

MILITARY TECHNICAL COLLEGE  
CAIRO-EGYPT



FIRST INTERNATIONAL CONF. ON  
ELECTRICAL ENGINEERING

## A STATISTICAL CONNECTIONIST APPROACH FOR FACE RECOGNITION

M. Shaarawy\*, H. Ismail\*\*, K. Hassanain\*\*\*

### ABSTRACT

Face recognition could be applied to a variety of practical applications and problems, including security and criminal identification systems. Face recognition using eigenface approach was motivated by information theory as it provides a practical solution. In this paper, the Principal Component Analysis (PCA) is used for eigenfaces (eigenvectors) computation. These eigenfaces present the extracted features for the faces to be recognized. A multilayer Artificial Neural Network (ANN) with back propagation adaptive learning algorithm is used for the classification phase. A number of experiments have been conducted on the system using the Olivetti Research Laboratory (ORL) database. Promising results have been achieved. Total performance accuracy on the data set used reached 98%.

### KEYWORDS

Face Recognition, Eigenfaces, Connectionist, and Principal Component Analysis.

### 1- INTRODUCTION

Biometry can be defined as the science of identifying an individual by analyzing his or her physical characteristics, for example, fingerprints, voiceprints, retina patterns, hand geometry, face recognition, etc. [1]. In the recent years, the demand for a high-performance, reliable personal identification access control systems based upon biometrics is growing rapidly. They are also referred to as physical attribute systems and personal characteristics systems. Biometric systems can be categorized broadly into identification systems and verification systems [2]. *Identification systems* are used to identify whether a

\* Associate Professor, Egyptian Armed Forces.

\*\* Ph.D., Egyptian Armed Forces.

\*\*\* Eng., Technical Research Department, Cairo, Egypt.

person matches anybody in an existing database of people (i.e. to search a database of stored templates). It is usually not necessary for recognition to be done in real-time and often only one image is available per person. While *Verification systems* are used to verify that a person matches their previous template (i.e. to prove that someone is who they say they are). Multiple images per person are often available for training and real-time recognition is required.

Biometric systems never provide an absolute certain identification. A human guard is not able to do that either. To evaluate the reliability of recognition, two parameters are computed: False Reject Rate (FRR) Type 1 errors and False Accept Rate (FAR) Type 2 errors [2]. *False-reject rate* is defined as deny admittance to a person, who should have been admitted, while *false-accept rate* is defined as emit admittance to a person who should not have been admitted. These parameters are also used in testing the performance of our verification system. Face recognition has the highest rate of acceptance among users of biometric systems, as most people are used to being photographed, and it is the manner in which people most recognize one another. However, achieving highly reliable recognition of a significant number of individuals under realistic viewing conditions has proved difficult. This is mainly because faces are complex objects that vary in appearance with time [4].

In this paper, a statistical connectionist approach for face recognition is addressed. Face recognition techniques can be categorized mainly into feature-based approach or image-based approach, some of the attempts in both approaches are discussed in section 2. Steps of eigenfaces computation and the design of a multilayer ANN trained with back propagation adaptive learning algorithm as a classifier are explained in section 3. Experiments and results are discussed in section 4. Conclusions are drawn in section 5.

## 2- RELATED WORK

A brief overview of the attempts in face recognition is given in this section. There are two categories to get a one-to-one correspondence between a stimulus (face to be recognized) and its stored representation (face in the database) [4]: feature-based approach and image-based approach.

### Feature-Based Approach

Typically in this approach, faces are represented either in terms of distances (spatial location of facial features) and their geometrical relations (angles and ratios between features) such as the eyes, nose, chin, etc. [3-8], or in terms of more complex functions like autocorrelation or moment invariants [6]. These features are detected either manually or by automatic detection algorithm [7]. The feature-based approach is referred to as geometrical feature approach [8], feature-based coding [3] and in ANN literature non-connectionist

model [6]. The main advantage of the approach is its efficient way of storing information. The main drawback of the approach includes [9-11]:

- Selecting a set of features that capture the appropriate information is not easy.
- The approach is dependent on the accuracy of the feature location algorithm.
- Geometric distortion due to varying perspectives in viewing faces that are not head-on will prove difficult.
- Computational time is considerable high.
- Obscurants like glasses will limit performance.

### **Image-Based Approach**

This approach is based on the direct processing [7], global or holistic analysis [12] of the gray level images (2D-pixel intensity array). In the ANN literature it is known as connectionist model [6] and in the statistical community it is known as image-based coding of the faces. The facial features are extracted implicitly. The image-based approach is more robust compared to the geometric approach under realistic viewing conditions. Its drawback includes that the images are sensitive to substantial variations in lighting conditions, head orientation, and size, which can be overcome by preprocessing the faces (normalization for size and position). There are various techniques, which can be classified under this, approach such as: Hidden Markov Models [13-14], Template Matching [15], Artificial Neural Networks [6], Principal Component Analysis [16-17], etc.

The use of Karhunen-Loève (K-L) expansion for the representation and recognition of faces has generated renewed interest. Sirovich and Kirby [16] represented images (cropped faces) using K-L expansion. They noted that the number of images  $M$  usually available for computing the covariance matrix of the data is much less than the row or column dimensionality of the covariance matrix, leading to the singularity of the matrix. They used a standard method from linear algebra that calculates only  $M$  eigenvectors that do not belong to the null space of the degenerate matrix. Once the eigenvectors (eigenpictures) are obtained, any image in the ensemble can be approximately reconstructed using a weighted combination of eigenvectors. Also Turk and Pentland [5] used eigenvectors (eigenfaces) for face detection and identification. Given the eigenfaces, every face in the database can be represented as a vector of weights. The weights are obtained by projecting the image into eigenface components by a simple inner product operation. The identification of a test image is done by representing it in a vector of weights and locating the image in the database whose weights are the closest to the weights of the test image using Euclidean distance.

### **3- SYSTEM DETAILS**

The proposed verification system used for face recognition is an image-based approach, it is composed of two stages: statistical extraction of the facial features using the PCA and an

ANN classifier for the classification phase. The experiments are achieved using the ORL database.

**Extraction of Facial Features (Eigenfaces)**

The first stage is based upon the computation of “eigenfaces” [5] that relies on the statistical analysis tool PCA. It is a common statistical method for analyzing data. It is a means of extracting the salient features of a mass of data, to draw any reliable conclusions about the presence or absence of correlation in the data [18]. In other words, to find a set of  $M$  orthogonal vectors in data space that accounts for much as possible of the data’s variance [19]. Projecting the data from their original  $N$ -dimensional space into  $M$ -dimensional subspace spanned by these vectors then performs a dimensionality reduction that often retains most of the intrinsic information in the data. Typically  $M \ll N$ , making the reduced data much easier to handle. In communication theory it is known as the Karhunen-Loève (K-L) transform. It is closely related to least square method, factor analysis, Singular Value Decomposition (SVD), and matched filtering [19]. In image processing literature it is commonly referred to as Hotelling transform [20].

The “eigenface” technique treats a normalized image of face as a high-dimensional vector, the distribution of the vectors is captured by a lower-dimensional linear subspace spanned by the principal eigenvectors of their covariance matrix. This compact lower-dimensional representation can then be used for coding and recognition of faces.

Follows are the steps for the computation of eigenfaces:

Step 1: Consider the training set of  $M$  face images, where each face is of dimension  $(N \times N)$  and its vector form is  $(N^2)$ .

$$f_1, f_2, f_3, \dots, f_M \tag{1}$$

Step 2: The average of the training set is calculated by

$$\Psi = \frac{1}{M} \sum_{i=1}^M f_i \tag{2}$$

Step 3: Each face differs from the average by the vector

$$\Phi_i = f_i - \Psi \quad i = 1 \dots M \tag{3}$$

Normally, the vectors  $\Phi_i$  are subject to PCA to seek a set of  $M$  orthogonal vectors which best describes the distribution of the data, thus a covariance matrix  $C$  defined by

$$C = \frac{1}{M} \sum_{i=1}^M \Phi_i \Phi_i^T \tag{4}$$

Would be constructed. This matrix is of dimension  $(N^2 \times N^2)$  and the calculation of  $N^2$  eigenvectors associated with the eigenvalues is a time consuming process. Since the number of faces in the training set  $M$  is less than the dimension of a face  $N^2$ , then there will be only  $M$  meaningful eigenvectors. So we can solve the eigenvectors of an  $(M \times M)$  matrix, than taking an appropriate linear combinations of the vectors.

Step 4: Construct a matrix  $A$  ( $N^2 \times M$ ), where

$$A = [\Phi_1 \Phi_2 \dots \dots \Phi_M] \quad (5)$$

And its transpose is  $A^T$ .

Step 5: Construct a matrix  $L$  ( $M \times M$ ), where

$$L = A^T A \quad (6)$$

Step 6: Compute the eigenvalues  $\lambda_j$  and its associated eigenvectors  $V_j$  of matrix  $L$ , for  $j = 1 \dots M$ , such that:

$$L V_j = \lambda_j V_j \quad (7)$$

Step 7: Sort the eigenvalues such that:

$$\lambda_1 > \lambda_2 > \lambda_3 \dots \dots > \lambda_M \quad (8)$$

So as to rank the eigenvectors  $V_j$  according to their usefulness in characterizing the variation among the training set of faces.

Step 8: These eigenvectors determine the linear combinations of the  $M$  differenced vectors  $\Phi_i$  to form the eigenfaces

$$U_j = \sum_{k=1}^M V_{jk} \Phi_k \quad j = 1 \dots M \quad (9)$$

In practice  $K$  eigenfaces, where  $K < M$ , that correspond to the  $K$  eigenvectors with the highest eigenvalues are sufficient for human face identification

Step 9: For each face image in the training / testing set. Transform it into its eigenface components, i.e. projecting the original face image onto the low-dimensional face space by

$$\omega_j = U_j^T (f - \Psi) \quad j = 1 \dots K \quad (10)$$

These weights form a vector

$$\Omega^T = [\omega_1 \omega_2 \dots \omega_k] \tag{11}$$

That describes the contribution of each eigenface in representing the face image. This vector can be used then for classification.

**Classification**

A three-layer ANN that learns with the backpropagation algorithm can be understood as a layered system of interacting nodes. Each node is a single processing element that acts on the data to produce a result. Typically, the nodes are arranged in layers and are each connected to the nodes in the preceding layer for input and the following layer for output [22]. Furthermore, each connection has an associated adjustable value called weights, see Fig.1, of a typically 3-layered ANN. The key to BackPropagation (BP) learning algorithm is its ability to change the values of the weights of the connections between the nodes in response to errors. The error function is defined by

$$E = \sum_p E_p = \frac{1}{2} \sum_p \sum_j (t_{pj} - o_{pj})^2 \tag{12}$$

Where summation is performed over all output nodes  $j$  and patterns  $p$ ,  $o_j$  is actual output and  $t_j$  is target or desired output for a given input pattern. For it to be possible to calculate the errors, the training data must contain a series of input patterns labeled with their target output patterns. The aim of minimizing the sum of the ANN's squared errors is to find a set of weights that processes training data accurately enough for the application. Saving these trained weights preserves what the ANN learned and enables it to function as a classifier without further training [21-23].

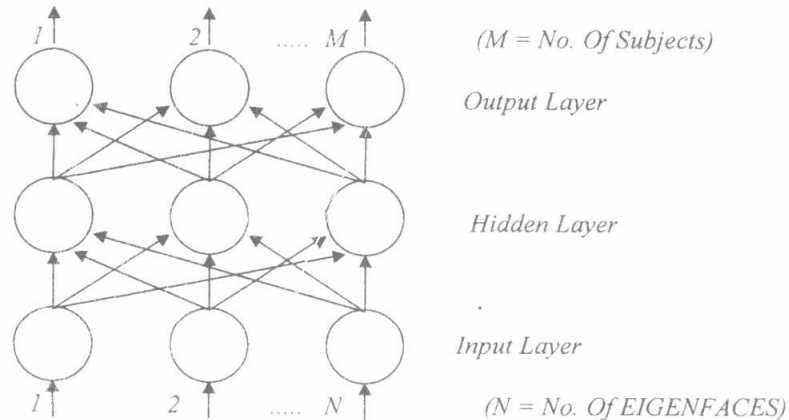


Fig.1. A typical three layer fully connected ANN:

**Data**

The ORL database consists of 400 images of 40 individuals which contains quite a high degree of variability in expression, pose, and facial details, taken between April 1992 and

April 1994 at the Olivetti Research Laboratory in Cambridge, UK. It is available free of charge, see <http://www.cam-orl.co.uk/facedatabase.html>. There are 10 different images of 40 distinct subjects. The images were taken at different times. There are variations in facial expression (open/closed eyes, smiling/non-smiling), and facial details (glasses/no glasses). All the images were taken against a dark homogeneous background with the subjects in an up-right, frontal position, with tolerance for some tilting and rotation of up to about 20 degrees. There is some variation in scale of up to about 10%. The images are grayscale with a resolution of 92x112. The images are normalized, thus a preprocessing phase is not required.

#### 4- EXPERIMENTAL RESULTS

In this section, some of the achieved experimental results are presented to test the performance of the proposed verification system. During experiments two problems raised. First the required number of eigenfaces, for adjusting the classifier parameters. Second, determining the number of neurons in the hidden layer, since they are application dependent. The experiments were conducted over the ORL database, which consists of 40 subjects with 10 faces per subject. The 40 subjects are divided into two groups, 20 subject per group. Fig.2, shows a sample of the ORL database, Fig.3, shows some of the computed eigenfaces. In these experiments, the first group is assumed for admittance through the proposed verification system, while the second group is not allowed for admittance.

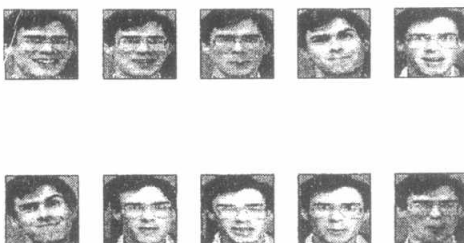


Fig.2. Sample of the ORL databases.



Fig.3. Some of the computed EIGENFACES.

Two strategies were adopted Quarter strategy and Half strategy. These strategies are based on taking quarter (half) the number of the training data set as Eigenfaces for quarter (half)

strategy, respectively. A fully connected 3 layer ANN with back propagation learning algorithm is used with learning parameters; learning rate = 0.01, momentum = 0.5, number of epochs = 5000, error rate = 0.001, no. of neurons in the output layer = 20, and no. of neurons in the input layer = no. of EIGENFACES. Some of the achieved results are discussed as follows:

1- The relation between the number of neurons in the hidden layer against the number of eigenfaces. As shown in Fig.4. For quarter strategy the number of faces rejected decreases as the number of training faces increases and vice versa, while for half strategy it seems that increasing the number of eigenfaces over the training data set doesn't necessarily achieves better performance. This may be due to the increased complexity of the ANN for half strategy and also for the non-linearity of the ANN learning rules.

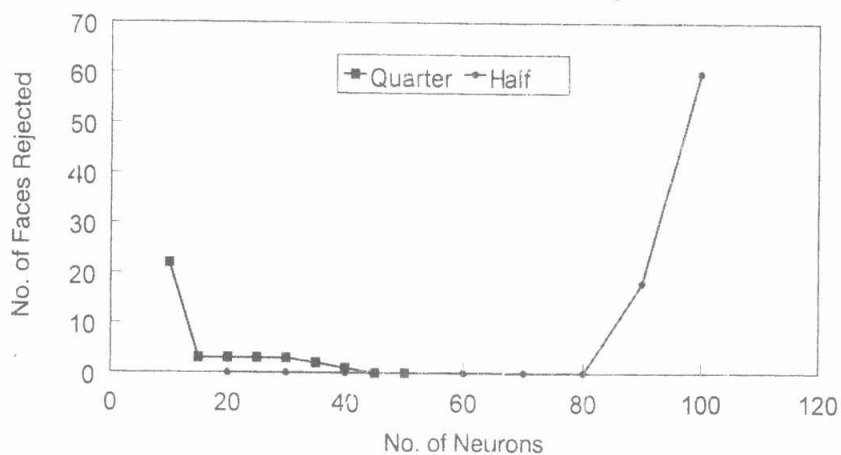


Fig.4. Number of neurons in the hidden layer versus the number of faces rejected

2- The FRR for a fixed (variant) number of neurons in the hidden for both quarter and half strategies. We noted that the FAR was zero for all experiments conducted, while the FRR decreases as the number of training faces per subject increases. Fig.5. Shows the FRR for quarter and half strategies for a constant number of neurons in the hidden layer, which were 50 neurons. While Fig.6. For a variant number of neurons taking the number of neurons equal to the number of eigenfaces computed over the training set.

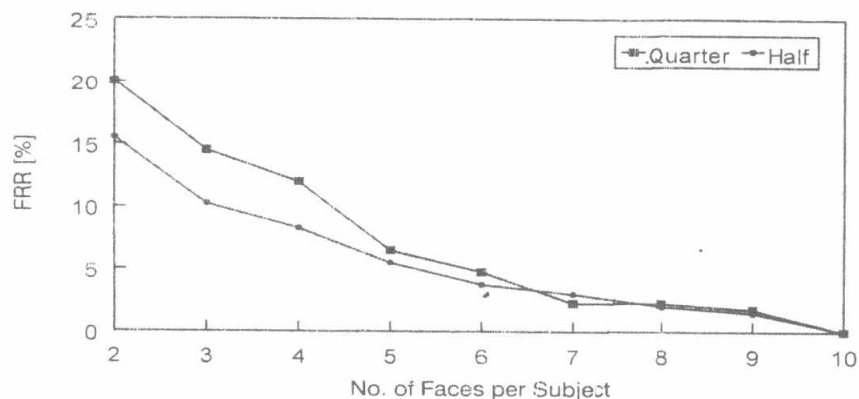


Fig.5.No. Of Faces vs. FRR a Constant number of neurons = 50



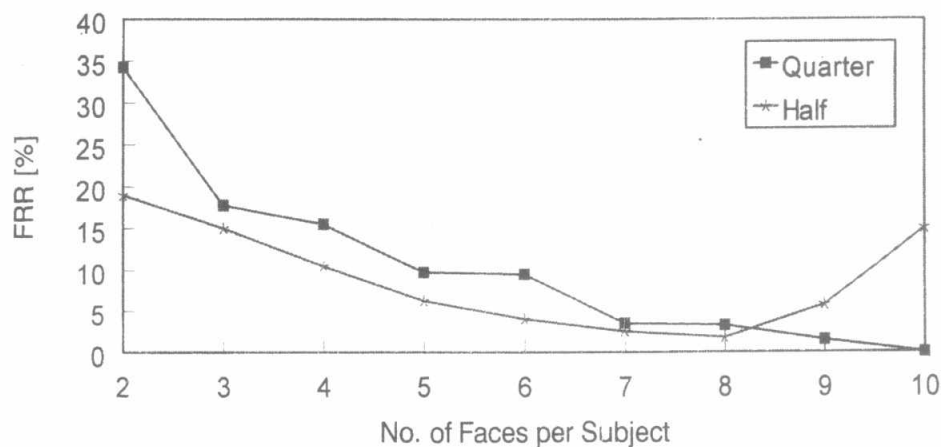


Fig.6. No. Of Faces vs. FRR a variant number of neurons.

## 5- CONCLUSIONS

The Eigenface approach to face recognition was motivated by information theory. It provides a practical solution, as it is relatively simple and fast for verification systems. We concluded that:

- There is a certain limit to the number of neurons in the hidden layer and it is not a direct proportional relation with the number of eigenfaces.
- Half the number of eigenfaces of the training data set is sufficient for recognition.
- The FAR for all experiments conducted were zero, therefore it is highly recommended for security applications.
- The FRR decreases as the number of training faces per subject in the training set increases.

## REFERNCES

- [1] Philip Walker, "Electronic Security Systems: Better Ways To Crime Prevention", Second edition, Butterworth publishers, (1988).
- [2] Lawrence J. Fennelly, "Effective Physical Security", Butterworth publishers, (1992).
- [3] Ingemar J Cox, Joumana Ghosn, and Peter N. Yianilos, "Feature-Based Face Recognition Using Mixture-Distance", Technical Report, NEC Research Institute, NJ, October (1995).
- [4] R. Brunelli, and T. Poggio, "Face Recognition through Geometrical Features", Proceedings of ECCV '92, pp. 792-800, (1992).
- [5] Matthew Turk and Alex Pentland, "Eigenfaces for Recognition", Journal of Cognitive Neuroscience, Vol. 3, No. 1, (1991).

- [6] Dominique Valentin, Herve Abdi, Alice J. O'Toole, and Garrison W. Cottrell, "Connectionist Models of Face Processing: A Survey", *Pattern Recognition*, Vol. 27, pp. 1209-1230, (1994).
- [7] Nathan Intrator, Daniel Reissfeld, and Yehezkel Yeshurun, "Face Recognition using a Hybrid Supervised/Unsupervised Neural Network", Technical Report, Department of Computer Science, Tel-Aviv University, Israel, (1995).
- [8] Seteve Lawrence, C. Lee Giles, Ah Chung Tsoi, and Andrew D. Back, "Face Recognition: A Convolutional Neural Network Approach", Technical Report UMIACS-TR-96-16, Institute for Advanced Computer Studies, University of Maryland, (1996).
- [9] T. Kanade, "Picture Processing by Computer Complex and Recognition of Human Faces", Technical Report, Kyoto University, Dept. of Information Science, (1973).
- [10] Ian Craw, David Tock, and Alan Bennett, "Finding Face Features", *Proceedings of European Conference on Computer Vision, ECCV-92*, (1992).
- [11] R. Brunelli, and T. Poggio, "Face Recognition: Features versus Templates", *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)*, 15(10):pp. 1042-1052, October (1993).
- [12] Joseph Wilder, "Face Recognition using Transform Coding of Gray Scale Projections and Neural Tree Network", *Artificial Neural Networks for Speech and Vision*, Edited by Richard J. Mammone, Published by Chapman & Hall, chapter 28, (1993).
- [13] F. Samaria, "Face Identification Using Hidden Markov Models", 1<sup>st</sup> Year Report, Cambridge University Engineering Department, (1992).
- [14] F. Samaria and A. Harter, "Parameterization of Stochastic Model for Human Face Identification", 2<sup>nd</sup> IEEE workshop on Applications of Computer Vision, Sarasota, Florida, (1994).
- [15] Alan L. Yuille, "Deformable Templates for Face Recognition", *Journal of Cognitive Neuroscience*, 3(1), pp. 59-70, (1991).
- [16] L. Sirovich and M. Kirby, "Low-Dimensional Procedure for the Characterization of Human Faces", *Journal of the Optical Society of America A*, 4(3), 519 – 524, (1991).
- [17] M. Kirby and L. Sirovich, "Application of Karhunen-Loève Procedure for the Characterization of Human Faces", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 12(1), (1990).
- [18] A. Jennings, J. Mckeown, "Matrix Computation", Published by John-Willy and Sons Ltd., 2<sup>nd</sup> edition, (1992).
- [19] John Hertz, Andres Krogh, and Richard G. Palmer, "Introduction To The Theory Of Neural Computation", *Lecture Notes, Volume 1*, (1991).
- [20] Rafael Gonzalez and Richard Woods, "Digital Image Processing", Addison-Wesley Publishing co., (1993).
- [21] Dan Hammerstrom, "Working with Neural Networks", *IEEE Spectrum*, July (1993).
- [22] Richard P. Lippmann, "An Introduction to Computing with Neural Nets", *IEEE ASSP Magazine*, April (1987).
- [23] R. Beale and T. Jackson, "Neural Computing: An Introduction", Published by Adam Hilger, (1990).