Proceedings of the 1st ICEECE & AMF, Vol. 1, No 1, 2021.

Development of a Face Recognition System Using Hybrid Genetic-Principal Component Analysis

Akinola Ayobami Ibikunle
Department of Computer Engineering Technology
Osun State College of Technology
Esa Oke, Nigeria
olaoluayobami@gmail.com

Israel Bukola Olusola
Department of Computer Engineering Technology
Osun State College of Technology
Esa Oke, Nigeria
ibolusola@gmail.com

***Abstract*— Humans have been using physical attributes such as face, voice gait and fingerprints to recognize each other for ages. With the recent technological advancement, face recognition is a branch of biometrics system which has received considerable interest because of its ease in collecting, analysing and recognising face images. It is a system which compares an unknown image against the trained images in a database in order to identify the image. It has a number of applications such as Automatic Teller Machine (ATM), credit card, physical access control, National Identity card and correctional facilities. It has been found to be one of the ways of controlling and reducing crime rate. The development and evaluation of the performance of a face recognition system using hybrid Genetic- principal component Analysis technique is presented. The system consists of three major subsystems. Initial preprocessing procedures are applied on the input face images selected from the ORL Database. Consequently, face features are extracted from the processed images by principal component analysis and finally face identification is carried out using Genetic algorithm. Image resolutions of 50 x 50, 70 x 70, 100 x 100 and 140 x 140 are used in training and testing the system. The identification rates obtained were 100%, 96.36%, 93.63% and 90.90% for 50 x 50, 70 x 70, 100 x 100 and 140 x 140 respectively. This experimental result revealed that the lower the resolution of the cropped images, the higher the number of the correctly identified face images. The reason is attributed to the fact that there is variation in the features considered for recognition for each resolution. Hence, this technique has been proved to be more robust and suitable for low resolution.**

***Keywords*—** Face recognition; Genetic Algorithm; Principal component analysis; Identification rate and ATM

I. INTRODUCTION

Face recognition is an aspect of pattern recognition which describes the classification of a face as either known or unknown by comparing a face with the images stored in the database [1]. It is desirable to have a system that has the capability of learning to recognize unknown faces. It is very easy for people to recognize and distinguish faces but computer finds it so difficult because of the processes involved. The main goal of face recognition is to match a given face image against the ones stored in the database. This is done by comparing the invariant features such as eye centers, nose, and width of eyebrows, obtained from the techniques, which captures the representative variability of the faces [2]. This technology has received considerable interest as widely accepted biometrics because of the ease in collecting face images of persons.

The major benefits of the technique over other biometrics approaches is that face images can be taken from a distance without the foreknowledge of the concerned individual as it might be required in the identification of criminals. It has become one of the major research areas in the applications such as public security, law enforcement and commerce, credit card verification, criminal identification, access control, digital library and information security. This technique has been faced with major challenges such as presence of structural components, poses, facial expression and emotion, occlusion and image orientation among others. There are a number of face recognition algorithms such as principal component analysis (PCA), Multi-linear principal component analysis (MPCA) and Linear Discriminate Analysis (LDA) [3].

II. LITERATURE REVIEW

Moussa *et al*. proposed a face recognition system based on an approach, which combined Genetic Algorithm and Discrete Cosine Transform – Principal Component Analysis (DCT-PCA) for feature selection and dimensionality reduction [4]. The approach allowed good discrimination between chromosomes and defined the genetic operators used in the classification. The results showed that the approach has the ability to select a small number of features while maintaining very satisfactory classification rates.

Furthermore, Rikhtegar *et al.* proposed a face recognition system, which solved the problem of variations in pose, illumination, and facial expression using convolutional neural network (CNN) and support vector machine (SVM) [5]. A genetic algorithm was employed to find the optimum structure of CNN. SVMs were used to replace the last layer of CNN in order to improve the performance of the system. Moreover, an array of SVMs was constructed based on error correction output code to gain an optimum balance between resiliency and complexity of the system.

Moustafa *et al.* proposed a deep learning method for Age-Invariant Face Recognition (AIFR) [6]. The proposed method extracted deep learning features using transfer deep learning, extracted from the unprocessed face images. The extracted face features were optimized using a Genetic Algorithm (GA) procedure designed to select the most relevant features to the problem of identifying a person based on his/her facial images over different ages. K-Nearest Neighbor (KNN) classifiers with different distance metrics were investigated to classify the facial images. The results showed a Manhattan distance KNN classifier achieve the best Rank-1 recognition rates of 86.2% and 96% on the standard Face and Gesture Recognition Research Network (FG-NET) and MORPH datasets respectively.

Dembani *et al.* proposed the detection of user facial expression using a binary format of edge feature in the genetic algorithm. In this work, fitness value of the chromosome was estimated and two-phase population updating increases the cluster selection accuracy [7]. Results showed that the proposed genetic algorithm-based clustering technique increased the precision value by 35.36 % as compared to the previous approach as well as recall and the accuracy of facial expression identification without any training. In this paper, a new approach to face recognition is proposed which evaluate the performance of face recognition system using a hybrid Genetic-Principal Component Analysis technique.

There are two concepts involved in the development of this face recognition system.

A. Principal Component Analysis (PCA)

This algorithm is applied to check for a face at each pixel in the input image. It reduces the dimension of image vectors in the database. This is realized by using cross-correlation in the frequency domain between the entire input images and Eigenvectors. PCA calculates Eigen values by using covariance analysis. The Eigenvector associated with the largest Eigen value is one that reflects the greatest variance in the image and defines the subspace. Hence the Eigenvector is the subspace dimensionality. Eigenvectors and Eigen features are identified using PCA.

B. Genetic Algorithm (GA)

These algorithms operate on a number of population consisting of some encoding of the parameter set simultaneously. In this study, genetic algorithm was used in the matching stage of the system as shown in Fig. 1. Images were projected into the subspace. Subsequently, the differences between the Euclidean distance of the training images and those generated by the function of the genetic algorithm were determined.

An N-population of 8-bits was generated where N = 456 corresponding to the Euclidean distance generated for the input image. Binary encoder was used to encode the chromosome and genetic algorithm functions were applied to determine the identification of the images.

III. METHODOLOGY

The face recognition system is a kind of system whose input can either be an image or video stream. The output of the system is a verification or identification of the subject or subjects that appear in the image. The system under study has three sub-units which take care of a three-step process of facial biometrics system. The system comprises of the image acquisition unit, features extraction unit and facial identification/recognition unit.

A. Image Acquisition

In this phase, input face images were selected from the Our Laboratory Database of Faces (ORL). The images were taken with respect to different expressions and gestures such as sadness, normal, disguise, happiness etc., along with certain face pose and lighting condition. A database was prepared for training and testing.

B. Feature Extraction

Feature extraction involves obtaining relevant facial features from human faces acquired using a particular approach. These features could be certain facial regions, variations, angles etc. Dimensionality reduction and feature selection were carried out under feature extraction. Principal component analysis (PCA) is the approach employed for the following reasons: to reduce dimension of the data to more tractable limit; to capture salient class features of the data and to eliminate redundancy

C. Facial Identitification/Recognition

This was the final stage which deals with the actual face recognition. It involves the system reporting an identity from the database. It comprised of a comparison method, a classification algorithm and an accuracy measure. Genetic algorithm was used as the matching technique. A certain threshold value was chosen to make the comparison. Hence, the system was used to classify the face introduced as either known or unknown.

The face recognition system was implemented using MATLAB. This study was subdivided into two basic processes: Learning process and Identification process. The learning process algorithm is as stated below:

Start

Acquire a set of face images (training image database)
Normalize the image (resizing and background)
Gray scale conversion (black and white color)
Conversion to vector form
Apply PCA for dimensionality reduction and feature selection of all training face images in the created database.
Form the Eigen faces also known as face space
Project the centered images into face space.
Store the projected images in the template called training template.

End.

The algorithm for identification phase is as stated below:

Start

Input an unknown image
Convert image to RGB and resize to be in the same dimension with images in the training database
Convert to vector form.
Centered the testing image
Project the image into face image
Calculate the Euclidean distance – the differences between the projection of all centered training images and the projected testing images.
Match the generated test template with trained template in the database using genetic algorithms
Classify the image into either known or unknown using threshold value.

End.

IV. RESULTS AND DISCUSSION

The training and testing set of face images are selected from the Our Laboratory Database of Faces (ORL) in order to reduce variation in poses. This database consists of forty (40) subjects. Each subject is an individual with ten (10) different facial expressions. The face recognition system is

tested four (4) times using separate databases created for each image resolution as each image is cropped and resized to different resolutions such as 50 x 50, 70 x 70, 100 x 100 and 140 x 140. Hence a database allotted to each resolution. The sample of the image is shown in Fig. 2. The training dataset consists of 290 images and the testing dataset consists of 110 images.

Table 1 shows the average training time and recognition time at different resolutions as well as the identification rates, number of testing images and the number of unidentified images. The time involved in training each image varied and depended largely on the resolution of the images used for training the database.

The recognition accuracy of the system varies along with the image resolutions. Table 2 shows that as the number of faces images increases the duration also increases. Hence, the relationship is linear. Figs. 3, 4, and 5 show graphs illustrating the relation of training time at different image resolutions; recognition time at different image resolutions and recognition accuracy at different image resolutions respectively.

The results of the proposed facial recognition system using hybrid Genetic-Principal Component Analysis (GPCA) show that the lower the facial features that are present in the training and testing images, the better the performance. The recognition performance of the system can be attributed to the fact that the more the features available for the processing. The higher will be the variation on the parameters used for selection process. It is discovered that as the image resolution increases the Euclidean distance used as input into the genetic algorithm also increases. Hence, different threshold have to be set to get a better performance of the system at each image resolution.

Besides, the performance of the system can also be tested to false rejection rate (FRR). FRR is the probability that the system will fail to identify an enrollee or verify the legitimate claimed identity of an enrollee. For example, using 70 x 70 image resolution, the recognition accuracy of the system is 96.36%, the FRR is 0.036 which shows that out of the 110 testing images, four (4) is rejected. This is attributed to the fact that as the resolution of the cropped images increases there is higher variation in the features to be considered for recognition.

The performance of the system is also evaluated using a retrieval process classified into four terms: True positive (Tp); True negative (Tn); False positive (Fp) and False negative (Fn) [8]. The retrieval process is classified under True positive if an image is part of the database and correctly retrieved by the system. If an image is part of the database but a different image is wrongly retrieved from the database by the system, it is then classified under false negative.

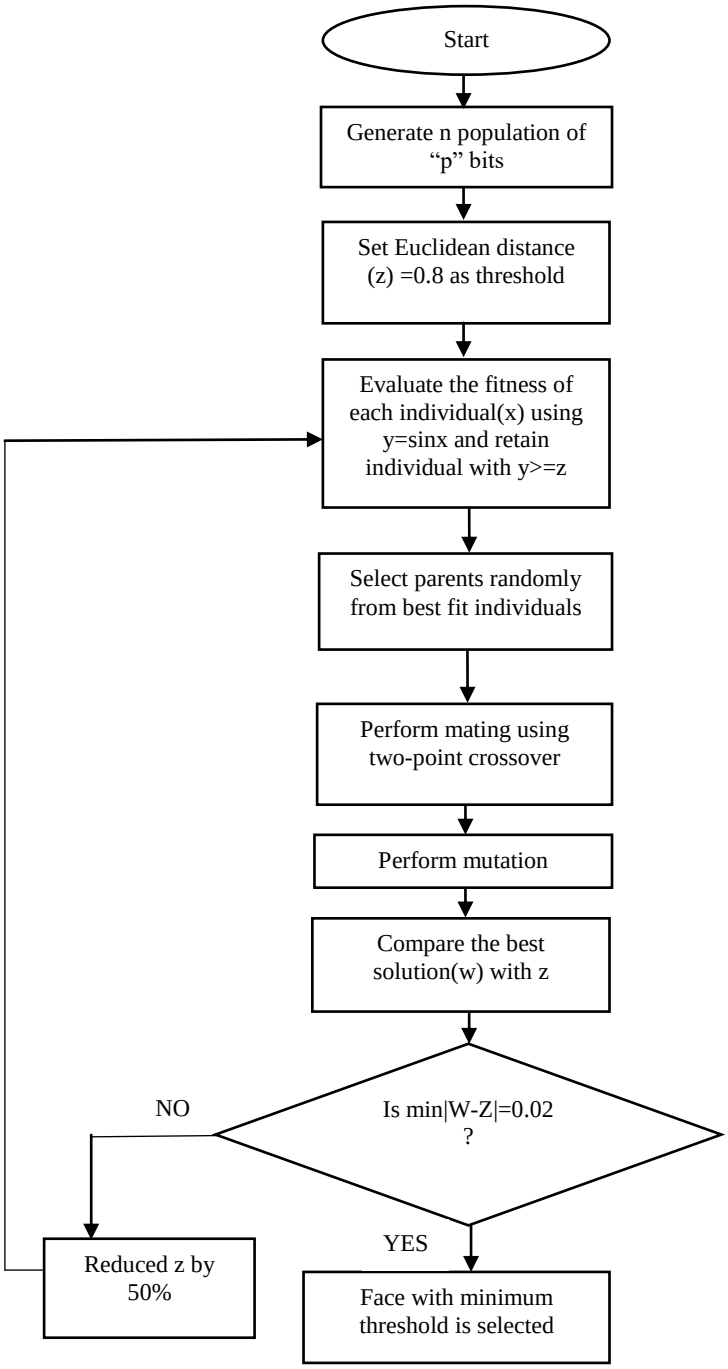

Fig. 1. Flow chart of matching stage using genetic algorithm

However, if an image is not part of the database but a different image is wrongly retrieved from the database by the system, the retrieval process is classified to be false positive [8]. If an image is not part of the database and correctly identified by the system as not belonging to the database, it is then classified as True negative. From Table 3, for image resolution 50 x 50, True positive results were recorded 110 times, no False positive, no True negative and no False negative, for image resolution 70 x 70, True positive results were recorded 106 times, no false positive, no True negative and False negative results were 4 times and soon.

Table 1: Average training time and recognition time at different image resolutions

	Total number of testing images	Total number of correctly identified images	Correct identification rate (%)	Training time(secs)	Recognition time(secs)
50x50	110	110	100.00	17.10	6.17
70x70	110	106	96.36	27.05	8.61
100x100	110	103	93.63	48.71	16.66
140x140	110	100	90.90	78.74	26.35

Fig. 2. Sample of the ORL Database for training and testing datasets

Table 2. The relationship between the number of images in the training databases and training time

Number of images in the database	Training time (secs)
50	0.90
100	2.12
150	3.40
200	4.80
300	8.80
350	11.32
400	15.35
456	20.26

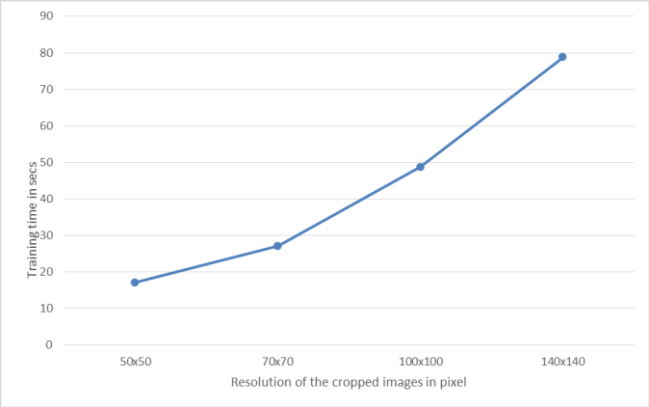

Fig. 3. Graph of training time at different image resolutions

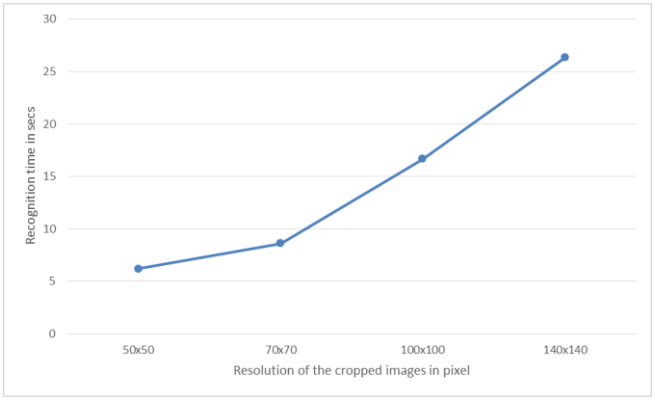

Fig. 4. Graph of recognition time at different image resolutions

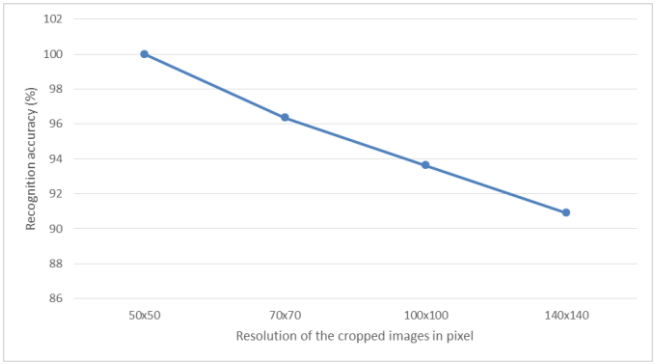

Fig. 5. Graph of recognition accuracy at different image resolutions

Table 3. True positive value, False positive value, True negative value and False negative value for each image resolution.

Image resolution	True positive	False positive	True negative	False negative
50 x 50	110	0	0	0
70 x 70	106	0	0	4
100 x 100	103	0	0	7
140 x 140	100	0	0	10

V. CONCLUSION

A prototype of a real time face recognition system using hybrid Genetic-Principal Component Analysis (GPCA) has been successful used to extract and recognize face images. This technique has proved to be more robust suitable for low resolution, variable lightning and different facial expressions.

The static frontal images used were selected from the ORL Database. The dimensionality reduction was done by the use of principal component analysis implemented in Matlab environment. An input image was introduced into the system to test the accuracy of the system. Euclidean distance was used as distance measure for each input image several Euclidean distance were got corresponding to the number of images in the training database. This was used as input into the Genetic algorithm.

The highest performance of the system was recorded for the lowest image resolutions. This can be attributed to the features available for the processing and thereby made the parameters available for the Genetic algorithm.

REFERENCES

[1] S. Fernandes and J. Bala, (2013). "Performance Analysis of PCA-based and LDA-Based Algorithms For Face recognition," *International Journal of Signal Processing Systems,* Vol. 1, No. 1, pp. 1–6

[2] S. Singh, M. Sharma, and D. N. S. Rao, (2012): "Accurate Face Recognition using PCA and LDA*,"International Journal of Management, IT and Engineering Vol. 2 No. 4 pp. 101-121*

[3] D. Yi, Z. Lei, and S. Z. Li, (2013) "Towards Pose Robust Face Recognition". *In Proceedings of the IEEE conference on Computer Vision and Pattern Recognition (CVPR).*pp. 3539–3545

[4] M. Moussa, M. Hmila, and A. Douik. (2018). "A Novel Face Recognition Approach Based on Genetic Algorithm Optimization," *Studies in Informatics and Control, Vol. 27 No. 1, pp. 127-134*

[5] A. Rikhtegar, M. Pooyan, and M. T. Manzuri-Shalmani. (2016). "Genetic algorithm-optimised structure of convolutional neural network for face recognition applications*".IET Computer Vision, Vol. 10, No. 6, pp. 559-566*

[6] A. A. Moustafa, A. Elnakib, and N. F. F. Areed. (2020)."Optimization of deep learning features for age-invariant face recognition". *International Journal of Electrical and Computer Engineering (IJECE). Vol. 10, No. 2, pp. 1833-1841*

[7] R. Dembani, W. Zheng, and M. Sun. (2020)."Unsupervised facial expression detection using genetic algorithm". *Neural Network World, Vol. 30, No. 1, pp. 65-75*

[8] A. Zubair and O. A. Alo (2020). Content-based Image Retrieval System using Second-Order Statistics. *International Journal of Computer Applications. Vol. 176, No. 36, pp. 1-9*

[9] The ORL database, Available: http://www.cam-orl.co.uk/face