Optik 124 (2013) 5599–5603

Contents lists available at ScienceDirect

Optik

journal homepage: www.elsevier.de/ijleo

Face recognition based on PCA image reconstruction and LDA

Changjun Zhou*, Lan Wang, Qiang Zhang, Xiaopeng Wei

Key Laboratory of Advanced Design and Intelligent Computing, Dalian University, Ministry of Education, Dalian 116622, China

ARTICLE INFO

Article history:
Received 13 November 2012
Accepted 1 April 2013

Keywords:
PCA
Image reconstruction
Residual images
LDA
Face recognition

ABSTRACT

Face recognition has become a research hotspot in the field of pattern recognition and artificial intelligence. Principal Component Analysis (PCA) and Linear Discriminant Analysis (LDA) are two traditional methods in pattern recognition. In this paper, we propose a novel method based on PCA image reconstruction and LDA for face recognition. First, the inner-classes covariance matrix for feature extraction is used as generating matrix and then eigenvectors from each person is obtained, then we obtain the reconstructed images. Moreover, the residual images are computed by subtracting reconstructed images from original face images. Furthermore, the residual images are applied by LDA to obtain the coefficient matrices. Finally, the features are utilized to train and test SVMs for face recognition. The simulation experiments illustrate the effectivity of this method on the ORL face database.

1. Introduction

Face recognition is a technology of using computer to analyze the face images and extract the features for recognizing the identity of the target [1]. The research of face recognition has great theoretical value, involving subjects of pattern recognition, image processing, computer vision, machine learning, physiology, and so on, and it also has a high correlation with other biometrics recognition methods. In recent years, face recognition is one of the most active and challenging problems in the field of pattern recognition and artificial intelligence. Face recognition has a lot of advantages which are not involved in biometrics recognition methods, such as nonaggressively, friendly, conveniently, and so on. Therefore, face recognition has a prospective application foreground, such as the criminal identification, security system, file management, entrance guard system, and so on.

The key of face recognition lies in feature extraction and the design of the classifier, and PCA method is one of the most popular methods in feature extraction. The idea of using principal components to represent human faces was developed by Kirby and Sirovich [2] and used by Turk and Pentland [3] for face detection and recognition, namely “Eigenfaces” method. PCA is also known as Karhunen–Loeve transformation. PCA method has the characteristic of descending dimension and elimination of correlation. PCA obtains the maximum variance of data, and it’s optimal in terms of minimum reconstruction error. However, by its nature, it is not suitable for classification problems since it does not make use of any class information in computing the principal components [4,5].

Linear Discriminant Analysis (LDA) is a well-known method for dimension reduction in pattern recognition. It projects the original high-dimensional data onto a low-dimensional space, where all the classes are well separated by maximizing the Raleigh quotient. LDA creates a linear combination of independent features which yields the largest mean differences between the desired classes. The basic idea of LDA is to find a linear transformation such that feature clusters are most separable after the transformation which can be achieved through scatter matrix analysis. In other words, the goal of LDA is to maximize the between-class scatter matrix measure while minimizing the within-class scatter matrix measure. However, applied in the practical environment, in general, the number of training samples is less than that of number of dimension of the feature space (Small Sample Size Problem), hence, the within-class scatter matrix is singular and the Linear Discriminant Analysis (LDA) method cannot be applied directly [6]. In order to solve the linear discriminant in the Small Sample Size Problem, many LDA methods have been proposed in recent years, The most popular method is so-called Fisherfaces which is proposed by Belhumeur, Hespanha, and Kriegman [7] in 1997. This method combines PCA and LDA for dimensionality reduction so as to make the within-class scatter matrix nonsingular, but some useful discriminatory information may be lost [5].

According to the above discussion, in this paper, we propose a novel method based on PCA image reconstruction and LDA for face recognition. First, the inner-classes covariance matrix for feature extraction is used as generating matrix and then eigenvectors from each person is obtained, then we obtain the reconstructed images. Moreover, the residual images are computed by subtracting reconstructed images from original face images. Furthermore, the residual images are applied by LDA to obtain the coefficient matrices. Finally, we use two classifiers for face recognition respectively,

* Corresponding author.
E-mail address: zhou-chang231@163.com (C. Zhou).

0030-4026/$ – see front matter © 2013 Elsevier GmbH. All rights reserved.
http://dx.doi.org/10.1016/j.ijleo.2013.04.108

including the minimum distance classifier and Support Vector Machine (SVM). The proposed methods not only overcome the Small Sample Size Problem caused by LDA, but also solve the incomplete separable sample problem caused by PCA. The simulation experiments illustrate the effectivity of this method on the ORL face database.

2. Relevant works

2.1. Principal Component Analysis (PCA)

Consider a training set with the following parameters: the training set is denoted by $A = (A_1, A_2, \cdots, A_N)$, the training set has N images, $j = 1, 2, \cdots, N$, belonging to c classes, $i = 1, 2, \cdots, c$, and the number of pixels in the image is n. The between-class scatter matrix, within-class scatter matrix, and total scatter matrix are defined as [8]:

$$S_b = \sum_{i=1}^{c} P(\omega_i)(\mu_i - \mu_0)(\mu_i - \mu_0)^T \tag{1}$$

where S_b is the between-class scatter matrix, $P(\omega_i)$ is the priori probabilities of ω_i, in general, $P(\omega_i) = (1/c)$, μ_i is the mean vector of class ω_i, $\mu_0 = (1/N)\sum_{j=1}^{N} A_j$ is the mean vector of all the samples.

$$S_w = \sum_{i=1}^{c} P(\omega_i) S_i \tag{2}$$

where S_w is the within-class scatter matrix, $S_i = E\{(A - \mu_i)(A - \mu_i)^T | A \in \omega_i\}$ is the covariance matrix of ω_i.

$$S_t = S_b + S_w = \frac{1}{N}\sum_{j=1}^{N} (A_j - \mu_0)(A_j - \mu_0)^T \tag{3}$$

The between-class scatter matrix represents the scatter of class means μ_i around the overall mean μ_0, and the within-class scatter matrix is the scatter of the samples around their respective class means μ_i.

The PCA-based face recognition methods mainly use total scatter matrix or within-class scatter matrix to extract facial features. If the total scatter matrix is used as generating matrix, the optimal projection matrix is equal to computing the maximal eigenvalues and the corresponding eigenvectors of S_t, and the optimal projection matrix $(X_1, X_2, \cdots, X_d)$ is the eigenvectors associated with the d largest generalized eigenvalues. Therefore, $(X_1, X_2, \cdots, X_d)$ describes the contribution of each eigenface in representing the input face image, then we can extract facial features through it.

2.2. Linear Discriminant Analysis (LDA)

The traditional Linear Discriminant Analysis (LDA) is proposed by R.A. Fisher [9], so it is also known as Fisher's Discriminant Analysis. LDA searches for a linear transformation such that the feature clusters are most separable after the transformation which can be achieved through scatter matrix analysis. Given the training set $V = (v_1, v_2, \cdots, v_m)$, $V \subset R^{n\times m}$ each column of contains pixel values of one face image, and each image belongs to classes, the between-class scatter matrix S_w and within-class scatter matrix S_b are defined as [10–12]:

$$S_w = \sum_{i=1}^{c}\sum_{j=1}^{N_i} (V_j - \mu_i)(V_j - \mu_i)^T \tag{4}$$

$$S_b = \sum_{i=1}^{c} N_i(\mu_i - \mu_0)(\mu_i - \mu_0)^T \tag{5}$$

where N_i is the number of samples in class V_i.

LDA method aims at searching for a group of basis vectors, which makes different class samples, have the largest between-class scatter and the smallest within-class scatter. If S_w is nonsingular, the optimal projection matrix is denoted by:

$$W_{opt} = \underset{W}{argmax} \frac{|W^T S_b W|}{|W^T S_w W|} = [W_1 W_2 \cdots W_{c-1}] \tag{6}$$

where $W_{opt} = [W_1 W_2 \cdots W_{c-1}]$ can be obtained via solving the generalized eigenvalue problem:

$$S_b W_i = \lambda_i S_w W_i, \quad i = 1, 2, \cdots, c-1 \tag{7}$$

Because the generalized characteristic equation is only truly optimized in $c-1$ dimensions as the S_b matrix lies, at most, in $c-1$-dimensional space, $W_{opt} = [W_1 W_2 \cdots W_{c-1}]$ is the eigenvectors associated with the $c-1$ largest generalized eigenvalues.

When S_w is nonsingular, the generalized characteristic Eq. (7) can be translated into:

$$S_w^{-1} S_b W_i = \lambda_i W_i, \quad i = 1, 2, \cdots, c-1 \tag{8}$$

In order to solve the Small Sample Size Problem which appears when using LDA in feature extraction, PCA+LDA was proposed by Belhumeur, Hespanha, and Kriegman [7]. Where, the optimal projection matrix W_{opt} in this situation can be formulated into solving the following optimization problems:

$$W_{opt}^T = W_{lda}^T W_{pca}^T \tag{9}$$

$$W_{pca} = \underset{W}{argmax} |W^T S_t W| \tag{10}$$

$$W_{lda} = \underset{W}{argmax} \frac{|W^T W_{pca}^T S_b W_{pca} W|}{|W^T W_{pca}^T S_w W_{pca} W|} \tag{11}$$

where S_t is the total scatter matrix, defined by: $S_t = \sum_{j=1}^{m} (V_j - \mu)(V_j - \mu)^T$.

2.3. Support vector machine (SVM)

Support Vector Machine (SVM) is a novel learning machine with many merits such as fast solving and strongly generalizing ability, so it is widely used in the field of pattern recognition. SVM is developed from the theory of limited samples Statistical Learning Theory (SLT) and VC-dimension (Vapnik–Chervonenkis Dimension), which is originally designed for binary classification. Based on statistical learning theory, Support Vector Machine (SVM) can solve the small-sample problems well by using the structural risk minimization principle [13].

The basic idea of SVM is to adoption of non-linear transform, so that after changing the characteristics of the sample space become linearly separable. And for linear non-separable data, SVM uses a device called kernel mapping to map the data in input space to a high-dimensional feature space in which the problem becomes linear. The most commonly used kernel functions are polynomial, RBF and sigmoid functions, as follows [14,15]:

(1) Polynomial function

$$K(x, x_i) = [(x, x_i) + 1]^d \tag{12}$$

The corresponding SVM is a classifier of polynomial of degree d.

$$f(x, \alpha) = sign\left(\sum y_i \alpha_i [x_i \cdot x + 1]^d - b\right) \tag{13}$$

(2) Radial basis function

$$K_r(|x-x_i|) = \exp\left\{-\frac{(x-x_i)^2}{\sigma^2}\right\} \tag{14}$$

The SVM is a Gaussian RBF classifier, the optimal decision function can be written as:

$$f(x) = sign\left(\sum_{i=1}^{n}\alpha_i K_r(|x-x_i|) - b\right) \tag{15}$$

where $K_r(|x-x_i|)$ depends on the distance between two vectors. In order to setting up the discriminant criterion of formula (15), we have to estimate some parameters, including the value of σ, the number N of the central point x_i, the central vector x_i, the value of α_i.

(3) Sigmoid function

$$K(x, x_i) = \tanh(\gamma x_i^T x_j + b) \tag{16}$$

We choose RBF kernel function for SVM classifier in our face recognition experiments.

3. Face recognition based on PCA image reconstruction and LDA

3.1. Image preprocessing

Image preprocessing is a core and key step of the face recognition system. Because of various disturbance (such as illumination changes, face rotation, and background interference) in the process of acquiring face images, the input images should be preprocessed first before feature extraction. Therefore, the input images are given to the image preprocessing to remove the illuminations, shades, lighting effects. The common ways for image preprocessing include histogram equalization, geometry normalization, image smoothing, mean centering, and so on [16].

To train a classifier, all samples are needed to perform histogram equalization. That is, it normalizes the histogram of face samples and makes faces more compact in the image space. Histogram equalization is probably the simplest and most common technique to lessen the effects of lighting variations. Discussions about histogram equalization can be found in [17]. Let $r_i, r_2, \cdots, r_{L-1}$ denote the gray value of image $f(x, y)$, and $n(r_i)$ is the probability of grayscale pixel r_i, so image histogram is defined as:

$$p(r_i) = \frac{n(r_i)}{N}, \quad i = 0, 1, \cdots, L-1 \tag{17}$$

where N is the total pixels of an image, L is the grayscale of image pixel, and

$$\sum_{i}^{L-1} p(r_i) = 1 \tag{18}$$

So, we can obtain the cumulative probability of image pixel distribution $P_f(r_i)$ and the pixel gray value of the output image s_i, as follows:

$$P_f(r_i) = \sum_{j=0}^{i} p(r_j) \tag{19}$$

$$s_i = T(r_i) = \sum_{j=0}^{i} p(r_j) = \sum_{j=0}^{i} \frac{n_j}{N}, \quad i = 0, 1, \cdots, L-1;\ 0 \le r_i \le 1 \tag{20}$$

Fig. 1. Some of the eigenfaces of ORL database.

3.2. PCA image reconstruction

The traditional PCA-based face recognition methods mainly use the total scatter matrix or within-class scatter matrix to extract facial features. Through this method, we can only obtain the common characteristics and ignore everyone's different face characteristics, but everyone's different face characteristics are the most useful information in face recognition. Hence, we propose a novel method based on PCA image reconstruction for face recognition. We uses the inner-classes covariance matrix $S_i = E\{(A-\mu_i)(A-\mu_i)^T | A \in \omega_i\}$ as generating matrix and then we obtain all the individual characteristics subspace W_i, $i = 1, 2, \cdots, m$. Namely, firstly, calculate the eigenvalues and the corresponding eigenvectors of the generate matrix S_i, secondly, put these eigenvalues in order of largest to smallest, and similarly put the corresponding eigenvectors in order of largest to smallest, thirdly, choose some of them to structure the characteristic subspace, and then obtain the characteristics subspace of all the 40 classes of face images. Some of the eigenfaces of ORL database is shown in Fig. 1.

Moreover, features vector H_{ij} are extracted by mapping the training image X_j to basis image W_i by Eq. (21), and then they are utilized to reconstruct the images by Eq. (22), finally, we can obtain the reconstructed images Y_{ij}. Some of reconstructed images of ORL database is shown in Fig. 2.

$$H_{ij} = (X_j - \mu_i) \times W_i \tag{21}$$

$$Y_{ij} = W_i \times H_{ij} + \mu_i \tag{22}$$

Finally, the residual image $\overline{X}$ is obtained by subtracting the reconstruction image Y_{ij} from the original image X_j, namely Eq. (23). The residual images still contain rich information for the individual identities, so we extract face features from these residual faces instead of the original faces.

$$\overline{X} = X_j - Y_{ij} \tag{23}$$

3.3. PCA image reconstruction and LDA for face recognition

After obtaining the residual images by performing PCA on the original images, the residual images are applied by LDA to obtain the coefficient matrices. Finally, the features are utilized to train and test SVMs for face recognition. Fig. 3 shows the flow char of PCA image reconstruction and LDA for face recognition.

The major steps in the face recognition are as follows:

Step 1. Preprocess all images in the database, mainly include histogram equalization, geometry normalization, image smoothing, and remove the illuminations, shades, lighting effects possibly;

Step 2. Randomly partitioned into a training set from face database and the rest is a testing set;

Step 3. Use the inner-classes covariance matrix $S_i = E\{(A-\mu_i)(A-\mu_i)^T | A \in \omega_i\}$ as generating matrix, and then obtain all the individual characteristics subspace W_i, $i = 1, 2, \cdots, m$. Namely, firstly, calculate the eigenvalues and the corresponding eigenvectors of the generate matrix S_i, secondly, put these eigenvalues in order of largest to smallest, and similarly put the corresponding eigenvectors in order of largest to smallest, thirdly, choose some of them

Fig. 2. Some of reconstructed images of ORL database.

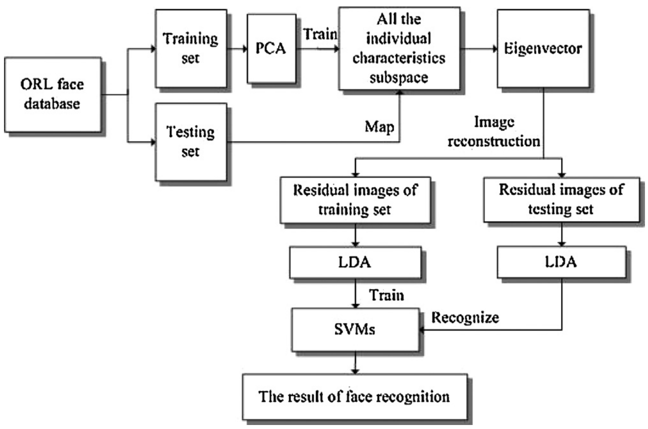

Fig. 3. The flow char of LDA and PCA image reconstruction for face recognition.

to structure the characteristic subspace, and then obtain the characteristics subspace of all the 40 classes of face images W_i, $i = 1, 2, \cdots, 40$;

Step 4. Obtain the encoding coefficients H_{ij} based on Eq. (21), and then reconstruct image Y_{ij} based on Eq. (22);

Step 5. Subtract the reconstruction image Y_{ij} from the original image X_j to get the residual image $\overline{X}$, $\overline{X} = X_j - Y_{ij}$;

Step 6. Perform LDA in residual images, compute LDA basic images W_{lda} and coefficient matrixs by Eqs. (9)–(11);

Step 7. Map the testing set to the characteristics subspace, similarly repeat Steps 4–6 for extract the testing set features;

Step 8. Utilize the train set's features to training SVMs, and utilize the test set's features and SVMs to recognize a face image.

4. Experimental results and discussion

4.1. Experiments on ORL database

The face recognition based on PCA image reconstruction and LDA is evaluated on the ORL database [18]. The ORL face database was composed of 400 images of size 112 × 92. There are 40 individuals, 10 face images for each person and that took at different time, varying lighting slightly, with unlike beam and expression, etc. All the images were taken against a dark homogeneous background.

Table 1
Comparison of different methods on ORL face database.

Recognition method	The average recognition rate (%)
CLDA [5]	97.20
PCA [19]	90.50
LDA [19]	93.82
ICA [19]	89.59
PCA + LDA [20]	94.06
PCA image reconstruction + LDA	97.48
PCA image reconstruction + LDA + SVM	97.74

Each image was linearly stretched to the full range of pixel values of [0, 255]. Fig. 4 shows some samples of one person in the ORL database.

For saving storage of computer and quickening elements' operation, on each benchmark, we reduced the face images size from 112 × 92 to 24 × 24 for efficiency. To evaluate the effectiveness of the algorithms better, we randomly choose 5 images among the 10 images of each person to compose the training set, and choose the rest 5 ones to compose the testing set. In this paper, we use two classifiers for face recognition respectively, including the minimum distance classifier and SVM. Furthermore, the experiments are repeated for fifty times, and the average recognition rate of fifty groups of data is reported. The results compare our methods with other methods on the ORL database is shown in Table 1.

In the experiments, there are 10 face images as extracting features, and then the rank of the inner-classes covariance matrix is 9 at the most. In other words, we can obtain 9 nonzero eigenvalues at the most, and the largest number of eigenfaces is 9. Where, the feature dimension d is chosen as 3, 4, 5, 6, 7, 8, 9. Table 2 shows the comparison of vary feature dimension results.

4.2. Discussion

The results of the experiments show that our proposed methods are better than other traditional face recognition methods. When the minimum distance classifier is used, the average recognition rate is 97.48%, and while using SVM, the average recognition rate is 97.74%. The recognition accuracy of our methods is much higher than that of traditional PCA and that of LDA and a little higher than that of CLDA method. Though the CLDA method of [5] uses the dual space of S_w to get the discriminant vectors without any loss

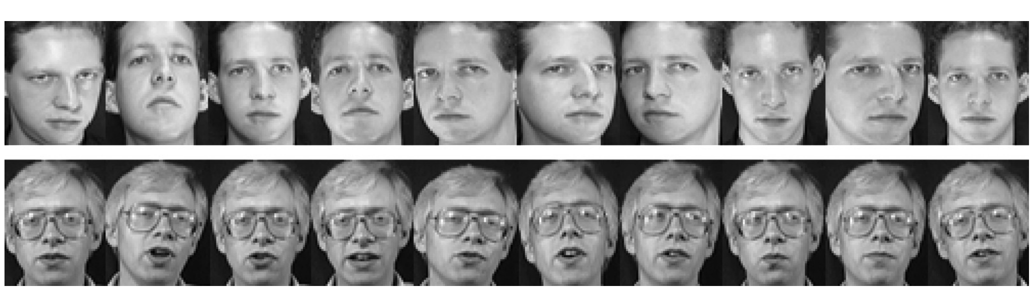

Fig. 4. Some samples of one person in ORL.

Table 2
Comparison of vary feature dimension based on ORL database.

Recognition method	Feature dimension						
	3	4	5	6	7	8	9
PCA image reconstruction + LDA	93.46	94.85	96.75	97.16	97.29	97.48	97.44
PCA image reconstruction + LDA + SVM	93.57	95.42	97.19	97.59	97.63	97.74	97.74

of effective discriminatory information, our methods effectively combine the advantages of PCA and LDA and can handle SSS problem better so as to improve face recognition accuracy. Besides, Table 2 indicates that the average recognition rate increases with the feature dimension, when the dimension to a certain degree, the average recognition rate increases slowly. Hence, the experiments on ORL show the efficiency of our proposed methods over the original face recognition methods.

5. Conclusions

In this paper, we propose a novel method based on PCA image reconstruction and LDA for face recognition. Our proposed methods effectively combine the advantages of PCA, LDA, and SVM. In this method, first, the inner-classes covariance matrix for feature extraction is used as generating matrix and then eigenvectors from each person is obtained, then we obtain the reconstructed images. Moreover, the residual images are computed by subtracting reconstructed images from original face images. Furthermore, the residual images are applied by LDA to obtain the coefficient matrices. Finally, we use two classifiers for face recognition respectively, including the minimum distance classifier and SVM. The simulation experiments illustrate the effectivity of this method on the ORL face database.

Acknowledgements

This work is supported by the National Natural Science Foundation of China (Nos. 31170797, 30870573, and 61103057), the Program for Changjiang Scholars and Innovative Research Team in University (No. IRT1109), the Program for Liaoning Excellent Talents in University (No. LR201003), the Key Project of Chinese Ministry of Education (No. 211036), and the Program for Liaoning Science and Technology Research in University (No. LS2010179).

References

[1] M.H. Yang, D.J. Kriegman, N. Ahuja, Detecting faces in images: a survey, IEEE Trans. Pattern Anal. Mach. Intell. 24 (1) (2002) 34–58.
[2] M. Kirby, L. Sirovich, Application of the KL procedure for the characterization of human faces, IEEE Trans. Pattern Anal. Mach. Intell. 12 (1) (1990) 103–108.
[3] M.A. Turk, A.P. Pentland, Eigenfaces for recognition, J. Cogn. Neurosci. 3 (1) (1991) 71–86.
[4] W. Zhao, R. Chellappa, A. Krishnaswamy, Discriminant analysis of principal components for face recognition, in: Third IEEE International Conference on Automatic Face and Gesture Recognition, 1998, pp. 336–341.
[5] G.F. Lu, J. Zou, Y. Wang, Incremental complete LDA for face recognition, Pattern Recogn. 45 (7) (2012) 2510–2521.
[6] A. Bansal, K. Mehta, S. Arora, Face recognition using PCA and LDA algorithm, in: Second International Conference on Advanced Computing Communication Technologies, 2012, pp. 251–254.
[7] P.N. Belhumeur, J.P. Hespanha, D.J. Kriegman, Eigenfaces vs. Fisher faces: recognition using class specific linear projection, IEEE Trans. Pattern Anal. Mach. Intell. 19 (7) (1997) 711–720.
[8] A. Eftekhari, M. Forouzanfar, H.A. Moghaddam, J. Alirezaie, Block-wise 2D kernel PCA/LDA for face recognition, Inform. Process. Lett. 110 (17) (2010) 761–766.
[9] R.A. Fisher, The use of multiple measurements in taxonomic problems, Ann. Eugen. 7 (1936) 178–188.
[10] A.H. Sahoolizadeh, B.Z. Heidari, C.H. Dehghani, A new face recognition method using PCA, LDA and neural network, Proc. World Acad. Sci. Eng. Technol. 31 (2008) 7–12.
[11] J.K. Li, B.J. Zhao, H. Zhang, Face recognition based on PCA and LDA combination feature extraction, in: 1st International Conference on Information Science and Engineering, 2009, pp. 1240–1243.
[12] S.N. Borade, R.P. Adgaonkar, Comparative analysis of PCA and LDA, in: International Conference on Business, Engineering and Industrial Applications, 2011, pp. 203–206.
[13] W.H. Chi, J.L. Chi, A comparison of methods for multiclass support vector machines, IEEE Trans. Neural Netw. 13 (2) (2009) 415–425.
[14] S. Viaene, B. Baesens, T. Van Gestel, Knowledge discovery in a direct marketing case using least squares support vector machines, in: International Journal of Intelligent Systems, 1999, pp. 112–120.
[15] N. Cristianini, J.S. Taylor, An Introduction to Support Vector Machines and Other Kernel-based Learning Methods, Cambridge University Press, New York, USA, 2004, pp. 82–108.
[16] R. Gross, V. Brajovic, An image preprocessing algorithm for illumination invariant face recognition, in: 4th International Conference on Audio- and Video Based Biometric Person Authentication, 2003, pp. 10–18.
[17] P.J. Phillips, Y. Vardi, Efficient illumination normalization of facial images, Pattern Recogn. Lett. 17 (8) (1996) 921–927.
[18] ATT Laboratories Cambridge, The ORL Database of Faces, 2005 http://www.cam-orl.co.uk/facedatabase.html
[19] G.Y. An, Q.Q. Ruan, Novel mathematical model for enhanced Fisher's linear discriminant and its application to face recognition, in: 18th International Conference on Pattern Recognition, 2, 2006, pp. 524–527.
[20] W. Zhou, X.R. Pu, Z.M. Zheng, Parts-based holistic face recognition with RBF neural networks, in: Third International Symposium on Neural Networks, 2006, pp. 110–115.