Neurocomputing 128 (2014) 415–420

Contents lists available at ScienceDirect

Neurocomputing

journal homepage: www.elsevier.com/locate/neucom

Optimizing principal component analysis performance for face recognition using genetic algorithm

Waled Hussein Al-Arashi [a,b], Haidi Ibrahim [a], Shahrel Azmin Suandi [a,c,*]

[a] Intelligent Biometric Group, School of Electrical and Electronic Engineering, USM Engineering Campus, Universiti Sains Malaysia, Seri Ampangan, 14300 Nibong Tebal, Pulau Pinang, Malaysia
[b] Electronic Engineering Department, Engineering Faculty, University of Science and Technology, Sanaá, Yemen
[c] Department of Information Systems, College of Computer Sciences and Information Technology, King Faisal University, P.O. Box 400, Al-Ahssa 31982, Saudi Arabia

ARTICLE INFO

Article history:
Received 14 January 2013
Received in revised form
7 May 2013
Accepted 14 August 2013
Communicated by S. Fiori
Available online 8 November 2013

Keywords:
PCA
Face recognition
Genetic algorithm
Principal component analysis

ABSTRACT

Principal Component Analysis (PCA) turns out to be one of the most successful techniques in face recognition systems as a statistical method for dimensionality reduction. Even so, it is yet not optimal from the perspective of classification because the underlying distribution among different face classes in the image space is unpredicted and not known in advance. Besides, in practical applications, a question always raised on how much data should be included in the training. In this paper, a technique that associates genetic algorithm (GA) to PCA is proposed to maintain the property of PCA while enhancing the classification performance. It reconsiders the available training data and tries to find the best underlying distribution for classification. ORL, and Yale A databases have been used in the experiments to analyze and evaluate the performance of the proposed method compared to original PCA. The experiment results reveal that the proposed method outperforms PCA in terms of accuracy and classification time.

1. Introduction

Face recognition, as biometric technologies among the six biometrics [1], has several advantages. It is natural, nonintrusive, and easy to use. Due to these reasons along with its potential for increasing commercial applications, it has been extensively researched for decades. However, developing a computer algorithm to recognize a face remains a difficult task due to it is a three-dimensional object and subjects to varying in many factors, which affect its performance. Many approaches have been proposed to build a face-recognition computer system. One of the most successful techniques that has been used in face recognition is Principal Component Analysis (PCA) [2–10] because of the ease of implementation, its reasonable performance level [4,5] and effectiveness in large databases [11]. Hence, many researchers focus on the improvement of PCA. Some works address the problem of selecting the best eigenvectors [4–6,8,12], which improve the performance of PCA by eliminating eigenvectors containing noise and decrease the time cost by compressing images. Others carried out the distance measurement methods [12–14] and the threshold used in the measurement [8].

Although PCA projections are optimal in terms of correlations or reconstruction errors from lower dimensional subspace, it is not from the perspective of classification [15]. For this reason, Linear Discriminant Analysis (LDA) [16], which is also known as Fisher Discriminant Analysis (FDA) [17], tries to maximize the separation between_class and minimize the within_class measure. One way to do this is to maximize the ratio of standard deviation of between_class to within_class. It is shown with some databases that LDA gives better results in terms of classifications. However, it is shown in [18] that LDA does not always give better accuracy. Therefore, PCA may outperform LDA in some circumstances. Beside this, it is also shown in [19] that PCA is less sensitive to different training data set. This is because it describes data better than LDA. Thus, this evidence points out to the likelihood, which PCA-based is more preferable compared to LDA-based. Due to this reason, LDA is not included in the experiments in this work. In addition, face images are prone to many variations due to different facial expressions, illuminations, poses and occlusion. Regarding these variations, it is also observed that variations due to illumination and viewing direction of the same class are larger than the variations due to change in different classes [20]. Thus, their

* Corresponding author at: Intelligent Biometric Group, School of Electrical and Electronic Engineering, USM Engineering Campus, Universiti Sains Malaysia, Seri Ampangan, 14300 Nibong Tebal, Pulau Pinang, Malaysia. Tel.: +60 4 5995814; fax: +60 4 5941023.
E-mail addresses: whma10_eee116@student.usm.my (W. Hussein Al-Arashi), haidi_ibrahim@ieee.org (H. Ibrahim), shahrel@eng.usm.my, shahrelsuandi@gmail.com, shahrel@usm.my (S. Azmin Suandi).

0925-2312/$ - see front matter © 2013 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.neucom.2013.08.022

distribution in the image space is unexpected and the underlying distribution of different classes is not known in advance [18]. Therefore, in practice, the available training data may be unsuitable or maybe not the foremost choice for building an effective face-recognition system. To behold this issue more expressively, Fig. 1 demonstrates the influence of training data distribution in calculating PCA component (eigenvectors) as well as the classification accuracy.

Fig. 1(a) displays the Eigenvectors calculation of random data in 2D space. Removing a point from the training data will affect the calculation of eigenvectors. For example, removing randomly the point corresponding to $x=16$, $y=20$ from class 1 in the previous data will change the direction of the eigenvectors. This change in direction may give better description toward the training data as depicted in Fig. 1(b). From these figures, it can be concluded that by reconsidering the available training data, different directions of eigenvectors can be calculated using PCA. Then, from these different eigenvectors, the eigenvectors which give the optimum subspace for classification can be selected. This indicates that not all available training data are favorable for recognition. Therefore, excluding some images from the available training data may lead to better recognition system. While, at the same time, the recognition time will also be more efficient.

To cope with these issues, in this paper, we propose a method that associates genetic algorithm (GA) to PCA to search the most suitable training data from the available ones. PCA technique is chosen because its' advantages mentioned above. By using GA combined with PCA, the best underlying distribution for classification can be determined. At the same time, this will increase the performance of PCA especially when the training data is large because the proposed method tries to find the best distribution which maximizes the between_class measure as in LDA while maintains the properties of PCA. Furthermore, the classification time is reduced because not all training data are used. On that account, there is no need to compare the probe image to all available training data. Along with this, it answers the question which always arises in practice of how many data should be included in the training stage. Therefore, the difficulty of ascertaining whether or not the available training data is appropriate for the recognition system is solved. On top of this, the proposed method is more applicable and suitable for real world face recognition applications.

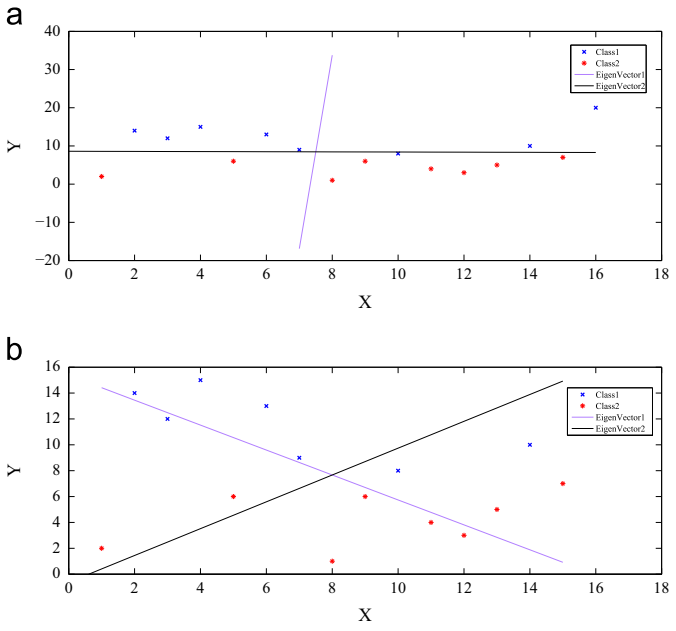

Fig. 1. The effect of training data distribution in calculating PCA component (eigenvectors). (a) Original data and (b) after removing one point ($x=16$, $y=20$).

The rest of this paper is organized as follows. In Section 2, PCA algorithm is reviewed. Section 3 describes the proposed method. Experimental results and the discussions are presented in Section 4. Finally, in Section 5, the paper is concluded.

2. Principal component analysis (PCA)

Let $\mathbf{X}$ represent a set of face images as follows:

$$\mathbf{X}=[x_1\ x_2\ \ldots\ x_M]$$

where x_i is a vector of face image with dimension N and M is the number of face images. The vector of face image is formed by concatenating the columns or rows of the image. The typical method of calculating the principal components is to find the eigenvectors and eigenvalues of the covariance matrix $\mathbf{C}$ [2] as given in the following:

$$\mathbf{C}=\sum_{i=1}^{M}(x_i-\overline{x})(x_i-\overline{x})^T \tag{1}$$

where x_i is a vector of face image and $\overline{x}$ is the average face image. The eigenvectors and the eigenvalues can be calculated as

$$\mathbf{C}v_i=\lambda_i v_i,\quad i=1,\ldots,N \tag{2}$$

where v_i is the ith eigenvector and λ_i is the corresponding eigenvalue which reflects the variance of the images. By selecting eigenvectors corresponding to the largest eigenvalues, we get a subspace which represents the image space in minimum Mean Square Error (MSE) manner. This subspace is called eigenvector or eigenface. By projecting the training data into the new space, we get the feature vector $y_i \varepsilon \mathbb{R}^m$ using the following equation:

$$y_i=\mathbf{V}^T(x_i-\overline{x}) \tag{3}$$

where $\mathbf{V}$ is the eigenvector matrix.

3. Searching optimal underlying distribution using genetic algorithm

3.1. Genetic algorithm

To find the most suitable training data from the available training data, genetic algorithm (GA) was used to search for the best data distribution which accomplishes the maximum classification accuracy. According to [21], GA is a stochastic algorithm that provides an efficient method to find globally the optimal solution in large space. One of the most significant features of GA is that it has a population of solutions in each cycle, which offers many advantages. GA begins with a random generation of a constant-sized population of n individuals called chromosomes. The fitness of each chromosome is evaluated. Then, a typical GA algorithm employs three distinctive operators, selection, crossover and mutation, which leads the populations towards convergence. Selection is the procedure of creating offspring from the current population by employing process similar to the natural selection in the biological systems. The aim of selection is to assert better performing, or fitter, individuals in the population in expectancy that their offsprings have a likelihood of promoting the information they include within the successive generations. The magnitude of the selection process has high impact on the convergence rate of GA. Along with this, selection approach should avert premature convergence by maintaining the diversity in the population. At the same time, it has to be balanced with other GA operations, i.e., crossover and mutation. Crossover is the procedure of picking two parents and exchange information between them,