The Search for the Criminals using PCA Algorithm

Ahmed Mohamed Abushaala and Elham Ali Abobaker
Faculty of Information Technology
Misurata University - Libya
P. O. Box 2664 , Misurata - Libya
E-mails: am_rata@yahoo.co.uk and elhamabobker@gmail.com

ABSTRACT

In this work**,** we are offering a system that would help a lot in facilitating a search for persons who have criminal records. Our search starts by entering the data or the pictures. In this system, we have built a database for many persons. Each one has five different pictures, which are taken from a database ORL submitted by the University of Cambridge, to configure a database training. There are many algorithms which are used in this field; however, we have chosen the Principal Component Analysis (PCA) algorithm to extract the features. The system has been programmed using C# 2008 because of its salient features of high efficiency and speed of implementation. After a number of tests on persons fabricated with different conditions, we get excellent results while using a database of pictures (ORL). Our results are then compared with the results of the Artificial Neural Networks algorithm with the same database (ORL) and the results of our algorithm are far better.

KEYWORDS

Principal Component Analysis (PCA), Artificial Neural Network (ANN), Jacobi Method, face recognition, database ORL.

1 INTRODUCTION

Because of the increasing concerns regarding the security issues about the accuracy of the computer systems in identifying the faces, the number of security systems and applications is increasing and evolving significantly, where the face plays a key role in social communication. The ability to identify faces manage to recognize the thousands of faces throughout our lives and familiar faces at a glance even after years of separation [1], in spite of the great changes in visual stimulation because of the conditions of vision and expression, aging and deviations such as glasses or changes in hairstyle. Identify the faces activities area since late 1980, because it can contribute to not only in theory but also in practical applications, such as the criminal, security systems and image processing, movies, and interaction between human and computer, etc. [2].

In spite of advance mathematical models to identify the faces, it is still a challenge and difficulty because the face is multidimensional, and is presented to change with the passage of time. However, in order to solve this problem, various techniques, for example, Artificial Neural Networks (ANN), and Principal Component Analysis (PCA), have been suggested [1].

2 SEARCH CRIMINALS BY SPECIFIED DATA

The system receives a specified data which describe some features on the criminal, like a color of the hair, his old, or his tall. Then after the search process the system shows all the criminals who have these specified data**.**

3 SEARCH CRIMINALS BY PICTURE

3.1 Principal Component Analysis (PCA) Algorithm

The Principal Component Analysis (PCA) was presented for the first time in the year 1901 and was done several modifications up to the year 1963, and continues to this day as one of the most important ways of numerous variables (multivariate techniques). PCA algorithm lies in reducing the large dimensions (dimensionality reduction) and extracts the feature (feature extraction) [3].

The Principal Component Analysis algorithm is considered as one of the most important and successful techniques that have been used in the identification of images and image compression. The PCA is classified as one of the statistical methods [1].

The application of the transformation of the principal component requires to calculate the covariance matrix for the characteristics, then to calculate the basic components which can be calculated by Jacobi methods [3].

The goal of the transformation of the principal component is to transform the amount of high data and characteristics-scattered to data that contain in the fundamental characteristics and ranks [4].

3.2 Mathematical Description for Principal Component Analysis

The analysis of the principal component mathematically requires finding covariance matrix for the images available in accordance with the following equation:

$$Covij = \frac{1}{mn}\sum_{k=1}^{m}\sum_{l=1}^{n}(Xi(k,l) - Mi)(Xj(k,l) - Mj) \quad (1)$$

Where: m, n represents the number of elements in the picture; *i* and *j* represent the number of packets; *k*, *l* is the location of the point in the picture. *Mi*, *Mj* is the arithmetic average of packets *i*, *j* sequentially.

We note from the equation no. (1) that the matrix is homogenous, that means, the element value over the main diagonal is equal to the equivalent element value under the main diagonal. Thus, covariance matrix for n number of pictures is matrix n * n can be arranged as follows:

$$C = \begin{bmatrix} C11 & \cdots & C1n \\ \vdots & \ddots & \vdots \\ Cn1 & \cdots & Cnn \end{bmatrix} \quad (2)$$

Where, Cij = Cji for each j ≠ i.

We can derive the Eigenvector from the covariance matrix using Jacobi method, that will be described in the following paragraph, to get the matrix D, where D is diagonal matrix and describes the principal component as following:

$$D = \begin{bmatrix} \lambda 11 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \lambda nn \end{bmatrix} \quad (3)$$

The diagonal element of D is called Eigenvalue for covariance matrix in which all the elements outside the diagonal of the matrix is zero, which indicates that these components are non-interrelated [5][6].

3.3 Theory of Jacobi

The following points show the steps of Jacobi theory:

1. Find the largest component of certain square matrix (let it C) so this element does not from the main diagonal (I ≠ K), and I, K are the row and the column in covariance matrix .

2. Find angle θ through the following:

   if $a_{II} \neq a_{KK}$

   $\theta = 1/2 \arctan(2\, a_{IK} / (a_{II} - a_{KK}))$

   and if $a_{II} = a_{KK}$ then

$$\theta = \begin{cases} -\frac{\pi}{4} & \text{when } a_{IK} < 0 \\ \frac{\pi}{4} & \text{when } a_{IK} > 0 \end{cases}$$

3. Do the rotation process on the matrix, and returning the product matrix C as following:

$$d_{ii} = a_{ii}\cos^2\theta + 2a_{ik}\sin\theta\cos\theta + a_{kk}\sin^2\theta$$

$$d_{kk} = a_{ii}\sin^2\theta + 2a_{ik}\sin\theta\cos\theta + a_{kk}\cos^2\theta$$

$$d_{ik} = d_{ki} = -(a_{ii} - a_{kk})\sin\theta\cos\theta + a_{ik}(\cos^2\theta + \sin^2\theta)$$

4. Repeat the steps 1 to 3 to the resulting array to get the values of the non-main diagonal elements is close to zero [7][8].

The following figure shows the flowchart of the Principal Component Analysis (PCA) algorithm.

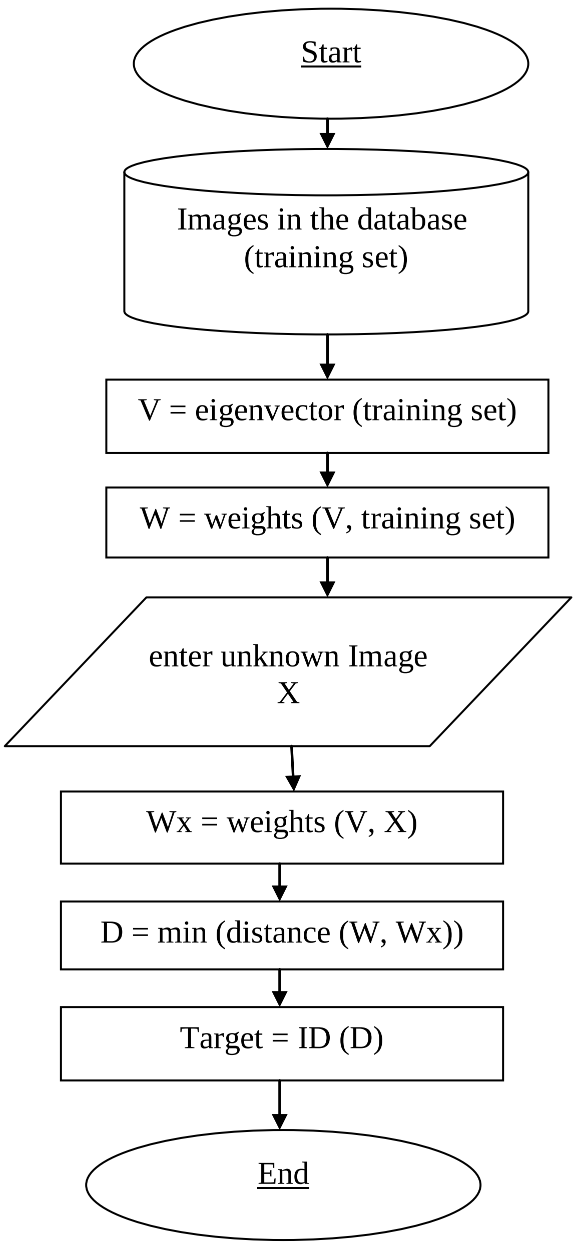

Figure (1) – The flowchart of steps of Principal Component Analysis Algorithm

4 THE RESULTS AND CONCLUSION

The forms of the images in the database ORL have been tested in PCA algorithm to draw features and identify the faces with least distance, and the number of samples, 195, appointed by the 39 classes (each class represents 5 photos of a particular person).

In order to calculate the rate of learning, we use the following equation:

$$\text{rate of learning} = \frac{\textit{total identified items}}{\textit{the total number of items}} * 100 \quad (4)$$

and the result was 100 % , which indicates that the number of samples that are known by the system is 195 samples.
The following figures show the results of the system.

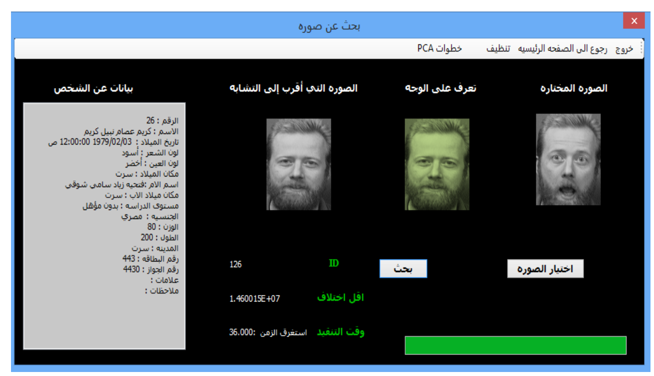

Figure (2) – The result of searching person-1using Principal Component Analysis algorithm

Figure (3) – The result of searching person-2 using Principal Component Analysis algorithm

Our results are compared with other system that used Artificial Neural Networks (ANN) algorithm with the same database (ORL), and its rate of learning was about 38%, which indicates that our system is far better.

5 REFERENCES

[1] العربية Schwarz Tiger مدونة , http://schwarztiger.wordpress.com/ التعرف - على - الوجوه – باستخدام - خوارزمية – principal/ , Blog Electronic Schwarz Tiger, 03/09/2013.

[2] Soliman, Karima , "Face recognition based on Artificial Neural Network (ANN)", a thesis submitted to Nottingham Trent University associated with Misurata University for the fulfillmen t of the part required for the degree of Master in Information Technology, 2009.

[3] Jacqueline Le Moigne, "Multi-Sensor Image Registration, Fusion and Dimension Reduction", 2003.

[4] P. Bajcsy and P. Groves, "Methodology For Hyperspectral Band Selection", Published in Photogrammetric Engineering and Remote Sensing journal, Vol. 70, No. 7, 2004.

[5] H. Eesa, "A PC_Based Multispectral Image Processing Software for Remote Sensing Applications", MSc. thesis, University of Mosul, Collage of Science, Computer Science Dept., 1998.

[6] T. Shumam,"An Integrated Software for Remotely Sensed Data Classification", MSc. thesis, University of Mosul, College of Science, Computer Sc. Dept., 2001.

[7] Froberg C.,"Introduction to Numerical Analysis", 2nd edition, Addison –Wesely, 1969.

[8] http://en.wikipedia.org/wiki/Jacobi_eigenvalue_algorithm, 25/10/2013.