Pattern Recognition 47 (2014) 1869–1879

Contents lists available at ScienceDirect

Pattern Recognition

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

An efficient approach for face recognition based on common eigenvalues

Vilas H. Gaidhane [a,*], Yogesh V. Hote [b], Vijander Singh [a]

[a] ICE Division, Netaji Subhas Institute of Technology (University of Delhi) New Delhi, New Delhi 110078, India
[b] Department of Electrical Engineering, Indian Institute of Technology (IIT), Roorkee-247667, India

ARTICLE INFO

Article history:
Received 25 April 2013
Received in revised form
2 November 2013
Accepted 27 November 2013
Available online 6 December 2013

Keywords:
Face recognition
Feature extraction
Polynomial coefficients
Companion matrix
Nullity

ABSTRACT

In this paper, a simple technique is proposed for face recognition among many human faces. It is based on the polynomial coefficients, covariance matrix and algorithm on common eigenvalues. The main advantage of the proposed approach is that the identification of similarity between human faces is carried out without computing actual eigenvalues and eigenvectors. A symmetric matrix is calculated using the polynomial coefficients-based companion matrices of two compared images. The nullity of a calculated symmetric matrix is used as similarity measure for face recognition. The value of nullity is very small for dissimilar images and distinctly large for similar face images. The feasibility of the propose approach is demonstrated on three face databases, i.e., the ORL database, the Yale database B and the FERET database. Experimental results have shown the effectiveness of the proposed approach for feature extraction and classification of the face images having large variation in pose and illumination.

1. Introduction

Over the last decade, the face recognition has become the most popular area of research in computer vision and demanding in the variety of applications [1]. Face identification among many faces is a biometric process by which captured face image is matched with the known face images. Identification of similar faces of people is a challenging problem due to the different conditions such as pose, light intensity, emotions and natural ageing. It is an active field of research and has become most popular due to its utility in different areas such as security applications, communication, human–computer interaction and biometrics, etc. The efficacy of such a system is an important issue and depends on the detection and evaluation of the similar faces properly.

Many researchers have worked on geometric-based as well as holistic-based face recognition approaches. In case of geometric-based approaches, the local features such as mouth, eyes, and nose are initially extracted from face images. These local statistics and locations are used for classification [2]. The most successful geometric-based methods are elastic bunch graph matching (EBGM) [3] and active appearance graph models (AAM) [4]. However, the performance of these techniques depends on the selection of facial landmarks, which are collected manually. This makes the system semi-automatic and labor consuming. Moreover, in these methods, the recognition performance deteriorates if the manual landmark selection process is replaced by an automatic landmark detection method [5].

In contrast to geometric-based methods, holistic-based approaches extract the holistic features of the whole face region. They consider all the pixels of a face image of size $n \times m$ and represented by a vector in an $n \times m$-dimensional space. However, these vector spaces are too large, which increases the computational complexity in face recognition system. A common way to overcome this problem is to the dimensionality reduction methods [6]. The most popular dimensionality reduction techniques are principal component analysis (PCA), which is a statistical approach used to express the faces in a subset of their eigenvector [7,8] and Linear discriminant analysis (LDA, also called as Fisher Discriminant Analysis) [9]. Moreover, Kernel PCA and independent component analysis (ICA) have been suggested for facial feature extraction [10]. Kim et al. [11] uses the Kernel PCA for face feature extraction and shows that the Kernel PCA method outperforms the conventional PCA-based methods. Bartlett et al. [12] proposed an ICA method for facial feature extraction and found better than PCA when cosines are used as a similarity measure for classification. However, Kernel PCA and ICA methods are computationally more expensive than the conventional PCA method.

In PCA-based face recognition, the 2D images must be transformed into the 1D image vector which leads to a higher image vector space. In order to overcome the dimensionality problem, a straightforward image projection method, called 2D principal component analysis (2DPCA) is developed for image feature extraction [13]. 2DPCA is based on 2D matrices of images rather

* Correspondence to: ICE Division, NSIT, Sector-3, Dwarka, New Delhi-110078, India. Tel.: +91 88 60302137; fax: +91 11 25099022.
E-mail address: vilasgd612@gmail.com (V.H. Gaidhane).

0031-3203/$ - see front matter © 2013 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.patcog.2013.11.027

than 1D image vectors. Therefore, 2DPCA method is superior to PCA, ICA and Kernel PCA in terms of computational complexity. However, this method requires more coefficients to represent an image than the PCA-based method. Sparse principal component analysis (SPCA) [14,35] is another development over the PCA and 2DPCA methods. LDA is another well-known feature extraction technique which is extensively used for feature extraction and dimension reduction in pattern recognition applications. The underlying deference between PCA and LDA is that PCA compute a vector that has the largest variance associated with the dataset. In PCA, the average of all images is subtracted from all normalized vectors to ensure that the eigenvectors corresponding to the largest eigenvalue represents the dimension in the eigenspace in which variance of vectors is highest in the sense of correlation. On the other hand, LDA used to compute a vector in the underlying space that best discriminate among classes. More precisely, LDA creates a linear combination of a number of independent features which yields the mean difference between the desired classes [6]. The main goal of LDA is to maximize the between-class measure and minimizing the within-class measure. However, it is very difficult to evaluate the scatter matrix due to its large size and less number of training images. Moreover, the singular within-class scatter matrix results into an impractical task. Therefore, a 2DLDA method [15] is presented, which calculates the scatter matrices and respective eigenvectors of an image more accurately without image matrix to 1D-vector conversion. Although, 2DLDA method is more efficient, it requires more number of coefficients for image feature representation than the PCA as well as LDA-based methods. To overcome these problems, an efficient technique commonly known as $(2D)^2$LDA is proposed by Noushath et al. [16] for accurate feature representation and recognition. They have presented an incredible work and carried out experiments on ORL and Yale database. This method needs less number of coefficients and reported better recognition accuracy than the PCA, 2DPCA and 2DLDA methods.

Further, based on the simplicity and performance of PCA and LDA methods, the combination of PCA and LDA (PCA+LDA) [17] and bidirectional PCA (BDPCA) plus LDA [18] are presented. Recent developments of the holistic-based approaches include the eigenfeature regularization and extraction (ERE) [19]. An efficient Bayesian approach using wavelet transform technique is also proposed by James and Annadurai [20], has very high face recognition ratio for ORL and Yale database B face databases. The discriminative common vector (DCV) method is a recently addressed discriminant method, which shows the promising results than the LDA algorithms [21]. The DCV is based on the Fisher's LDA and it is suitable for the small sample database. For large sample, an improved DCV using support vector machine is reported, which shows the effective face recognition results [22].

Although, the face recognition for the face images has achieved a nearly 100% success, its performance and computational complexity is still the issue of research. The holistic-based methods though provides the fast recognition, the discrimination among the faces is not enough in the case of large databases. Moreover, in feature-based approaches, the extraction of actual location of facial features under the varying conditions such as illumination and pose is again a challenging problem. Therefore, a simple and computationally efficient method is needed for the face recognition application.

After detailed study on various techniques in control engineering and image processing, it is observed that the control technique such as algorithm on common eigenvalues [26] can be useful in image processing applications. Therefore, in this paper, an attempt has been made to use this technique for face recognition. The proposed technique is based on the properties of polynomial coefficients, companion matrix and common eigenvalues between two images. The polynomial coefficients are calculated directly from the 2D face images. These coefficients represent the important features of images and further used to construct a reduced dimension companion matrix commonly called as a feature matrix. Finally, the similarity among the face images is obtained by calculating the nullity between the training and testing feature matrices.

The rest of the paper is structured as follows. A brief overview of existing face recognition methods and proposed approach is presented in Sections 2 and 3, respectively. In Section 4, details of face databases are explained, which is used for experimentations. Section 5 presents the result of experiments and analysis. Finally, conclusions are drawn in Section 6.

2. Overview of existing face recognition methods

The feature extractions have been extensively used in face recognition. In traditional holistic approaches, 2D images must be initially transformed into 1D image vectors, which lead to a higher-dimensional vector space. Therefore, it is very difficult to calculate the covariance matrix due to its large size and the small number of training images. This problem is overcome by 2DPCA method, which is based on 2D image matrices. A 2D image transformation has major advantages in image feature extraction and dimensionality reduction. The 2D transforms of image $\mathbf{R}(x,y)$results in a transformed image $\mathbf{R}'(p,q)$ which can be defined as [18,23,36]

$$\mathbf{R}'(p,q)=\sum_{x=0}^{m-1}\sum_{\mathbf{y}=0}^{n-1}\mathbf{R}(x,y)\,\mathbf{K}(x,\mathbf{y};p,q) \tag{1}$$

where $\mathbf{K}(x,\mathbf{y}\,;p,q)$ represents the transform kernel. The above transform is called as separable if its kernel can be rewritten as [24].

$$\mathbf{K}(x,y;p,q)=\mathbf{K}_{col}(x,\ p)\,\mathbf{K}_{\mathbf{r}ow}(y,\ q) \tag{2}$$

In separable transform, the transformed image matrix $\mathbf{R}'(p,q)$ can be obtained as

$$\mathbf{R}'(p,q)=\mathbf{K}_{col}\mathbf{R}(x,\ y)\mathbf{K}_{\mathbf{r}ow}^{\mathrm{T}} \tag{3}$$

In the case of holistic PCA, the image matrix first transformed into 1D vector ($\mathbf{r}_{vec}$) and the total scattering matrix for M training sample can be obtained as

$$\mathbf{S}_{\mathrm{T}}=\frac{1}{M}\sum_{i=1}^{M}(\bar{\mathbf{r}}_{veci}-\bar{\mathbf{r}}_{vec})(\bar{\mathbf{r}}_{veci}-\bar{\mathbf{r}}_{vec})^{\mathrm{T}} \tag{4}$$

where $\bar{\mathbf{r}}_{veci}$,$\bar{\mathbf{r}}_{vec}$ represent the mean of ith class and the whole training images, respectively. The 2DPCA can be obtained as

$$\mathbf{R}'=\mathbf{R}\mathbf{K}_{\mathbf{r}ow} \tag{5}$$

In bidirectional PCA (BDPCA) [18], if M training samples of size $m\times n$, then the ith image $\mathbf{R}_i$ can be represented as a n set of $1\times m$ rows vectors

$$\mathbf{R}_i=\begin{bmatrix}\mathbf{r}_{veci\ 1}\\ \mathbf{r}_{veci\ 2}\\ \vdots\\ \mathbf{r}_{veci\ n}\end{bmatrix} \tag{6}$$

Then the total scattering matrix for M training sample can be represented as

$$\mathbf{S}_{\mathrm{T}\mathbf{r}ow}=\frac{1}{mn}\sum_{i=1}^{m}\sum_{j=1}^{n}(\bar{\mathbf{r}}_{vecij}-\bar{\mathbf{r}}_{vecj})\,(\bar{\mathbf{r}}_{vecij}-\bar{\mathbf{r}}_{vecj})^{\mathrm{T}} \tag{7}$$

The row eigenvectors corresponding to the first largest eigenvalues of total scattering $\mathbf{S}_{\mathrm{T}\mathbf{r}ow}$ can be used to construct the row projection matrix $\mathbf{W}_{\mathbf{r}ow}$. Similarly, the column projection matrix