{\pi}{2}) \cup [\frac{\pi}{2}, \pi) \end{cases} \text{ and } G(\xi) = H(\xi + \pi).$$

Then, we have

$$\hat{\phi}(\xi) = H(\xi/2)\hat{\phi}(\xi/2), \quad \hat{\psi}(\xi) = G(\xi/2)\hat{\phi}(\xi/2).$$

It is easy to see that $\phi(x)$ is an orthonormal scaling function and $\psi(x)$ is an orthonormal wavelet.

The necessary and sufficient condition for a translation invariant function $k(x, y) = k(x - y)$ to be a Mercer kernel is that its Fourier transform is non-negative. Under this consideration, the Shannon wavelet Mercer kernel $k(x, y)$ defined on $R^d \times R^d$ can be constructed as follows:

$$k(x, y) = \frac{1}{d} \sum_{i=1}^{d} [\psi((x_i - y_i)/\theta_i)]^p, \tag{2}$$

where $\psi(x)$ is Shannon wavelet function defined in (1), $x = (x_1, \cdots, x_d)^T, y = (y_1, \cdots, y_d)^T \in R^d$, $p \in Z^+$ and $\theta_i > 0\,(i = 1, \cdots, d)$ are kernel parameters.

3 Proposed SKDA Algorithm

This section reports Shannon wavelet kernel based subspace FDA method for face identification. Details are discussed as follows.

3.1 Some Notations

Let d and C be the dimensionality of original sample feature space and the number of sample classes respectively. The total original samples $X = \{X_1, X_2, \cdots, X_C\}$, the jth class X_j contains N_j samples, namely $X_j = \{x_1^j, x_2^j, \cdots, x_{N_j}^j\}$, $j = 1, 2, \cdots, C$. Let N be the total number of original training samples, then $N = \sum_{j=1}^{C} N_j$. If $\Phi(x)$: $x \in R^d \to \Phi(x) \in F$ is the kernel nonlinear mapping, where F is the mapped feature space, denote $df = \dim F$, the total mapped sample set and the jth mapped class are given by $\Phi(X) = \{\Phi(X_1), \Phi(X_2), \cdots, \Phi(X_C)\}$ and $\Phi(X_j) = \{\Phi(x_1^j), \Phi(x_2^j), \cdots, \Phi(x_{N_j}^j)\}$ respectively. Also, the mean of the mapped sample class $\Phi(X_j)$ and the global mean of the total mapped sample $\Phi(X)$ are given by $m_j = \frac{1}{N_j} \sum_{x \in X_j} \Phi(x)$ and $m = \frac{1}{N} \sum_{j=1}^{C} \sum_{x \in X_j} \Phi(x)$ respectively. In feature space F, the within-class scatter matrix S_w^Φ, between-class scatter matrix S_b^Φ and total scatter matrix S_t^Φ are defined respectively as:

$$S_w^\Phi = \frac{1}{N} \sum_{j=1}^{C} \sum_{x \in X_j} (\Phi(x) - m_j)(\Phi(x) - m_j)^T = \Phi_w \Phi_w^T,$$

$$S_b^\Phi = \frac{1}{N} \sum_{j=1}^{C} N_j (m_j - m)(m_j - m)^T = \Phi_b \Phi_b^T,$$

$$S_t^\Phi = \frac{1}{N} \sum_{j=1}^{C} \sum_{x \in X_j} (\Phi(x) - m)(\Phi(x) - m)^T = \Phi_t \Phi_t^T,$$

where Φ_w, $\Phi_t \in R^{d \times N}$ and $\Phi_b \in R^{d \times C}$.

The Fisher index $J_\Phi(W)$ in mapped feature space F is defined by

$$J_\Phi(W) = \det(W^T S_w^\Phi W)\,[\det(W^T S_b^\Phi W)]^{-1}, \tag{3}$$

where $W \in F^{df \times m}$. The objective of FDA is used to find a optimal projection in mapped feature space F that minimizes within-class distance and simultaneously maximizes between-class distance.

3.2 SKDA Strategy

Let $S_{wt}^{\Phi} = \Phi_w^T \Phi_t \in R^{N\times N}$. By performing singular value decomposition, there exist two orthonormal matrices $U, V \in R^{N\times N}$ and a diagonal matrix $\Lambda = \mathrm{diag}\{\sigma_1, \cdots, \sigma_r, 0, \cdots, 0\} \in R^{N\times N}$ with $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_r > 0$, such that $S_{wt}^{\Phi} = U\Lambda V^T$. Since $S_w^{\Phi} = \Phi_w \Phi_w^T$, we have $(\Phi_t V)^T S_w (\Phi_t V) = \mathrm{diag}\{\sigma_1^2, \cdots, \sigma_r^2, 0, \cdots, 0\} \in R^{N\times N}$. Rewrite the term $\Phi_t V$ as $\Phi_t V = [y_1, y_2, \cdots, y_r, y_{r+1}, \cdots, y_N]_{df\times N}$, where y_i is the ith column of $(\Phi_t V)_{df\times N}$. Denote $Y = [y_{r+1}, y_{r+2}, \cdots, y_N]$. It can be seen that Y is a $df \times (N-r)$ matrix and satisfies $Y^T S_w^{\Phi} Y = 0_{(N-r)\times(N-r)}$.

Having determined the null subspace of S_w^{Φ}, the projection is then determined in the subspace $\overline{N(S_b^{\Phi})}$, the complementary subspace of S_b^{Φ}. Thus, the second step is to discard the null space of S_b^{Φ} to ensure that the numerator of the Fisher index will not be zero.

To this end, we define $\hat{S}_b = Y^T S_b^{\Phi} Y$, and then $\hat{S}_b = (Y^T \Phi_b)(Y^T \Phi_b)^T$, $Y^T \Phi_b$ is a $(N-r)\times C$ matrix. By singular value decomposition, there exist two orthonormal matrices $(U_b)_{(N-r)\times(N-r)}$ and $(V_b)_{C\times C}$, such that $Y^T \Phi_b = U_b \Lambda_b V_b^T$, where $\Lambda_b = \begin{bmatrix} \Sigma_b \\ 0 \end{bmatrix} \in R^{(N-r)\times C}$ and $\Sigma_b = \mathrm{diag}\{\tau_1, \cdots, \tau_m, 0, \cdots, 0\} \in R^{C\times C}$ with $\tau_1 \geq \tau_2 \geq \cdots \geq \tau_m > 0$. Rewrite $U_b = [u_1, \cdots, u_m, u_{m+1}, \cdots, u_{N-r}] \in R^{(N-r)\times(N-r)}$ and denote $A = [u_1, \cdots, u_m]_{(N-r)\times m}$ and $D_m = \mathrm{diag}\{\tau_1, \cdots, \tau_m\}$, we have $A^T \hat{S}_b A = D_m^2$, namely, $(YA)^T S_b (YA) = D_m^2$. Let $W = (YA)_{df\times m}$, then

$$W^T S_w^{\Phi} W = 0_{m\times m}, \quad W^T S_b^{\Phi} W = D_m^2.$$

Thereby, W is the optimal SKDA projection matrix, by which the Fisher index $J(W)$ (3) reaches maximum.

3.3 SKDA Algorithm Design

Based on above analysis, the proposed SKDA algorithm is designed as follows.

Step 1: Compute the $N \times N$ matrix $S_{wt} = \Phi_w^T \Phi_t$ via the following formula:

$$\Phi_w^T \Phi_t = [N \cdot \mathbf{K} - \mathbf{K} \cdot \mathbf{1}_{NN} - N \cdot \Lambda_N \cdot \mathbf{K} + \Lambda_N \cdot \mathbf{K} \cdot \mathbf{1}_{NN}]/N^2,$$

where the kernel matrix $\mathbf{K} = \left(k(x_j^i, x_k^l)\right)_{j=1,\cdots,N_i;k=1,\cdots,N_l}^{i=1,\cdots,C;l=1,\cdots,C}$, $\mathbf{1}_{mn}$ denotes a $m \times n$ matrix with all terms equal to 1, $\Lambda_N = \mathrm{diag}[\Lambda_{N_1}, \cdots, \Lambda_{N_C}]$ is a N by N block diagonal matrix, and Λ_{N_i} is a $N_i \times N_i$ matrix with all terms equal to $1/N_i$, $i = 1, \cdots, C$.

Step 2: Perform singular value decomposition $S_{wt} \overset{svd}{=} U\Lambda V^T$, where $U, V \in R^{N\times N}$ are two orthonormal matrices, $\Lambda = \mathrm{diag}[\sigma_1, \cdots, \sigma_r, 0, \cdots, 0] \in R^{N\times N}$ with $\sigma_1 \geq \cdots \geq \sigma_r > 0$.

Step 3: Rewrite $V = [v_1, \cdots, v_r, v_{r+1}, \cdots, v_N]$ and denote $\tilde{V} = [v_r, v_{r+1}, \cdots, v_N] \in R^{N\times(N-r)}$ and $Y = (\Phi_t \tilde{V})_{df\times(N-r)}$.

Step 4: Compute $Z = (Y^T \Phi_b)_{(N-r)\times C} = \tilde{V}^T \Phi_t^T \Phi_b$, it yields that

$$\Phi_t^T \Phi_b = [N \cdot \mathbf{K} \cdot D_C - \mathbf{K} \cdot \mathbf{1}_{NC} \cdot D - \mathbf{1}_{NN} \cdot \mathbf{K} \cdot D_C + \frac{1}{N} \cdot \mathbf{1}_{NN} \cdot \mathbf{K} \cdot \mathbf{1}_{NC} \cdot D]/N^2,$$

where $\mathbf{K}$ is the kernel matrix, $D_C = \mathrm{diag}[D_{N_1}, \cdots, D_{N_C}]$ and D_{N_i} is a $N_i \times 1$ matrix with all terms equal to $1/\sqrt{N_i}$ $(i = 1, \cdots, C)$, $D = \mathrm{diag}[\sqrt{N_1}, \cdots, \sqrt{N_C}]$.

Step 5: If the norm of one row in the matrix Z is too small (say less than $1e{-}6$), then discard this row in matrix Z. Accordingly, discard the corresponding column in matrix $\tilde{V}$. Denote the modified matrices Z,$\tilde{V}$and Y as Z',$\tilde{V}'$ and Y' respectively, then do singular value decomposition $Z' \stackrel{svd}{=} U_b \Lambda_b V_b^T$, where U_b and V_b all are orthonormal matrices, $\Lambda_b = (\sum_b, 0)^T \in R^{(N-r')\times C}$ and $\sum_b = \mathrm{diag}[\tau_1, \cdots, \tau_s, 0, \cdots, 0]_{C\times C}$. Usually, $s = C - 1$.

Step 6: Rewrite $U_b = [u_1, \cdots, u_s, u_{s+1}, \cdots, u_{N-r'}]$, where u_i is the ith column of orthonormal matrix U_b. Denote $A = (u_1, \cdots, u_s)_{(N-r')\times s}$ and $W = (Y'A)_{df\times s}$, then W is the SKDA optimal projection matrix.

Step 7: For any testing sample $\Phi(x)$, we get its enhanced feature vector as

$$\begin{aligned}(W^T \Phi(x))_{s\times 1} &= (Y'A)^T \Phi(x) = A^T \tilde{V}'^T \Phi_t \cdot \Phi(x) \\ &= \frac{1}{\sqrt{N}} A^T \tilde{V}'^T [\Phi(x_j^i)^T \cdot \Phi(x) - m^T \cdot \Phi(x)]_{N\times 1}^{1\le i\le C; 1\le j\le N_i},\end{aligned}$$

where $\Phi(x_j^i)^T \cdot \Phi(x) = k(x_j^i, x)$ and $m^T \cdot \Phi(x) = \frac{1}{N} \sum_{i=1}^{C} \sum_{j=1}^{N_i} k(x_j^i, x)$.

---

4 Experimental Results

In this section, two popular and available human face databases, namely FERET and CMU PIE databases, are selected to evaluate the performance of our proposed SKDA algorithm. In the following experiments, Shannon wavelet kernel is with parameters $(p, \theta) = (1, 3.5)$.

4.1 Face Image Databases

For FERET database, we select 120 people, 6 images for each individual. Face image variations in FERET database include pose, illumination, facial expression and aging. Images from one individual are shown in Figure 1.

CMU PIE face database includes totally 68 people. There are 13 pose variations ranged from full right profile image to full left profile image and 43 different lighting conditions, 21 flashes with ambient light on or off. In our experiment, for each people, we select 56 images including 13 poses with neutral expression and 43 different lighting conditions in frontal view. Several images of one people are shown in Figure 2.

For all images in above two face databases, the following preprocessing steps are performed.

Fig. 1. Images of one person from FERET database

Fig. 2. Images of one person from CMU PIE face database

- All images are aligned with the centers of eyes and mouth. The orientation of face is adjusted (on-the-plane rotation) such that the line joining the centers of eyes is parallel with x-axis.
- All the original images with resolution 112x92 are reduced to wavelet feature faces with resolution 30x25 after two-level D4 wavelet decomposition [14].
- All training and testing samples in above two face databases are further normalized as follows:

$$x^* = \frac{x - mean(x)}{std(x)},$$

where x is a sample vector for training or testing, $mean(x)$ is the expectation of x and $std(x)$ is the standard deviation of x.

4.2 Results on FERET Database

This subsection reports the results of proposed SKDA method on FERET face database. We randomly select n (n=2 to 5) images from each people for training, while the rest $(6-n)$ images of each individual are selected for testing. The experiments are repeated 10 times and the average accuracies are recorded in table 1 and shown in the Figure 3. The identification rate of SKDA method increases from 75.90% with training number 2 to 93.92% with training number 5. While for RBF kernel based SFDA (RKDA), the corresponding identification accuracies increase from 71.35% with training number 2 to 92.08% with training number 5 respectively. Comparing with other kernel-based FDA methods, namely, GDA [2] with

Table 1. Comparison on different algorithms on FERET database

TN	2	3	4	5
GDA [2]	71.27%	82.31%	87.71%	92.58%
KDDA [4]	69.45%	82.69%	88.04%	93.25%
RKDA	71.35%	81.06%	87.00%	92.08%
SKDA	75.90%	84.75%	90.29%	93.92%

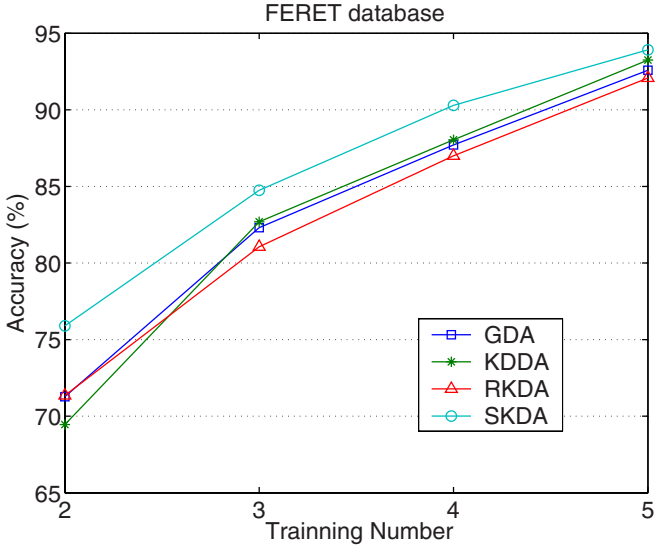

Fig. 3. Performance on FERET face database

RBF kernel and KDDA [4] with RBF kernel, the identification accuracies of GDA and KDDA methods increase from 69.45% and 71.27% with training number 2 to 93.25% and 92.58% with training number 5 respectively.

The results show our proposed SKDA methods gives the best performance for all cases on FERET database.

4.3 Results on CMU PIE Face Database

The subsection reports the results of the proposed SKDA method on CMU PIE database. We randomly select 14 images from each people for training (14×68=952 images for training), while the rest of images of each individual are selected for testing (42×68=2856 images for testing). The experiments are repeated 10 times and the average accuracies of rank 1 to rank 4 are recorded and shown in table 2 and plotted in Figure 4. The identification rate of proposed SKDA method increases from 78.31% with rank 1 to 83.31% with rank 4, while the identification accuracies of RKDA algorithm increase from 77.79% with rank 1 to 82.63% with rank 4. Comparing with other kernel-based methods, namely,

GDA [2] with RBF kernel and KDDA [4] with RBF kernel, the identification rates of GDA and KDDA methods increase from 77.86% and 77.64% with rank 1 to 83.22% and 83.07% with rank 4 respectively. The results demonstrate that our proposed SKDA method gives better performance on CMU PIE database.

Table 2. Comparison on different algorithms on CMU PIE database

Rank	1	2	3	4
GDA [2]	77.86%	80.49 %	82.10 %	83.22%
KDDA [4]	77.64%	80.39%	81.95%	83.07%
RKDA	77.79%	80.26%	81.67%	82.63%
SKDA	78.31%	80.57%	82.11%	83.31%

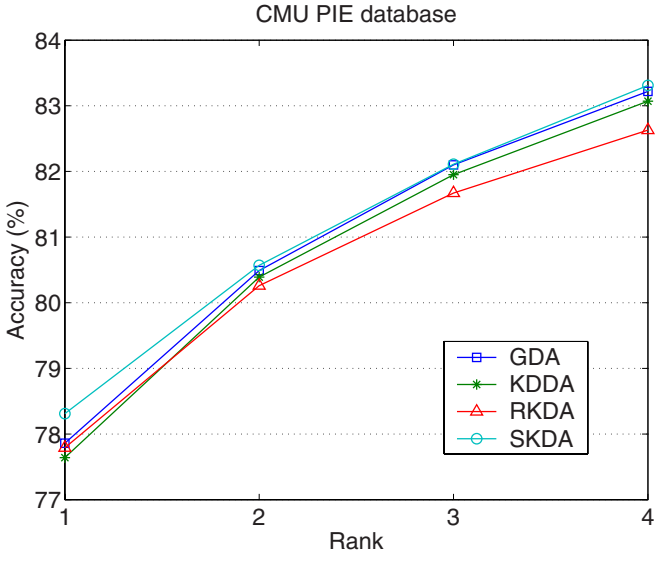

Fig. 4. Performance on CMU PIE face database

5 Conclusions

Based on Shannon wavelet kernel, this paper proposes and develops a novel subspace KDA algorithm for nonlinear feature extraction for face identification. Two human face databases, namely FERET database and CMU PIE database, are selected for evaluation. The results are encouraging on FERET and CMU PIE face databases. Experimental results show that the proposed SKDA algorithm gives better performance than existing state-of-the-art RBF kernel based FDA algorithms.

Acknowledgement

This project was supported by the Science Faculty Research grant of Hong Kong Baptist University RGC Earmarked Research Grant HKBU-211306 and NSF of China (60373082) and NSF of Guangdong province (06105776). The authors would like to thank for the US Army Research Laboratory for contribution of the FERET database and CMU for the CMU PIE database.

References

1. Belhumeur, P.N., Hespanha, J.P., Kriegman, D.J.: Eigenfaces vs. fisherfaces: Recognition using class specific linear projection. IEEE Transactions on Pattern Analysis and Machine Intelligence 19(7), 711–720 (1997)
2. Baudat, G., Anouar, F.: Generalized discriminant analysis using a kernel approach. Neural Computation 12(10), 2385–2404 (2000)
3. Martinez, A.M., Kak, A.C.: PCA versus LDA. IEEE Transactions on Pattern Analysis and Machine Intelligence 23(2), 228–233 (2001)
4. Lu, J., Plataniotis, K.N., Ventsanopoulos, A.N.: Face recognition using kernel discriminant analysis algorithms. IEEE Trans. on Neural Network 14(1), 117–126 (2003)
5. Chen, W.S., Yuen, P.C., Huang, J.: A New Regularized Linear Discriminant Analysis Methods to Solve Small Sample Size Problems. Int. J. Pattern Recognit. Artif. Intell. 19(7), 917–936 (2005)
6. Huang, J., Yuen, P.C., Chen, W.S., Lai, J.H.: Component-based subspacec linear discriminant analysis method for recognition of face images with one training sample. In: Optical Engineering, vol. 44(5) (2005)
7. Yang, J., Frangi, A.F., Yang, J.Y., Zhang, D., Jin, Z.: KPCA plus LDA: a complete kernel Fisher Discriminant framework for feature extraction and recognition. IEEE Trans. Pattern Anal. Mach. Intell. 27(2), 230–244 (2005)
8. Chen, W.S., Yuen, P.C., Huang, J., Dai, D.Q.: Kernel machine-based one-parameter regularized Fisher discriminant method for face recognition. IEEE Transactions on Systems, Man and Cybernetics, Part B 35(3), 659–669 (2005)
9. Chen, W.S., Yuen, P.C., Huang, J., Lai, J.H.: Face Classification based on Shannon Wavelet Kernel and Modified Fisher Criterion. In: Proceeding of the 7th IEEE international conference on automatic face and gesture recognition, April 10-12, 2006 pp. 467–474 (2006)
10. Huang, J., Yuen, P.C., Chen, W.S., Lai, J.H.: Choosing Parameters of Kernel Subspace-LDA for Recognition of Face Images under Pose and Illumination Variations. In: IEEE Transactions on Systems, Man and Cybernetics, Part B (2007) (Accepted to be published)
11. Xiong, H.L., Swamy, M.N.S., Ahmad, M.O.: Two-dimensional FLD for face recognition. Pattern Recognition 38(7), 1121–1124 (2005)
12. Mallat, S.: A theory for multiresolution signal decomposition: the wavelet representation. IEEE Pattern Anal. and Machine Intell 11(7), 674–693 (1989)
13. Meyer, Y.: Wavelets and operators. Cambridge Univ. Press, Cambridge (1993)
14. Daubechies, I.: Ten lectures on wavelets, CBMS-NSF conference series in applied mathematics, SIAM Ed (1992)