FACE RECOGNITION USING ARTIFICIAL NEURAL NETWORK

Sentayehu Endeshaw and Kumudha Raimond
Department of Electrical and Computer Engineering
Addis Ababa University

ABSTRACT

Face recognition (FR) is one of the biometric methods to identify the individuals by the features of face. Two Face Recognition Systems (FRS) based on Artificial Neural Network (ANN) have been proposed in this paper based on feature extraction techniques. In the first system, Principal Component Analysis (PCA) has been used to extract the features from face images and classify them using ANN. In the second system, combination of Gabor Filter (GF) and PCA have been used for feature extraction and ANN for classification. The influence of different ANN parameters also has been studied in this work. Experiments have been carried out by using Olivetti Research Laboratory (ORL) face database. The results confirmed the feasibility of the methodologies followed in this work. Further, the systems performed very efficiently when subjected to new unseen images with a false rejection rate of 0% during testing.

Keywords: *Face Recognition, Biometrics, Artificial Neural Network, Gabor Filter, Principal Component Analysis.*

INTRODUCTION

Security has become an important social issue. To maintain security, identification and authentication methods are used in various areas of day-to-day activities like entrance control in buildings, access control for automatic teller machines, withdrawing money from a bank account, etc.

There are different forms of authentication methods like identity cards, PIN codes, human recognition (e.g. a concierge in a building), passports, keys, batches, or passwords. More sophisticated applications use fingerprints or retina scans, basic voice recognition, or a combination of the aforementioned techniques. However, except for face and voice recognition, other methods require the user to remember a password, to enter a PIN code, to carry a batch and all these means are prone to being lost or forgotten, whereas fingerprints and retina scans suffer from low user acceptance.

Some of the above mentioned techniques can also be grouped under biometric techniques, which is defined as "The identification of an individual based on biological traits, such as fingerprints, iris patterns, and facial features" [7]. As mentioned earlier, most biometric techniques are inconvenient due to the necessity of interaction with the individual who is to be identified or authenticated. FR on the other hand, can be a non-intrusive technique. This is one of the reasons for an increased interest in FR. Further, it also received significant attention recently due to increasing public concerns for security, especially, due to many events of terror around the world after September 11th 2001 [5].

LITERATURE SURVEY

A formal method of classifying faces was first proposed by Francis Galton in 1888 [1, 2]. Later, works on FR remained largely dormant. However, during 1990's, the research interest in FR has again grown up significantly.

Lately, FR technology has numerous commercial and law enforcement applications ranging from static matching of controlled format photographs such as passports, credit cards, photo ID's, driver's license etc to uncontrolled video images [12]. FRS that has emerged is capable of achieving a recognition rate of greater than 90% under controlled conditions [13].

Researches in this area are going on for many years; as a result, the current status of FR technology is well advanced. Some of the findings are mentioned below:

FR can be done for still images as well as for video which has its origin in still-image FR. Different approaches of FR for still images can be categorized into three main groups such as holistic approach (the whole face region used as input data), feature-based approach (local features on face such as nose, and eyes are segmented and used), and hybrid approach (both local feature and whole face) [16].

The issue of "which features, humans use for FR" has been studied and it has been argued that both holistic and feature information are crucial for perception and recognition of faces. However, holistic descriptions may not be used if dominant features such as big ears, a small nose, etc are present [12].

Some studies concluded that the low frequency components are needed for the global description of the individual while the high frequency components are required for finer details [16].

Hair, eyes, mouth, face outlines are more important than nose for perceiving and remembering faces. It has also been found that the upper part of the face is more useful than the lower part of the face for recognition. The role of aesthetic attributes such as beauty, attractiveness and pleasantness has also been studied, with the conclusion that the more attractive the faces are, the better are their recognition rates. Also, recent studies show that an inverted face is much harder to recognize than a normal face [12].

Also, a study on the direction of illumination [15] showed the importance of top lighting; it is easier to recognize faces illuminated from top to bottom than the faces illuminated from bottom to top.

Number of statistical and Artificial Intelligence techniques have been applied for FRS. Standard statistical pattern recognition techniques and/or ANN approaches are employed for matching faces after extracting features using feature-based approach [12]. One of the well known geometrical-local feature based methods is the Elastic Bunch Graph Matching (EBGM) technique.

The other approach, the holistic one, conceptually related to template matching, attempts to identify faces using global representations and pattern classifiers [6]. One of the methods to extract features in a holistic system is by applying statistical methods such as PCA to the whole image.

As seen above, there are a lot of research findings for FR. However, successful applications under real world conditions remains a challenge as face images are subjected to a wide range of variations like pose or view angle, illumination, occlusion, facial expression and individual differences [12].

Apart from the challenge that is stated above, the other most important problem in FR is the dimensionality problem. Appropriate methods should be applied to reduce the dimension of the studied space. Working on higher dimension causes overfitting where the system starts to memorize. Also, computational complexity would be an important problem when working on large databases.

So, this work proposes an ANN based FRS, wherein the focus is on images with wide range of variations (in pose, lighting and expressions) to address the conditions imposed by many real applications and on two different feature reduction techniques to address the dimensionality problem.

FEATURE EXTRACTION TECHNIQUES

Principal Component Analysis

The Principal Component Analysis (PCA) is one of the most successful techniques that have been used in image recognition and compression applications. PCA is a statistical method and its purpose is to reduce the large dimensionality of the data space (observed variables) to the smaller intrinsic dimensionality of feature space (independent variables), which are needed to describe the data economically.

Images of faces, being similar in overall configuration, will not be randomly distributed in image space and thus can be described by a relatively low dimensional subspace. The main idea of PCA is to find the vectors which best account for the distribution of face images within the entire image space.

A 2-D facial image can be represented as 1-D vector by concatenating each row (or column) into a long thin vector. If there are M face vectors, it can be represented as $\varphi_1, \varphi_2, \varphi_3 . \varphi_M$. The average face of the set is defined by

$$\psi = \frac{1}{M} \sum_{n=1}^{M} \varphi_n \tag{1}$$

Each face differs from the average by the vector

$$\Phi_i = \varphi_i - \psi \tag{2}$$

This set of very large vectors is then subjected to PCA, which seeks a set of (M-1) orthonormal vectors, which best describes the distribution of the data. The k^{th} vector, u_k, is chosen such that

$$\lambda_k = \frac{1}{M} \sum_{n=1}^{M} (u_k^t \Phi_n)^2 \tag{3}$$

is maximum, subject to

$$u_l^t u_k = \delta_{lk} = \begin{cases} 1, & \text{if } l = k \\ 0, & \text{otherwise} \end{cases} \tag{4}$$

for $l < k$, which constrains the vectors to be orthogonal.

The vectors u_k and scalars λ_k are the significant M eigenvectors and eigenvalues, respectively, of the covariance matrix

$$C=\frac{1}{M}\sum_{n=1}^{M}\Phi_n\Phi_n^t \tag{5}$$

The significant eigenvectors are the ones used as input features for the classifier [10].

Gabor Filter

GFs have been used during several years for extracting features from images. 2-D Gabor filter is a product of an elliptical Gaussian in any rotation and a complex exponential representing a sinusoidal plane wave. The sharpness of the filter is controlled on major and minor axis by γ and η. The filter response can be normalized to have a compact closed form [8].

$$\psi(x,y,f_0,\theta)=\frac{f_0^2}{\pi\eta\gamma}e^{\frac{f_0^2}{\gamma^2}x'^2+\frac{f_0^2}{\eta^2}y'^2}\;e^{j2\pi f_0 x'}$$
$$\text{where } x'=x\cos(\theta)+y\sin(\theta) \tag{6}$$
$$y'=-x\sin(\theta)+y\cos(\theta)$$

Where f_0 is the central frequency of the filter, θ is the rotation angle of both the Gaussian major axis and the plane wave, γ is the sharpness along the major axis and η is the sharpness along the minor axis (perpendicular to the wave). The aspect ratio of the Gaussian is $\lambda = \eta / \gamma$.

Typical Gabor feature, such as Simple Gabor feature space, consists of responses calculated with Gabor filters at several different orientations and scales (frequencies): a filter bank. A filter bank consisting of several filters needs to be used because relationship between responses provides the basis for distinguishing objects since using different orientations and scales ensures invariance; objects can be recognized at various different orientations, scales and translations.

Give a bank of *'m'(a*b)* GFs $\{\psi_{u,v}(x,y),\ u=0,\ldots a,\ v=0,\ldots b\}$, image features at different locations, frequency and orientation can be extracted by convolving the image $I(x, y)$ with the filters:

$$O_{u,v}(x,y)=\left|I(x,y)*\psi_{u,v}(x,y)\right| \tag{7}$$

The selection of discrete rotation angles θ_l (from equ.6) has already been demonstrated in [4], where it was shown that the orientations must be spaced uniformly.

$$\theta_l=\frac{l2\pi}{n},\quad \text{where } l=\{0,1,2,\quad n-1\} \tag{8}$$

Where θ_l is the lth orientation and n is the total number of orientations to be used.

In the selection of discrete frequencies fk, exponential sampling must be used, that is

$$f_l=k^{-l}f_{\max},\quad \text{where } l=\{0,1,2,\ldots,m-1\} \tag{9}$$

Useful values for k (frequency scaling factor) include $k = 2$ for octave spacing and $k = \sqrt{2}$ for half-octave spacing.

The resultant GF feature set consists of the convolution results of an input image $I(x, y)$ with all of the 'm' GFs.

$$S=\{O_{u,v}(x,y): u\in\{0,\ldots a\}, v\in\{0,\ldots b\}\} \tag{10}$$

Figure 1 shows the magnitudes of Gabor representations of a sample face image with $a = 5$ (scales) and $b = 8$ (orientations). A series of row vectors $O_{u,v}$ are concatenated together to generate a discriminative Gabor feature vector:

$$G(I) = O = (O_{0,0}\, O_{0,1}\ \ldots\ O_{4,7}) \tag{11}$$

For an image of size 64×64 and 40 GFs (a=5 and b=8), the convolution result will give 64×64×5×8 =163,840 features. There may be a lot of redundant information and therefore a feature selection mechanism should be used to choose the most useful features for classification [9].

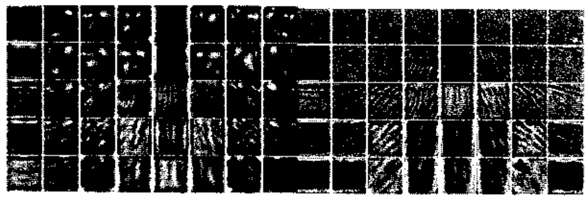

Figure 1 Convolution Result (Magnitude and Real part) of an image with 40 GFs

ARTIFICIAL NEURAL NETWORK

There has been an explosion of interest in applying ANN to the FR. ANN has several advantages that are desired in FR, including learning and adapting ability, parallel distributed computation, robustness, etc. There is an expectation that ANN techniques can lead to the realization of truly automated FRS.

In this work, Multiple layer perceptron (MLP) trained using Back Propagation Algorithm (BPA) has been chosen for the FR purpose on account of it's capability to solve difficult and diverse problems.

BPA is intended for training layered (i.e. nodes are grouped in layers), feed forward (i.e. the arcs joining nodes are unidirectional, and there are no cycles) nets, as shown in Fig. 2. It consists of an input layer, one or two hidden layers and an output layer. This approach involves supervised learning, which requires a 'teacher' that knows the correct output (desired output) for any input, and uses gradient descent on the error to train the weights.

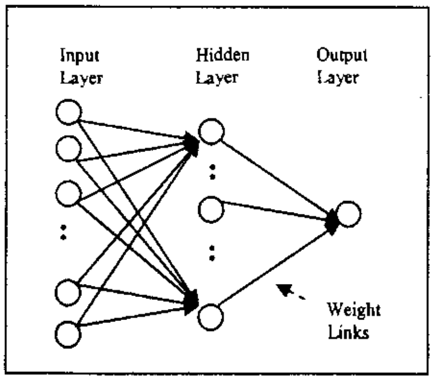

Figure 2 Architecture of multiple layer perceptron

The propagation rule, also called a summation or aggregated function, is used to combine or aggregate inputs passing through the connections from other neurons. It can be expressed as

$$Net_j = \sum_{i=1}^{n} a_i W_{ij}\ ,\ Net_k = \sum_{j=1}^{m} O_j W_{jk} \qquad (12)$$

where, i is an input neuron, j is a hidden neuron, and k is an output neuron, n is the number of input neurons in the input layer, m is the number of hidden neurons in the hidden layer, W_{ij} and W_{jk} denote weight values from input to hidden neuron, and from hidden to output neuron respectively. Net_j and Net_k represent the aggregation at the hidden and output nodes respectively. The transfer function, also called as activation function, is used to produce the output based on the value of aggregation at the hidden and output layer nodes. One of the popular transfer functions is called sigmoidal function, which can be expressed as:

$$O_j = \frac{1}{1+\exp^{-\eta Net_j}}\ ,\ O_k = \frac{1}{1+\exp^{-\eta Net_k}} \qquad (13)$$

where η is the learning rate parameter and O_j and O_k denote the output of the hidden and output neurons respectively. The actual output of the network at the output layer is compared with the desired output and the error is propagated back to adjust the weights of the network and this process is repeated until the error percentage falls into a reasonable range. The detailed description of the algorithm is available in [3].

PROPOSED SYSTEM

The following are the components of the proposed FR system (Fig. 3). Two systems have been proposed in this work. First system uses PCA for feature extraction and ANN for classification (System 1) and the second system uses GF and PCA for feature extraction and ANN for classification (System 2).

Figure 3 Proposed FR System

Face Images

ORL database contains photographs of faces taken between April 1992 and April 1994 at the ORL in Cambridge, UK. There are 10 different images of 40 distinct persons available in the database. The set of 10 images of 20th and 40th persons are shown in Fig.4.

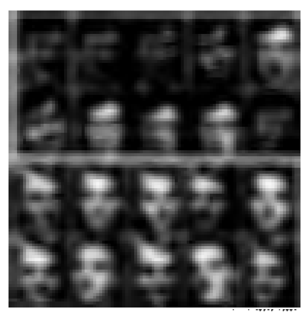

Figure 4 The set of 10 images of the 20th and 40th persons

The images were taken at different times. There are variations in facial expressions (open/closed eyes, smiling/non-smiling) and facial details (glasses/no glasses). All of the images were taken against a dark homogeneous background with the subjects in an upright, 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 92 * 112.

Face Extraction Techniques

This step is essential to extract the key features of the face images.

Two distinct models are developed to extract the features from the face images. The first model uses PCA and the second model uses a combination of GF followed by PCA.

Principal Component Analysis

While doing PCA, one has to clarify the number of training patterns available to PCA and the dimensionality reduction required from PCA. ORL face database has 400 images and each image has 10304 (112*92) pixels. The number of pixels defines the number of dimensions. As described in section III A, the next step is to find out the mean of the face images followed by covariance calculation and its eigenvectors and eigenvalues. Then reorder the eigenvectors so as to be in a descending order of "importance" (first set of eigenvectors are those whose corresponding eigenvalues are the maximum). Then depending on the dimensionality reduction level required, appropriate number of less important eigenvectors can be eliminated.

In this work, the first 100, 150, 200 and 300 principal components (eigenvectors) of the covariance matrix have been investigated. The first 200 eigenvectors have been selected for further work as it provides sufficient dimensionality reduction and optimal mean square error (MSE) when it is used to reconstruct the original signal as shown in Table 1. Lower dimensions less than 100 components have not been investigated as it is evident from Table 1 that lower dimensions will yield higher MSE. Though the MSE for 300 eigenvectors is very small, it is not selected for further work. This is due to the fact that the increase in vector size will intensify the computations and increase the time for developing an optimal ANN model. So, there is a trade-off between the feature vector size and the computations. Hence, in this work, average eigenvector of size 200 is used for training and testing the ANN model.

Table 1: MSE While Reconstructing the Original Image

Number of Eigenvectors	Reconstruction MSE
100	13.449
150	11.082
200	9.0277
300	4.7985

Gabor Feature Representation

PCA is an efficient technique to reduce the dimensionality, but has the drawback of being more sensitive to image variations. Its performance is dependent on the accuracy of normalization, and the process has no inherent invariance to translation, scale or rotation [11]. So, Gabor wavelet transform have been performed before PCA to provide a greater level of invariance than fouhd using grey-level pixel information. Gabor coefficients, the output of GF can be used as input data for PCA to provide a greater level of illumination invariance than found using grey-level pixel information [14].

The Gabor wavelet representation of an image is the convolution of the image with a family of GFs as described in section III B. To facilitate the GF representation, the ORL images are scaled to 128 X 128 using a bicubic interpolation. In this work, 40 GFs have been used. The convolution output yielded feature vectors of length 655360 (real and imaginary parts). Each output is first down sampled by a factor ρ=64 to reduce the dimensionality of the original vector space (10240 length) before giving the vectors as input to PCA.

Table 2 shows the MSE while trying to reconstruct the GF image for different eigenvector lengths. 200 eigenvectors of the covariance matrix formed from Gabor feature vectors are chosen (the selection is based on the same reason mentioned for PCA) as those with the largest associated eigenvalues. A face space is constructed using these 200 eigenfaces and each Gabor feature vector is projected to it. The dimension of the Gabor feature vector (10240) is thus reduced to 200 and it can be used as a feature vector to train and test ANN.

Table 2: MSE While Reconstructing GF Image

Number of Eigenvectors	Reconstruction MSE
100	0.055052
150	0.032981
200	0.018768
300	0.0044954

Neural Network Classifier

Figure 5 shows the functional block diagram of ANN classifier that has been used in this work.

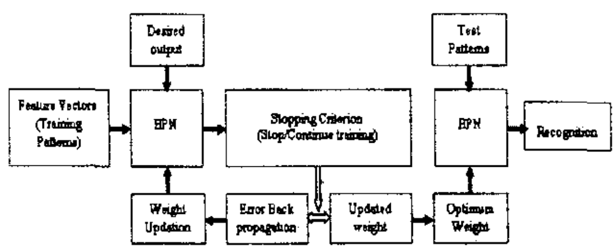

Figure 5 Functional block diagram of neural network classifier

Network Architecture Selection

As discussed in section IV, MLP which is one of the popular ANN models is used in this work. MLP is trained using BPA. MLP is a three layer feed forward network to be trained in supervised learning mode. The first layer of the network is the input layer, which is used to distribute the training patterns uniformly to the subsequent layers and since the number of nodes in this layer must be equal to the number of input data points per pattern, it has been set to 200, based on the length of the feature vector of the training and testing patterns.

The second and third layers are the hidden layer and the output layer respectively which play the role of classification or generating decision boundaries used in the categorization of the training patterns into their true classes via the adaptive adjustment and convergence of the input-to-hidden layers and hidden-to-output layers weighting coefficients to an acceptable value determined by the output layer MSE value at the last iteration.

The number of nodes in the hidden layer is determined in conjunction to the number of input layer nodes, output layer nodes and other network parameters together with the evaluation of the compromise between learning performance (convergence speed) and test performance (acceptance of the solution arrived). Based on these facts, it has been set to 100.

The output layer nodes are also used to monitor the learning performance of the network at each epoch. The number of nodes "M" required in this layer is determined by the number of persons "N" to be classified. It can be shown that those two parameters do have logarithmic relationships so that the number of nodes at the output layer can be found by using the formula $M = \text{ceil}(\log_2 N)$. So, in the present case for $N = 40$, $M = 6$. The final outputs are passed through a thresholding function of value 0.5. The output of the neuron will be equal to '1' if the activation value is $\geq$ 0.5, else it will be equal to '0'. The threshold value has been selected experimentally.

Initial input-to-hidden and hidden-to-output layers' weighting coefficients are generated randomly between -0.5 and 0.5, the bias term as usual is set to 1, the learning rate (η) and momentum (α) have been set experimentally.

Training, Testing and Validation Data Sets

In this work, standard deviation of feature vectors has been used to divide the dataset. Feature vectors (10 vectors) that represent each person are sorted in the increasing order of their standard deviation. Representative vectors are then taken for training, validation and testing as shown in Fig. 6.

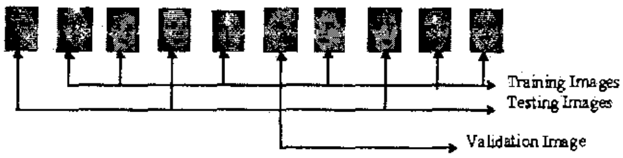

Figure 6 Sorted in increasing order of standard deviation

Thus, 6 images have been used for training, 1 for validation and 3 for testing purpose.

Stopping Criterion

In this work, the stopping criterion during training is determined by the validation set. After every 10 epochs during training, the network is tested with the validation data. The training is stopped as soon as the error on validation set increased more than the previous validation check even if it has not reached the preset epoch number. In the present case, the epoch number is set to 2000. However, the training is stopped whenever the validation error increases and the corresponding weight vectors are used for the testing purpose. In most of the cases in this work, the minimum validation error occurs at the preset epoch number.

Optimal Learning and Momentum Constants

Using MLP with 100 hidden neurons, different combinations of $\eta \in \{0.01, 0.1, 0.5, 0.9\}$ and $\alpha \in \{0, 0.1, 0.5, 0.9\}$ has been simulated to observe their effect on network convergence in both the proposed systems. Each combination has been trained with the same set of initial random weights and the same set of 240 input-output patterns (40 persons x 6 images/person), so that the results of the proposed systems can be compared directly. The performance of the proposed systems with different values of η and α is presented in the following sections 4.1 and 4.2.

System 1

The effect of η and α is shown in the following tables (Table 3) for System 1.

Table 3: Performance of the Network for Different Values of α and η for System 1

η = 0.01			
α	Minimum Training Error	Minimum Validation Error	Recognition Rate
0	0.0076408	0.24327	63.3
0.1	0.006312	0.24297	62.5
0.5	0.0023899	0.24281	61.7
0.9	0.00026806	0.2427	63.3

η = 0.1			
α	Minimum Training Error	Minimum Validation Error	Recognition Rate
0	0.00026847	0.24253	63.3
0.1	0.00023587	0.24277	63.3
0.5	0.00011599	0.24488	63.3
0.9	0.000017513	0.20568	68.3

η = 0.5			
α	Minimum Training Error	Minimum Validation Error	Recognition Rate
0	0.000038885	0.23646	65
0.1	0.000034389	0.23459	65
0.5	0.000017343	0.2075	71.7
0.9	1.2429	1.2498	0

η = 0.9			
α	Minimum Training Error	Minimum Validation Error	Recognition Rate
0	0.000019343	0.20672	69.2
0.1	0.000017167	0.19615	70.8
0.5	**8.8058E-06**	**0.14851**	**76.7**
0.9	1.25	1.25	0

System 2

Again the effect of η and α is shown in the following tables (Table 4) for System 2.

Table 4: Performance of the Network for Different Values of α and η for System 2

η = 0.01			
α	Minimum Training Error	Minimum Validation Error	Recognition Rate
0	0.012518	0.17232	70.8
0.1	0.011205	0.18245	69.1
0.5	0.0067885	0.18208	69.1
0.9	0.0044579	0.16629	70.8

η = 0.1			
α	Minimum Training Error	Minimum Validation Error	Recognition Rate
0	0.0044568	0.18479	71.6
0.1	0.0044148	0.1488	72.5
0.5	0.0042898	0.1442	71.6
0.9	0.000021415	0.11245	77.5

η = 0.5			
α	Minimum Training Error	Minimum Validation Error	Recognition Rate
0	0.0044148	0.1488	77.5
0.1	0.0021196	0.15965	71.6
0.5	0.0021015	0.1034	78.7
0.9	1.1246	1.2495	0

η = 0.9			
α	Minimum Training Error	Minimum Validation Error	Recognition Rate
0	0.0041874	0.11754	79.2
0.1	0.0021016	0.10629	74.2
0.5	**9.13E-06**	**0.090955**	**88.3**
0.9	1.25	1.25	0

In both the systems, it can be observed that as the η value increases, the network learning and validation errors also show improvement by large percentage. Further, a combination of η = {0.9} and α = {0.5} gave good recognition rate in both the systems and the same values are used for further work.

Testing

During testing, pre-processed test patterns without its associated target are propagated through the trained network. These patterns are weighted (successively by the optimum input-to-hidden layer and hidden-to-output layer weighting coefficients), summed and aggregated at each node of the hidden and output layers. In general, the performance of the proposed system 2 (GF + PCA + ANN) for the optimal (best) parameter values is summarized in Table 5. System 1 yielded a rate of 76.6%.

Table 5: Recognition Performance of the Proposed Method on ORL Face Database

Person No.	No of Test + Validation Images	Recognized Images	Rejected Images	Recognition Rate
1	4	2	2	50%
2	4	4	0	100%
3	4	4	0	100%
:				
:				
38	4	3	1	75%
39	4	4	0	100%
40	4	4	0	100%
		Average Recognition Rate		88.33 %

Effect of increased number of epochs

The amount of training also influences the performance of the network. The default number of epoch set initially is 2000. The following table (Table 6) shows the effect of more training time/epochs of system 2. Though trained for more time, it did not show much improvement in the recognition rate.

Table 6: Far and FRR Values at Different Threshold Values

Number of Epochs	Elapsed Time(Hr)	Recognition Rate (%)
2000	0.75	88.33
5000	1.805833333	83.33333333
7000	3.6475	83.33333333
10000	5.875277778	85

False Rejection Rate and False Acceptance Rate

It is necessary to measure the accuracy of the proposed system in order to evaluate the FRS. When performing recognition for verification or authentication purposes, one typically attempts to obtain a score above a fixed threshold. If the score is above or equal to the threshold, the person is recognized, otherwise the person is not recognized.

Accuracy for verification applications is often characterized in terms of two probabilities at a given threshold:

1. False Acceptance Rate (FAR): The chance that an imposter will be erroneously recognized (obtain a matching score equal to or higher than the threshold).
2. False Rejection Rate (FRR): The chance that an authorized person will not obtain a score equal to or above the threshold.

Figure 7 Images from FERET database to measure the FRR of the proposed system

To determine the accuracy of the proposed system 2, new images shown in fig. 7 have been used. These images are from FERET face database. The results (Table 7) show the potentiality of the proposed system and it has rejected all the new images and FAR is 0% up to the thresholding limit of 0.5. Also, the table shows FRR corresponding to the test images (120 test images) of the ORL database.

So, the best values for FRR and FAR of the proposed system are 11.7% and 0% respectively. The experimental threshold value is about 0.5 in order to achieve good rejection and acceptance rate. The corresponding graph (FAR Vs FRR) is shown in Fig.8.

Table 7: FAR and FRR Values at Different Threshold Values

Threshold	FRR(%)	FAR(%)
0	100	0
0.1	20	0
0.2	16.66667	0
0.3	13.33333	0
0.4	12.5	0
0.5	**11.66667**	**0**
0.6	12.5	7.692308
0.7	15.83333	38.46154
0.8	15.83333	53.84615
0.9	19.16667	61.53846

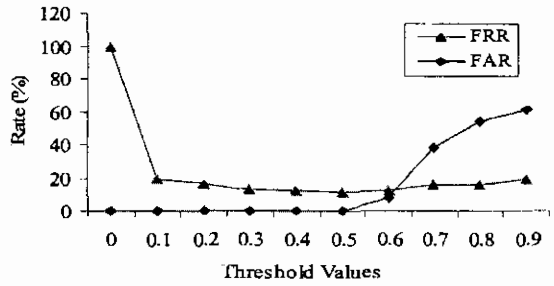

Figure 8 FAR Vs FRR

CONCLUSIONS AND RECOMMENDATIONS

Experiments have been carried out on FRS by using ORL face database, the images of which vary in illumination, expression, pose and scale. The result shows the feasibility of the methodologies followed in this work. System 1 achieved a recognition rate of 76.6% whereas model 2 FRS achieved 88.3% of correct classification and performed very efficiently when subjected to new unseen images with a false rejection rate of 0% during testing. The high recognition rate of model 2 shows the efficiency of GF in feature extraction.

There are much space for the improvements of the current system:

- Adaptation of the learning rate parameter during training may yield better classification results.
- Can investigate the effect of using more than one hidden layer.
- The algorithm can be improved in order to recognize more complicated images; colored or with different backgrounds.
- Can make it on-line recognition system by combining the software with hardware devices, like cameras.
- In order to further speed up the algorithm, number of GFs could be decreased with an acceptable level of performance reduction in recognition.

REFERENCES

[1] F. Galton, "Personal identification and description 1," *Nature*, 21 June1888, pp.173-177.

[2] F. Galton, "Personal identification and description-II," *Nature*, 28 June 1888, pp. 201-203.

[3] Haykin S., "Neural Network - A comprehensive foundation," Macmillan college publishing company, Newyork, 1994.

[4] Hyun Jin Park and Hyun Seung Yang, "Invariant object detection based on evidence accumulation and Gabor features," Pattern Recognition Letters, 22:869–882, 2001.

[5] Ilker Atalay, "FR Using Eigenfaces," *Istanbul Technical University*, January 1996.

[6] J. Huang, "Detection Strategies For FR Using Learning and Evolution," *PhD. Thesis*, George Mason University, May 1998.

[7] Johan Blommé, "Evaluation of biometric security systems against artificial fingers," *Institutionen för Systemteknik*, 2003.

[8] Joni-Kristian Kämäräinen, "Feature extraction using gabor filters, Ph.D Thesis, Lappeenranta University of Technology, Finland, 2003.

[9] Linlin Shenm, Li Bai, Bardsley Daniel, Yangsheng Wang, "Gabor Feature Selection for Face Recognition Using Improved AdaBoost Learning," Lecture Notes in Computer Science, Springer Berlin / Heidelberg, 2005.

[10] M. Turk and A. P. Pentland, "FR using eigenfaces," *IEEE Conf. Computer Vision and Pattern Recognition*, 1991.

[11] P. J. B. Hancock, V. Bruce, and A. M. Burton, "A comparison of two computer-based face recognition systems with human perceptions of faces," *Vision Research*, 1997.

[12] R. Chellappa, C. Wilson, and S. Sirobey, "Human and machine recognition of faces: A survey," *Proceedings of IEEE*, vol. 83, May 1995.

[13] Ralph Gross Jianbo Shi Jeff Cohn, Quo vadis, "Face Recognition?," *Robotics Institute, Carnegie Mellon University*, 2001.

[14] S.J. McKenna, S. Gong, and J.J. Collins, "Face tracking and pose representation," In R. B. Fisher and E. Trucco, editors, *Proceedings of British Machine Vision Conference*, Edinburgh, BMVA Press, 1996, pp. 755-764.

[15] V. Bruce, "Identification of Human Faces," Image Processing and its Applications, *Conference Publication No. 465, IEEE*, 1999, pp. 615-619.

[16] Zhao, W. Chellappa, R. Phillips, P. J. Rosenfeld, A., "Face recognition: A literature survey," ACM Computing Surverys (CSUR), Vol. 35, Issue 4, 2003, pp. 399-458.