PCA Based Face Recognition System

Md. Rakibul Alam Khan, Halima Akhter

ABSTRACT: "Face Recognition" is a common research work. Now a day's various techniques are available and in this paper Principal Component Analysis (PCA) have used. PCA use feature extraction and data representation technique which widely used in pattern recognition. In a research paper of Manal Abdullah, Majda Wazzan and Sahar Bo-saeed [1] they have shown the eigrnfaces for the PCA algorithm where eigen values have sorted and those who are less than a specified threshold have eliminated. In PCA a set of eigenface have been computed for a database of face image and then test image has been compared to the images in the database. A specific threshold value has been applied to the image result, which work like a filter for false image detection. A good experimental result has found from image data base. For testing database, we have used OLR and Local database. Threshold value has been applied on trial and error method.

Keywords: *PCA, Face recognition, test image, database,* eigenface

i. Introduction

There are many parts of pattern matching technique, face recognition is one of them. A face recognition system can identify the test image from stored database images. Face recognition is mainly pattern recognition. Face recognition is a very complex form of pattern recognition. The domain of face recognition lies on image processing. It uses PCA which is a statistical method which works on 1-D image converted from 2-D image and then measure the distance of feature vector.

Face Recognition can be classified into two different classes

1. Geometric.

2. Photometric.

This document concentrates on Feature Based Face recognition systems.

Face recognition has a wide range of applications. They are mostly used in biometrics, for reliable personal identification. Application of face recognition can be classified into two types [3] :

1. Trying to find a face within a large database of faces. In this approach, the system returns a possible face from the database.

2. Real time face recognition, where getting the test image of a person from live camera and then matching it with large scale images strode in database and give the result.

The data used plays a very important role in face recognition. Most of the techniques in this document use the ORL database at the Olivetti Research Laboratory in Cambridge, United Kingdom, and we also test it on local database.

ii. Features of the Face Measurement & methods of face recognition

Each human face has approximately 80 nodal points. Some of these measured by the software are:

1. Distance between the eyes
2. The width of the nose
3. The depth of the eye sockets
4. The shape of the cheekbones
5. The length of the jaw line

There are so many methods for face recognition [4]. So many analysis has been made to improve face recognition system. Some technique work well for 3D face recognition system other works well for 2D face recognition system. Some of these are given below:

1) PCA- (Principal Component Analysis).
2) ICA-(Independent Component Analysis).
3) EP – (eigen space-based approach).

4) EBGM- (Elastic Bunch Graph Matching).
5) Kernel Methods.
6) Trace Transform.
7) DTC (Discrete cosine transform).

Our concern is to analyze the PCA- Principal Component Analysis in details.

iii. Data Compression and Data Reconstruction

Training images are the images that we store in the database .Training images are calculated in PCA by taking its grayscale pixel values in the m×n matrix where m×n is the image dimension. We find some data from these images which will be ready to be compared with the test image data.

The gray-scale pixel values of training images are kept in the matrix I_i, where i = 1,2,3,4……N.

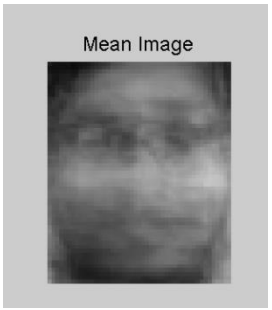
Mean Image [5] is the average image vector, which is calculated by adding all images and dividing the added images by the number of image numbers. The mean image is a column vector such that each entry is the mean of all corresponding pixels of the training images.

$$M=\frac{1}{N}\sum_{i=1}^{N} x_i$$

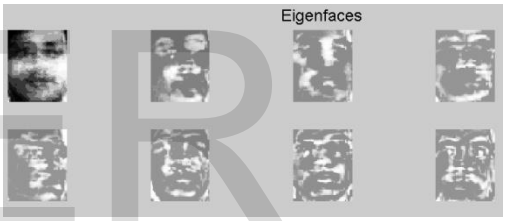
$$I_i = \begin{bmatrix} a_{11} & a_{12} & .. & an \\ a_{21} & a_{22} & .. & a_{2n} \\ ... & .. & .. & .. \\ a_{m1} & a_{m2} & .. & a_{mn} \end{bmatrix} \quad mXn \longrightarrow \begin{bmatrix} a_{11} \\ \vdots \\ a_{1n} \\ \vdots \\ a_{2n} \\ \vdots \\ a_{mn} \end{bmatrix}$$

$$mnX1 = x_i$$

Where x_i is the column vector of I_i.



Eigenfaces are a set of eigenvectors which PCA recognizes as human face. These are then resampled at the same pixel resolution. [6]. Eigenface has a ghost like appearance. These Eigenfaces are arranged in special patterns. These single patterns are important features of the face that will be calculated and stored. The image of the eigenface looks like a shadow of a human face. Since these eigenvectors have a face like appearance, they are called Eigenfaces. Some Eigenfaces of our local database are given below:



Euclidean distance [7] between two points that is given by the Pythagorean formula. By using this formula as distance, Euclidean space becomes a metric space. The Euclidean distance between points p and q is the length of the line segment pq. In Cartesian coordinates, if p = (p_1, p_2,..., p_n) and q = (q_1, q_2,..., q_n) are two points in Euclidean *n*-space, then the distance from p to q is given by:

$$\mathrm{d}(\mathbf{p},\mathbf{q}) = \sqrt{(p_1-q_1)^2+(p_2-q_2)^2+\cdots+(p_n-q_n)^2} = \sqrt{\sum_{i=1}^{n}(p_i-q_i)^2}.$$

In the Euclidean plane, if p = (p_1, p_2) and q = (q_1, q_2) then the distance is given by

$$\mathrm{d}(\mathbf{p},\mathbf{q}) = \sqrt{(p_1-q_1)^2+(p_2-q_2)^2}.$$

L1-Norm:

L1 distance [8] is measured by the following equation: $d(x,y)=|x-y|=\sum_{i=1}^{k}|x_i - y_i|$

This is the summation of all the difference of two vectors.

L2-Norm:

The L2-norm [8] is determined by:

$$d(x,y)=\|x-y\|=\sum_{i=1}^{k}(x_i - y_i)^2$$

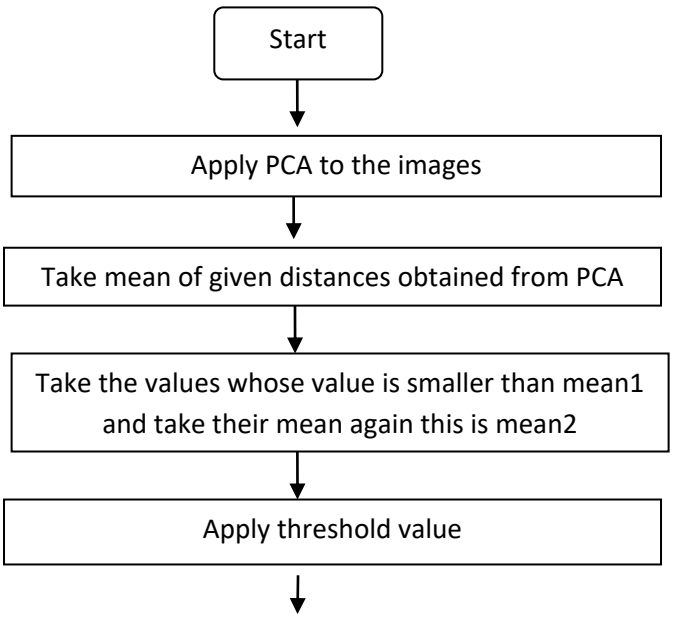
This is the summation of all the square difference of two vectors.

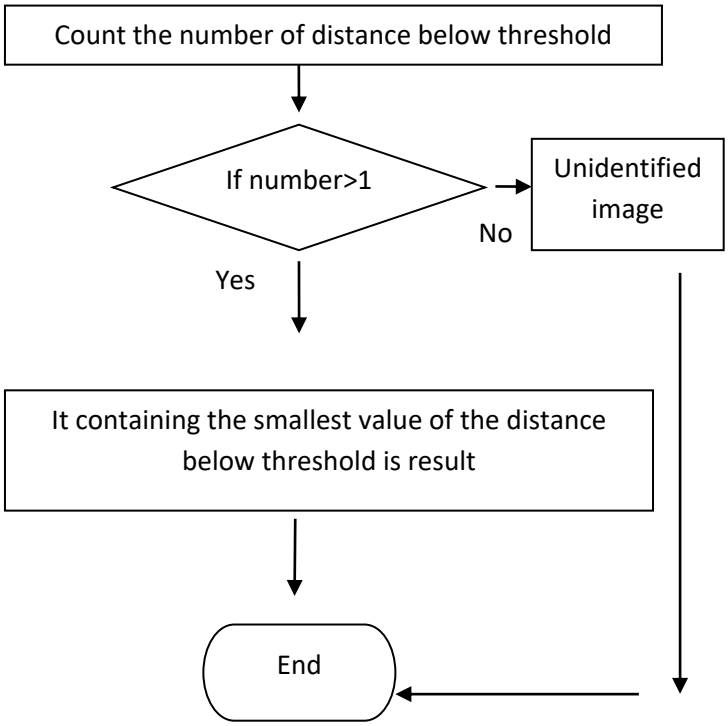
iv. Methods and Methodology

Our aim is to introduce a new technique to find matched face images with a database as well as detect non-face images. A PCA-based face recognition system assumes the lowest distance is the face. If a non-face object like a ball or coconut is given for matching, it also predicts it matches with somebody. Our proposed algorithm minimizes these non-faces false detection and also improves the original detection rate. For this we have considered the LI norm, L2 norm and Euclidean distance. We have tried the ORL database and our local database with test images and non-face objects.

v. Algorithm for Face Recognition

The algorithm we proposed is based on distance measured in PCA. Flow Chart of our proposed algorithm is given below:



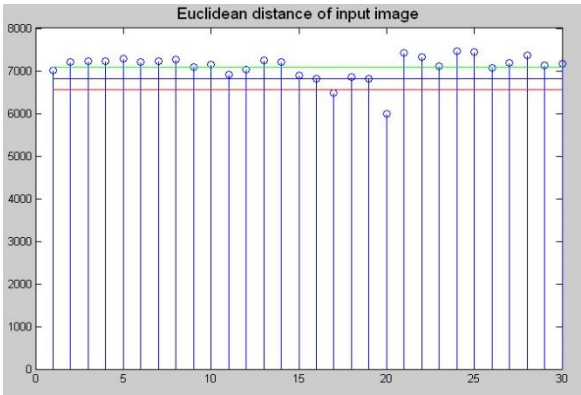


We first took the average of all the distances of the faces. We will call it mean 1. It will divide the distance into two parts upper and lower. As the upper part is not in our interest, we will discard all the values of the upper part. Below the mean all the values are more significant. Take again the mean of these values. It will also divide the values into two sections. Now we have to determine the threshold value. We have tried to find out the relation between the number of images and threshold values. From our observation we have seen that as the number of images increases, more and more value will come under the mean 2 .So we take Log_2 (n)+constant. We choose this constant value heuristically. Now for the threshold value we subtract (mean 2*Log Log_2 (n) + constant)/100 from the mean 2. There will be some distance value below the threshold value. We have to take the smallest value of all .And this is our result.

PRR = Log (N) + 1.5, where N is the number of database images.

Threshold =Mean 2-(Mean2*PRR)/100

If any face that is not belongs to the database should be above the threshold value. One reason for the counting the number of the distance below the threshold is we have seen images from the same person should have nearest distance of the lowest distance. We have experimented with image databases and found some results, which is very hopeful.



vi. Experimental Results

Tests Performed Using The Local Face Database:

We have tried it on our local database. It consists of 5 images of 6 person total of 36 images of different orientations. 3 unknown image of 3 individual persons and 10 non-face image. We applied a different PRR value to see how it behaves on change of this value. The experiment data have given in table 1 to Table 7.

From the above data, we can see L1 and L2 both have high recognition rate but low percentage in false detection, but Euclidean distance has low recognition rate but high rate in false detection. In normal PCA there is no level that can determine whether the test image belongs to this database or not .So we have tried to find out the process that can tell us whether the face belongs to this database or not. We have applied threshold value, above this value; no distance of face will belong to this database. We applied a heuristic here. So we tried different values like 0.5, 1, 1.5 etc with $Log_2(n)$ where n is the total number of training images. From the tables we can see that using $Log_2(n)+1.5$ gives a moderate result. L1 finds 3 images out of six and 1 image is false detected .1 nonface image is not matched and 9 have false detected images is not matched in unknown face and 2 are false detected. For L2 4 images have matched, 2 images not matched. For an unknown face all three images were false detected. And for non face, 2 images are not matches out of 10 images and 8 have false detected. For Euclidean distance success rates of known face are little low. 2 images are matched out of 6 and 4 are not matched. All of the unknown face are not matched which is the correct result. And 9 images are not matched in the database and 1 have false detection. So after observing the results of our local database, we decided to see how it behaves in an international database. We have also tried PCA and our algorithm on ORL database.

Tests Performed Using The ORL Database:

For experiments, the first five images were taken as a training set and the rest of the images were taken as test images of each individual. There are 40 person's images in the database and each has 10 images of a different pose. Therefore, the total number of images is 400. And there are also 10 non-face test image. In the experiment, the same image wasn't given for the test. The original images have dimensions of 92×112 pixels. We also tried on lower dimension images from the ORL database of 80×80 dimensions. Next, we took the first 3 images as training; the test images were the same. After the experiment, we got some good results.

Conclusion

In this paper, we have given an overview of face recognition research activities. After completing our experiment, we have gained various results from different portions of face recognition. We first have explained our face recognition algorithm. The performance of our algorithm is good. It will work well under various constraints.

We have tried it for ORL database and our database. False detection rate has been reduced in both L1 and L2 distance measurements without harming the original success rate of PCA. The time required to compute the whole operation is also moderate. We have experimented with different numbers of test and training images. We have also experimented with different dimensions of test and training images. Now it is a matter of great joy that we have found good results by experimenting based on different dimensions of training and test images and different numbers of test and training images including non-face object.

Future Work :

In the previous PCA there was a limitation like data redundancy for the cause of taking different kinds of poses of the same image, but our proposed working plan also doesn't remove this limitation. Our future plan is to take only one pose image for every person, so that data redundancy will not be occur.

References:

[1] Manal Abdullah, Majda Wazzan and Sahar Bo-saeed (2012). Optimizing Face Recognition Using PCA,International Journal of Artificial Intelligence & Applications (IJAIA), 3(2),23-31.

[2] Jonathon Shlens, A Tutorial on Principal Component Analysis(2009), Center for Neural Science, New York University; Version 3.01.

[3] Eleyan, A., & Demirel, H. (2007). PCA and LDA Based Neural Networks for Human Face Recognition. Face Recognition. https://doi.org/10.5772/4833

[4] Shahrin Azuan Nazeer and Marzuki Khalid (2009), PCA-ANN Face recognition System based on Photometric Normalization Techniques, Telekom Research & Development Sdn. Bhd., Universiti Teknologi , Malaysia.

[5] Wendy S. Yambor (2000), Analysis of pca-based and fisher discriminant-based image recognition algorithms.

[6] http://en.wikipedia.org/wiki/Eigenface

[7]https://en.wikipedia.org/wiki/Euclidean_distance

[8] Hyeonjoon moon and P Jonathon Phillips (2001), Computational and performance aspect of PCA Face recognition algorithms, 30, 303-321.

	Known image (6)			Unknown image (3)			Object image(10)		
	Match image	No match	False detection	Match image	No match	False detection	Match image	No match	False detection
L1	2	2	2	0	0	3	0	0	10
L2	5	1	0	0	0	3	0	0	10
Ed	3	2	1	0	3	0	0	4	6

Table 1: Local database result of our proposed algorithm using Log(n)

	Known image (6)			Unknown image (3)			Object image(10)		
	Match image	No match	False detection	Match image	No match	False detection	Match image	No match	False detection
L1	3	2	1	0	0	3	0	0	10
L2	5	0	1	0	0	3	0	1	9
Ed	3	3	0	0	3	0	0	8	2

Table 2: Local database result of our proposed algorithm using Log(n)+.5

	Known image (6)			Unknown image (3)			Object image(10)		
	Match image	No match	False detection	Match image	No match	False detection	Match image	No match	False detection
L1	3	2	1	0	0	3	0	1	9
L2	4	2	0	0	0	3	0	2	8
Ed	2	4	0	0	3	0	0	9	1

Table 3: Local database result of our proposed algorithm using Log(n)+1

	Known image (6)			Unknown image (3)			Object image(10)		
	Match image	No match	False detection	Match image	No match	False detection	Match image	No match	False detection
L1	3	2	1	0	1	2	0	1	9
L2	4	2	0	0	0	3	0	2	8
Ed	2	4	0	0	3	0	0	9	1

Table 4: Local database result of our proposed algorithm using Log(n)+1.5

	Known image (6)			Unknown image (3)			Object image(10)		
	Match image	No match	False detection	Match image	No match	False detection	Match image	No match	False detection
L1	3	2	1	0	1	2	0	1	9
L2	3	3	0	0	1	2	0	2	8
Ed	0	6	0	0	3	0	0	10	0

Table 5: Local database result of our proposed algorithm using Log(n)+2

	Known image (6)			Unknown image (3)			Object image(10)		
	Match image	No match	False detection	Match image	No match	False detection	Match image	No match	False detection
L1	2	3	1	0	1	2	0	1	9
L2	3	3	0	0	1	2	0	3	7
Ed	0	6	0	0	3	0	0	10	0

Table 6: Local database result of our proposed algorithm using Log(n)+2.5

	Known image (6)			Unknown image (3)			Object image(10)		
	Match image	No match	False detection	Match image	No match	False detection	Match image	No match	False detection
L1	2	3	1	0	2	1	0	2	8
L2	3	3	0	0	1	2	0	3	7
Ed	0	6	0	0	3	0	0	10	0

Table 7: Local database result of our proposed algorithm using Log(n)+3