Optik 125 (2014) 5916–5919

Contents lists available at ScienceDirect

Optik

journal homepage: www.elsevier.de/ijleo

Face recognition based on PCA and logistic regression analysis

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 22 September 2013
Accepted 1 May 2014

Keywords:
PCA
Logistic regression
Linear regression
Face recognition

ABSTRACT

Face recognition is an important research hotspot. More and more new methods have been proposed in recent years. In this paper, we propose a novel face recognition method which is based on PCA and logistic regression. PCA is one of the most important methods in pattern recognition. Therefore, in our method, PCA is used to extract feature and reduce the dimensions of process data. Afterwards, we present a novel classification algorithm and use logistic regression as the classifier for face recognition. The experimental results on two different face databases are presented to illustrate the efficacy of our proposed method.

1. Introduction

Face recognition is an important research hotspot in the fields of pattern recognition and artificial intelligence, and it has attained great success in recent years. However, face recognition is also a challenging topic, because in real world face images are formed with the interaction of multiple factors on the different conditions, including background interference, illumination changes, face rotation, etc. [1]. Therefore, the various face recognition methods were presented to solve the problems, such as Principle Component Analysis (PCA) [2], Linear Discriminant Analysis (LDA) [3], Discrete Cosine Transform (DCT) [4], Independent Component Analysis (ICA) [5], Support Vector Machines (SVM) [6], and so on.

In this paper, we focus on the two problems which are how to extract feature and how to design novel classifiers for face recognition. Correspondingly, our research includes two technical components. Firstly, how to extract feature, we know that, PCA, LDA and ICA are common methods for feature extraction. The idea of using principal components to represent human faces was developed by Sirovich and Kirby [7] in 1987 and used by Turk and Pentland [8] in 1991 for face detection and recognition, namely "Eigenfaces" method. PCA can be also called as Karhunen–Loeve transformation. Besides, PCA has the characteristic of descending dimension and elimination of correlation. It can obtain the maximum variance of data, and it is optimal in terms of minimum reconstruction error [2].

Secondly, how to design novel classifiers for face recognition, there are lots of popular methods for classifiers, such as the minimum distance classifier, nearest neighbor classifier (NN), KNN, SVM, and so on. In recent years, SVM is widely used in the file of face recognition, because it is a novel learning machine with many merits such as fast solving and strongly generalizing ability. However, SVM is better applicable to solve binary classification problems, and it is difficult for solving the multiple classification problems [9]. Therefore, in 2010, Naseem proposed a novel method for face recognition which is a linear regression classification (LRC) algorithm [10]. Linear regression is a common statistical method. In his method, the regression coefficients are estimated by using the least square estimation method, and then the decision is made by the minimum distance between the original vector and the projected vector. Moreover, in 2012, a robust linear regression classification algorithm (RLRC) [11] which it estimates the regression parameters by using the robust Huber estimation was presented to address the problem of robust face recognition under illumination variation and random pixel corruption. In the same year, Huang and Yang [12] proposed an improved principal component regression classification (IPCRC) to overcome the problem of multicollinearity in the LRC. The IPCRC removes the mean of each image before performing principal component analysis and drops the first principal components. The projected coefficients are then executed by the linear regression classification algorithm. Then, in 2013, Huang and Yang [13] proposed linear discriminant regression classification (LDRC) algorithm to boost the effectiveness of the LRC for Face Recognition.

And, above all, linear regression classification is widely used, but it also has its limitations. Thus, we extend the linear regression classification, namely, in our method, we present a novel classification algorithm and use logistic regression as the classifier for face recognition. In essence, logistic regression is one kind of linear regression, and it only adds a layer of function mapping on the results of the mapping feature. Logistic regression can be not only

* Corresponding author.
E-mail address: zhangq26@126.com (Q. Zhang).

http://dx.doi.org/10.1016/j.ijleo.2014.07.080
0030-4026/

used to predict the probability but also used for classification. In the end, we can verify the effectiveness of the proposed methods based on two different face databases.

This paper is organized as follows. Section 2 summarizes the preliminary knowledge about related feature extraction and classification algorithms. Section 3 presents our novel approach based on Logistic Regression Classifier. In Section 4, experimental results based on two face databases are presented and discussed. Finally, we conclude in Section 5.

2. Related algorithms

2.1. Principal component analysis

Assume we have a training set $A = (A_1, A_2, \ldots, A_N)$ with N images, $j = 1, 2, \ldots, N$, belonging to c classes, $i = 1, 2, \ldots, c$, and the number of pixels in the image is n. Thus, the between-class scatter matrix, within-class scatter matrix, and total scatter matrix are defined as [2,14]:

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

where S_b is the between-class scatter matrix, $P(\omega_i)$ is the priori probabilities of ω_i, as usual, $P(\omega_i) = 1/c$, 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.

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

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.

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

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.

In our method, we use the total scatter matrix 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 $(X_1, X_2, \ldots, X_d)$ is the eigenvectors associated with the d largest generalized eigenvalues. Therefore, $(X_1, X_2, \ldots, X_d)$ describes the contribution of each eigenface in representing the input face image, then we can extract facial features through it. We will choose different d in the following experiments as needs.

2.2. Logistic regression analysis

In essence, logistic regression is one kind of linear regression, and it only add a layer of function mapping on the results of the mapping feature. Logistic regression has emerged as the conventional statistical technique of choice in the development of new models and also in the testing of existing instruments by Hosmer and Lemeshow in 1989. Many of its applications can be found in the fields of psychiatry and psychology. LR applies maximum likelihood estimation after transforming the dependent into a logit variable [15,16].

Logistic regression can not only be used to predict the probability, but also used for classification. Logistic regression can be used to classify individuals in the target categories through the logistic function. It is related to the probability of the chosen outcome event.

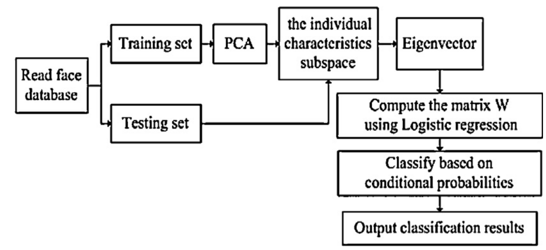

Fig. 1. The flow chart of our proposed method.

Let there be N number of classes with p_i training images from the ith class, $i = 1, 2, \ldots, N$. Through the projection of the image space onto the face space and normalizing each image vector, the matrix $W = [W_1, \ldots, W_i, \ldots, W_N]$ contains all feature vectors from classes. In order to apply regression analysis to estimate class specific model, we develop a class-specific model X_i as follows [10,13]:

$$X_i = [W_{i,1}, \ldots, W_{i,j}, \ldots, W_{i,p_i}] \in \Re^{L \times p_i} \quad (4)$$

Where, each vector $W_{i,j}$ is a column vector in size of $L \times 1$. Each vector $W_{i,j}$ spans a subspace of $\Re^L$, which is called the column space of X_i. Thus, in the training level, the class is represented by a vector space, which is also called the regressor or predictor for each class.

If y belongs to the ith class, it should be represented as a linear combination of the training images from the same class (lying in the same subspace), and can be defined as

$$y = X_i \beta_i, \quad i = 1, 2, \ldots, N \quad (5)$$

where β_i is the vector of regression parameters. The goal of the linear regression is to find to minimize the residual errors, and β_i can be estimated using least-squares estimation:

$$\widehat{\beta_i} = (X_i^T X)^{-1} X_i^T y \quad (6)$$

In the above, a linear combination of the training images from the same class can be defined as $y = X_i\beta_i$, $i = 1, 2, \ldots, N$; however, based on logistic regression, it can be translated into:

$$y = g(X_i \beta_i), \quad i = 1, 2, \ldots, N \quad (7)$$

$$g(z) = \frac{e^z}{1 + e^z} \quad (8)$$

$$y = \frac{e^{X_i \beta_i}}{1 + e^{X_i \beta_i}} \quad (9)$$

where, a linear combination of the features z can be expressed as

$$z = w_0 + w_i x_i + w_2 x_2 + \cdots + w_i x_i \quad (10)$$

Then, the conditional probabilities can be defined as

$$P(y = 1|x) = g(z) = g\left(\sum_{i=0}^{N} w_i x_i\right) \quad (11)$$

Thus, we use gradient descent method to obtain the weight of each feature w.

3. Proposed face recognition method

By combining the advantages of both the PCA and logistic regression, we propose a new method, and use PCA to extract feature and reduce the dimensions of process data. Afterwards, we use logistic regression as the classifier for face recognition. Fig. 1 depicts overall procedure of our proposed method.

First, we preprocessed the input images, mainly including histogram equalization, geometry normalization, in order to remove the illuminations, shades, and lighting effects possibly, and then

Fig. 2. Some samples images in Yale database.

Table 1
Comparison of different methods on Yale face database.

Recognition method	The recognition rate (%)
PCA [19]	90.50
LDA [19]	89.29
LocLDA [21]	78.10
DSNPE1 [22]	90.61
DSNPE2 [22]	89.88
Our proposed method	93.33

partitioned into a training set from face database and the rest is a testing set.

Then, we use the total scatter matrix as generating matrix to extract feature, and then obtain all the individual characteristics subspace W_i, $i = 1, 2, \ldots, m$. Namely, firstly, calculate the eigenvalues and the corresponding eigenvectors of the generate matrix S_t, secondly, put these eigenvalues in order of largest to smallest, and similarly put the corresponding eigenvectors in order of largest to smallest, the optimal projection matrix $(X_1, X_2, \ldots, X_d)$ is the eigenvectors associated with the d largest generalized eigenvalues, thirdly, choose some of them to structure the characteristic subspace, besides, we will choose different d in the following experiments as needs.

Finally, logistic regression will be used as the classifier for face recognition. In order to apply regression analysis to estimate class specific model, we develop a class-specific model $X_i = [W_{i,1}, \ldots, W_{i,j}, \ldots, W_{i,p_i}] \in \Re^{L \times p_i}$, and we use gradient descent method to obtain the weight of each feature W. Afterward, we compute the conditional probabilities for classification and output the recognition results.

4. Experimental results and discussion

To illustrate the efficacy of our proposed method, we compared the performances on two standard databases, i.e., Yale database [17] and the ORL database [18].

4.1. Yale database

The Yale face database contains images with major variations, including changes in illumination conditions, subjects wearing eyeglasses and different facial expressions. This database involves 165 frontal facial images, with 11 images of 15 individuals. Fig. 2 shows some sample images of two individuals of this database.

To evaluate the effectiveness of the algorithms better, each image is scaled down to the size of 100 × 100 pixels, and we use the first five images of every person as a training set, with the person's other six images as a testing set. Table 1 shows the result of the recognition rate of different methods on Yale face database.

The traditional face recognition methods have lower the recognition. Besides, LocLDA [21] considers an efficient LDA which can incorporate the local geometrical structure, and then presents a LocLDA method whose recognition rate is 78.10%. DSNPE [22] proposed a new learning algorithm called discriminant sparse neighborhood preserving embedding which adds the discriminant information into sparse neighborhood preserving embedding, but its recognition rate is only 90.61%. Thus, our method which uses logistic regression performs better effectiveness compared to the above methods.

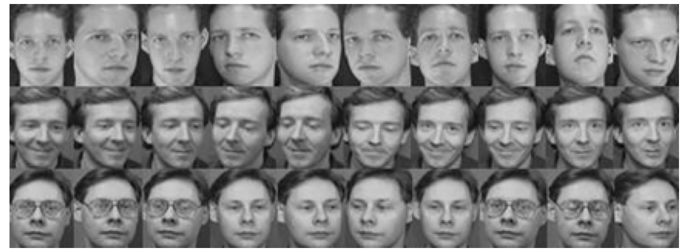

Fig. 3. Some samples in ORL.

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

Recognition method	Feature dimension				
	10	20	30	40	50
PCA+LDA [23]	0.780	0.835	0.895	0.930	–
Our proposed method	0.920	0.935	0.940	0.945	0.960

4.2. ORL database

The ORL face database was composed of 400 images (40 individuals and 10 face images for each person) of size 112 × 92. The images are taken at different time, varying lighting slightly, and facial details (open or closed eyes, smiling or non-smiling). All the images were taken against a dark homogeneous background. Fig. 3 shows some samples in the ORL database.

In ORL face database, each set of 10 images of a person was similarly partitioned into a training set of the first five images and a testing set of the other five images. In the experiments, we choose different feature dimension d as needs. Table 2 shows the comparison of vary feature dimension results, and it indicates that the recognition rate increases with the feature dimension. Besides, the results which compare our methods with other traditional methods on the ORL database are shown in Table 3.

The experimental results show that our proposed method is better than other traditional face recognition methods. While the PCA is used, the average recognition rate is 90.50%, and while using LDA, the average recognition rate is 93.82%. The recognition accuracy of our method is much higher than that of LR [10] whose recognition rate is 93.50%. Besides, the recognition accuracy of our method is a little higher than that of Gumus E [25] which uses wavelet transform.

Hence, this paper presents that our method has better recognition performance than PCA only based methods as well as other representative face recognition methods, such as LDA and linear regression.

Table 3
Comparison of different methods on ORL face database.

Recognition method	The recognition rate (%)
LR [10]	93.50
PCA [20]	90.50
LDA [20]	93.82
NS [24]	93.50
Gumus E [25]	95.30
Our proposed method	96.00

5. Conclusions

In this paper, we propose a novel face recognition method which is based on PCA and logistic regression. PCA is one of the most important methods in pattern recognition. Therefore, in our method, PCA is used to extract feature and reduce the dimensions of process data. Afterwards, we present a novel classification algorithm and use logistic regression as the classifier for face recognition. Our proposed method effectively combines the advantages of PCA and logistic regression. The experimental results on two different face databases are presented to illustrate the efficacy of our proposed method.

Acknowledgments

This work was supported by the National Natural Science Foundation of China (Nos. 31170797, 61103057, 31370778, 61370005), the Program for Changjiang Scholars and Innovative Research Team in University (No. IRT1109), and the Key Project of Chinese Ministry of Education (No. 211036).

References

[1] C. Tian, G. Fan, X. Gao, Multiview face recognition: from TensorFace to V-TensorFace and K-TensorFace, IEEE Trans. Syst. Man Cybern. B: Cybern. 42 (2) (2012) 320–333.
[2] 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.
[3] G.F. Lu, J. Zou, Y. Wang, Incremental complete LDA for face recognition, Pattern Recognit. 45 (7) (2012) 2510–2521.
[4] D. Zhong, I. Defe, DCT histogram optimization for image database retrieval, Pattern Recognit. Lett. 26 (14) (2005) 2272–2281.
[5] M.S. Bartlett, T.J. Sejnowski, Independent components of face images: a representation for face recognition, in: Proceedings of the Fourth Annual Joint Symposium on Neural Computation, 2007.
[6] O. Deniz, M. Castrillon, M. Hernandez, Face recognition using independent component analysis and support vector machines, Pattern Recognit. Lett. 24 (2003) 2153–2157.
[7] M. Kirby, L. Sirovich, Application of the KL procedure for the characterization of human faces, IEEE Trans. Pattern Mach. Intell. 12 (1) (1990) 103–108.
[8] M.A. Turk, A.P. Pentland, Eigenfaces for recognition, J. Cogn. Neurosci. 3 (1) (1991) 71–86.
[9] W.H. Chi, J.L. Chi, A comparison of methods for multiclass support vector machines, IEEE Trans. Neural Netw. 13 (2) (2009) 415–425.
[10] I. Naseem, R. Togneri, M. Bennamoun, Linear regression for face recognition, IEEE Trans. Pattern Anal. Mach. Intell. 32 (11) (2010) 2106–2112.
[11] I. Naseem, R. Togneri, M. Bennamoun, Robust regression for face recognition, Pattern Recognit. 45 (1) (2012) 104–118.
[12] S.M. Huang, J.F. Yang, Improved principal component regression for face recognition under illumination variations, IEEE Trans. Signal Process. Lett. 19 (4) (2012) 179–182.
[13] S.M. Huang, J.F. Yang, Linear discriminant regression classification for face recognition, IEEE Trans. Signal Process. Lett. 20 (1) (2013) 91–94.
[14] A. Eftekhari, M. Forouzanfar, H.A. Moghaddam, J. Alirezaie, Block-wise 2D kernel PCA/LDA for face recognition, Inf. Process. Lett. 110 (17) (2010) 761–766.
[15] Y.Y. Liu, M. Yang, M. Ramsay, S. Xiao, J.W. Li, A comparison of logistic regression, classification and regression tree, and neural networks models in predicting violent re-offending, J. Quant. Criminol. 27 (4) (2011) 547–573.
[16] M. Hajmeer, I. Basheer, Comparison of logistic regression and neural network-based classifiers for bacterial growth, Food Microbiol. 20 (1) (2003) 43–55.
[17] Yale University, The Yale Database of Faces, http://cvc.yale.edu/projects/yalefaces/yalefaces.html
[18] ATT Laboratories Cambridge, The ORL Database of Faces, 2005 http://www.cam-orl.co.uk/facedatabase.html
[19] Z. Wang, X. Li, Face recognition based on improved PCA reconstruction, Intell. Control Autom. (2010) 6272–6276.
[20] 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, vol. 2, 2006, pp. 524–527.
[21] X. Shu, Y. Gao, H. Lu, Efficient linear discriminant analysis with locality preserving for face recognition, Pattern Recognit. 45 (5) (2012) 1892–1898.
[22] J. Gui, Z. Sun, W. Jia, Discriminant sparse neighborhood preserving embedding for face recognition, Pattern Recognit. 45 (8) (2012) 2884–2893.
[23] J.K. Li, B.J. Zhao, H. Zhang, Face recognition based on PCA and LDA combination feature extraction, in: First International Conference on Information Science and Engineering, 2009, pp. 1240–1243.
[24] J.X. Mi, D.S. Huang, B. Wang, The nearest-farthest subspace classification for face recognition, Neurocomputing 113 (2013) 241–250.
[25] E. Gumus, N. Kilic, A. Sertbas, Evaluation of face recognition techniques using PCA, wavelets and SVM, Expert Syst. Appl. 37 (9) (2010) 6404–6408.