International Journal of Computer Information Systems and Industrial Management Applications.
ISSN 2150-7988 Volume 6 (2014) pp. 150-159

PERFORMANCE ANALYSIS OF FACE RECOGNITION SYSTEMS BASED ON FALSE REJECTION OR FALSE ACCEPTANCE OF PROBE IMAGE

Adebayo, Kolawole John[1], Onifade, Olufade F. Williams[2] and Yisa, Fatai Idowu[3]

[1,3] Computer Science Department, Oduduwa University,
Km 25, Ife roundabout, Ipetumodu, Ife North LG, Osun State, Nigeria
Collawolley3@yahoo.com [1], al_fattah2003@yahoo.com [3]

[2] Computer Science Department, University of Ibadan,
Ibadan, Oyo State, Nigeria
fadowilly@yahoo.com

***Abstract*: Face recognition like other biometrics systems involves some basic processes, which includes biometric feature acquisition / enrollment which in this case would be faces of human to be recognized, normalization of these enrolled features in order to standardize the training set and lastly is the recognition which involves mapping the enrolled features collected to features of people to be recognized i.e the probe images. Several comparisons have been made on some face recognition systems with variations in each of the result, even when the same algorithms is used in those experiments. This variation has in no small measure rubbish the authenticity of these algorithms leading to the common problem of either false acceptance or false rejection on the target object. This paper presents a comparative analysis of the performance of some selected face recognition systems, namely the PCA, 2DPCA and ICA. The algorithms were implemented and tested exhaustively to evaluate the performance of these algorithms under different face databases and similarity distance metrics in respect to the recognition accuracy. We statistically present the results obtained.**

Keywords: Principal Component Analysis, 2 dimensional PCA, Biometric, ICA, False Acceptance Rate, False Rejection Rate, Face recognition**.**

I. Introduction

Face recognition as a field of computer vision and pattern recognition has gained much attention in the industry and academia in the recent times; this is unsurprising because it possesses the merits of both high accuracy and low intrusiveness. One common application is to identify or verify the person of a given face in still or video images [1].

Face recognition problem is typically a problem of simple comparison in which facial image of a subject (probe) is compared to images of several subjects stored in a database (training images), in this scenario, a match means the subject of the probe image belongs to the set of images in the training database. This comparison is done in either of two ways i.e. one-to-one mappings or one-to-many mapping.

Face recognition like other biometrics systems involves some basic processes, these includes biometric feature acquisition / enrollment which in this case would be faces of human to be recognized, normalization of these enrolled features in order to standardize the training set and lastly is the recognition which involves mapping the enrolled features collected to features of people to be recognized. In this work, this test sets are referred to as the probe images. Recognition of images in itself follows some steps, this include segmentation of faces from cluttered scenes (especially if the image consists of several subjects, extraction of important features from the face region and finally, decision making on whether a person is who he claims to be or not. Segmentation is done by creating an edge map whose edges are connected together using some heuristics. The edges are then map into an elliptical shape using for example Hough transform. For the extraction of desired features, any of the algorithms based on the two common modalities used in face recognition is used i.e. Holistic e.g. PCA and feature based approach. In holistic feature, each feature is a characteristic of the whole face while partial features as considered in the feature based approach includes measurements of important key points like the nose, mouth, eyes etc. and the distances between these key points in a face.

Some sophisticated commercial systems have been developed over years which have achieved appreciable level of success, apparently, most of them are based on subspace projection in which data from high-dimensional space are reduced to a low-dimensional space and a distance metric for classification (such as nearest neighbor rule, mahalanobis etc) is then used in the low dimension space. While many of these commercial systems' algorithms have been shrouded in secrecy (including the several pre-processing and post-processing steps adopted in order to enhance recognition accuracy and improve system performance), few non-proprietary algorithms have been implemented and discussed widely in the academia, some given excellent results while being constantly re-modeled to enhance their performance. Some of these face recognition algorithms proposed by researchers includes the Principal Component Analysis [2], Local Feature Analysis, Linear Discriminant Analysis and Fisher face which are all based on dimensionality reduction [1]. Also neural networks [6], elastic bunch graph theory, 3D morphable models [5] and multi-resolution analysis are some other techniques usually used to mention a few. Each of the proposed face recognition algorithms has typically overcome the shortcomings of one another, thereby extending the application areas for face recognition systems. Important applications of face recognition are seen in biometrics i.e. computer security and human computer interaction. Most especially, it has been the most successful application employed in surveillance systems. Biometrics is an automated method of identity verification or identification based on the principle of measurable physiological or behavioral characteristics such as finger-print, iris pattern, facial characteristics or a voice sample [7]. Several state-of-the-art biometric techniques have been developed over the years which use a variety of human characteristics for identification and recognition. These include fingerprint, signature, iris, retina, hand, voice and facial recognition [7] [8]. Each biometric trait has its strengths and weaknesses, and the choice of a specific trait depends upon the requirements of the application. Among them all, face recognition is frequently used to discriminate authorized and unauthorized persons, as they are least intrusive with high public acceptability. One common application area that shows the unmatched ability of face recognition and its superiority to other biometric approaches is the surveillance system. Here, because people must be monitored without them being aware they are being monitored, a system that is totally non-intrusive must be put in place. Other biometric systems may need user participation and cooperation and as such are ineligible for such sensitive applications. This has especially prompted researches in face recognition systems.

This paper explores the performance of some selected face recognition algorithms. In particular, Principal Component Analysis (PCA) [2], a variation of PCA, termed 2DPCA [15] and Independent Component Analysis were used and analyzed in the experiments. Principal component analysis, also known as Karhunen Loeve expansion, is a classical feature extraction and data representation technique widely used in the areas of pattern recognition and computer vision. Sirovich and Kirby [2] first used PCA to efficiently represent pictures of human faces. They opined that any face image could be reconstructed approximately as a weighted sum of a small collection of images that define a facial basis (eigen images), and a mean image of the face. This influenced the work of Turk and Pentland [3] where the groundbreaking Eigenfaces method for face recognition was proposed. PCA has stayed to be one of the foremost implemented face recognition algorithms and has been widely investigated in the academia. PCA is widely used in face recognition systems since though less computationally intensive, gives good recognition accuracy. Also, 2DPCA presents a more efficient approach to dimensionality reduction as compared to ordinary PCA; this improves its efficiency over PCA. It is a straightforward image projection technique developed for image feature extraction. As opposed to conventional PCA, 2DPCA is based on 2D matrices rather than 1D vector. That is, the image matrix does not need to be previously transformed into a vector. Instead, an image covariance matrix can be constructed directly using the original image matrices. In contrast to the covariance matrix of PCA, the size of the image covariance matrix using 2DPCA is much smaller. As a result, 2DPCA has two important advantages over PCA. First, it is easier to evaluate the covariance matrix accurately. Second, less time is required to determine the corresponding eigenvectors [4]. ICA is one algorithm that has been explored widely in the literatures. While PCA decorrelates the input data using second-order statistics and thereby generates compressed data with minimum mean-squared re-projection error, ICA minimizes both second-order and higher-order dependencies in the input. ICA attempts to find the basis along which the data (when projected onto them) are statistically independent. It has been shown to perform better than the PCA in some literatures. These three algorithms are holistic based / appearance based approach. The block system diagram of an appearance based system that our experiment was patterned after for all the algorithms used can be seen in figures I and II of the appendix. Figure 1 gives an illustration of general subspace appearance-based face recognition system while figure II gives the matching face of such a system The system is quite simple, input a probe image and project the image into low dimensional space as shown in the Figures, calculate the distance between the projected image (probe) and the images in the database and pick the one with lowest distance as the recognized image. We have used three different popular distance metrics in our experiments which gives varying results as will be seen later.

The comparison made in this work is based on the fact that very few works have been done to experimentally compare some of these algorithms in terms of the false acceptance rate (FAR) and false rejection rate (FRR) based on the available literatures. The algorithms investigated in this work were chosen because they form the most widely implemented and available face recognition systems in the academia. Particularly, PCA and ICA have enjoyed continuous use and have been modified severally in the literatures for performance gain and to model a more efficient system. We believe a good analysis can present to the general public how good and efficient the algorithms are, when deployed on different databases. The study can guide the public in choosing an optimal algorithm.

The rest of the paper is divided into four sections. In the next section, we reviewed the related works, this is followed by the methodology employed, where we give a brief summary of the algorithms; next is the section describing the experimental setup and results. Finally, we present the interpretation of the result and the conclusion.

II. Related Works

We now give a brief survey of some work done on face recognition across several algorithms.

In ref [1], the authors gave a comprehensive survey of several face recognition techniques which include detailed description and classification of the algorithms both for still and video-based recognition and should be consulted for further review.

Reference [3] proposed a method using PCA which detects the head of an individual in a complex background and then recognizes the person by comparing the characteristics of the face to those of known individuals. In reference [12], the use of PCA and Gabor Filters was suggested. Firstly, Gabor Filters, Log Gabor filters and Discrete wavelet transform were used to extract facial features from the original image on predefined fiducial points. PCA was then used to classify the facial features optimally and reduce the dimension. The approximation coefficients in discrete wavelet transform was extracted and was then used to compute the face recognition accuracy instead of using all the coefficients. They suggested the use of combining these methods in order to overcome the shortcomings of PCA.

Reference [13] used supervised and unsupervised learning algorithms in their system. The supervised learning has been carried out with the using a bi-layered artificial neural network having one input, two hidden and one output layer. The gradient descent with momentum and adaptive learning rate back propagation learning algorithm was used to implement the supervised learning in a way that both the inputs and corresponding outputs are provided at the time of training the network, this gave an inherent clustering and optimized learning of weights which allowed an efficient results. The unsupervised learning was implemented with the help of a modified Counter propagation network. The work done in [14] builds multiple eigenspaces in terms of illumination directions and train illumination direction-specific neural networks on the feature coefficients projected in the corresponding eigenspaces. All illumination direction-specific neural networks are then combined by a neural network ensemble module. In the test phase, using an input image with an arbitrary illumination direction, their proposed ensemble architecture could complete recognition in a uniform way where they feed the input image into different channels corresponding to different illumination directions and obtaining a final decision from the ensemble module. They claimed a better result than the conventional approach.

The work done in [15] based on 2DPCA, uses 2D features obtained directly from original vector space of a face image rather than from a vectorized one dimensional (1D) space. Also, in [4], Jian Yang et al two dimensional PCA as a new approach to appearance based face representation and recognition system, they claimed their system was suitable for image feature extraction. They claimed the developed system is more computationally more efficient than the PCA. They gave some reasons why they believed the introduced approach i.e. 2DPCA is more suitable than the PCA. First, they claimed that 2DPCA is more suitable for small sample size problems (like face recognition) since its image covariance matrix is quite small. Since image representation and recognition based on PCA (or 2DPCA) is statistically dependent on the evaluation of the covariance matrix (although for PCA the explicit construction of the covariance matrix can be avoided), the advantage of 2DPCA over PCA is that the 2DPCA evaluates the covariance matrix more accurately. Bruce draper et al in [16] compared ICA and PCA in the context of a baseline face recognition system. Their paper shows how the relative performance of

PCA and ICA depends on the task statement, the ICA architecture, the ICA algorithm,and (for PCA) the subspace distance metric. They then explored the space of PCA/ICA comparisons by systematically testing two ICA algorithms and two ICA architectures against PCA with four different distance measures on two tasks (facial identity and facial expression). They showed that the FastICA algorithm configured according to ICA architecture II yields the highest performance for identifying faces, while the InfoMax algorithm configured according to ICA architecture II is better for recognizing facial actions. In all their experiments, PCA proved to perform well but not as well as ICA. However, they did not test the effects of registration errors or image pre-processing schemes on recognition accuracy in the comparison. In [17], Delac et al also worked on comparing three face recognition algorithms namely PCA, LDA and ICA using the FERET face database. Their work presented an independent, comparative study of these appearance-based face recognition projection methods and their accompanied four distance metrics (L1, L2, cosine, and Mahalanobis) in completely equal working conditions. Their experimental setup yielded 16 different algorithms that were compared. They gave some hypothesis as below: (1) No claim can be made about which is the best combination for the different expression task since the differences do not seem to be statistically significant (although LDA+COS seems to be promising), (2) PCA+L1 outperforms ICA1 and LDA with illumination changes task at all ranks and outperforms ICA2 from rank 27 further on, (3) COS seems to be the best choice of metric for ICA2 and gives good results for all probe sets, (4) ICA2+COS combination seems to be the best choice for temporal changes task, (5) In many cases L2 produced lower results than L1 or COS even though it is the most used, (6) L1 and COS metrics produced best overall results across all algorithms and should be further investigated. They concluded by saying when tested in completely equal working conditions, no algorithm (projection–metric combination) can be considered the best time and the choice of appropriate algorithm can only be made for a specific task.

III. Methodology

Despite the fact that few works have cross examined some face recognition algorithms, especially some holistic based algorithms as we opined to discuss in this work, very few of the works have agreed on recognition achieved by these algorithms. For instance, while some works claimed to have the best result from PCA, some have argued that ICA outperforms it in their experiments; others have claimed that LDA also outperforms PCA which was disproved in other works as having the least accuracy. Some works as in [17] extended their approach to include varying distance metrics for subspace projection which thy believed also affect the result obtained in earlier works since they mostly didn’t conduct their experiment with respect to the distance metrics available i.e. we cannot totally say an algorithm significantly outperforms another while different distance metrics have been used. Some other issues that have affected the difference in the result obtained has been the issue of pre-processing of images and the number of images trained per class. For example, while one work might have used several pre-processing algorithms to enhance performance of recognition, another might have simply used only histogram equalization without bothering about using all the other pre-processing steps used in the former work, obviously, the result that will be presented from the two works would differ even if they had used the same face database, algorithm and distance metrics for their comparison. In fact, it has been shown that ICA architecture I gives a different result to ICA architecture II [16]. Lastly, a baseline comparison should employ the same face database, for instance, if author A used FERET database for his work while author B used ORL face database for his work, their results would apparently be different slightly or could even give a significant statistical difference.

Based on these analysis and shortfalls causing the disparity apparent in the results of the earlier works, we are motivated to conduct our own original research on comparing and analyzing three holistic based algorithms with each algorithm given the same environment to thrive, for example, we have elected to use only histogram equalization [18] as a simple illumination normalization step across all database for our experiments. Also we are employing both the FERET along with its standard tests (gallery and probe sets) and ORL face database for our analysis. These two databases have been mostly used in earlier works comparing face recognition systems. While FERET database gives the appeal of a large dataset with many classes and fewer images (both gallery and probe) per class, the ORL face database on the other have a small dataset with fewer number of class and more images per class. This is expedient for the experiments as we can now analyze these two databases and get baseline accuracy for the algorithms used. Also, to make our work simple but realistic, we have considered only three distance metrics, that is, the mahalanobis, Euclidean and the cosine distance metrics. The three will be discussed later.

Below, we briefly discuss the three algorithms used for our analysis.

ALGORITHMS DESCRIPTION

A. *Principal Component Analysis (PCA)*

A method of extracting features in a holistic system is by applying statistical methods such as Principal Component Analysis (PCA) to the whole image. PCA can also be applied to a face image locally; in that case, the approach is not holistic. Irrespective of the methods being used, the main idea is the dimensionality reduction based on extracting the desired number of principal components of the multi-dimensional data. The goal is to extract the relevant information of a face and also capture the variation in a collection of face images and encode it efficiently in order for us to be able to compare it with other similarly encoded faces.

A method usually used is the Eigenface Method by Turk and Pentland [3] which is based on the Karhunen-Loeve expansion. The work in [3] is motivated by the ground breaking work of Sirovich and Kirby in [2] and is based on the application of Principal Component Analysis to human faces.

Principal component analysis provides a method to efficiently represent a collection of sample points, reducing the dimensionality of the description by projecting the points onto the principal axes, an orthonormal set of axes pointing in the directions of maximum covariance in the data. It minimizes the mean square error for a given number of dimensions and provides a measure of importance for each axis. The algorithm is as follows:

1. Let's assume the face images in our database is x_1, x_2, x_3, x_4 ……., x_M then find the mean image which is
   $$\Psi = \frac{1}{m}\sum_{n=1}^{m} X_n \qquad (1)$$
2. Next, we have to know how each face differs from the mean image above like this $\Phi_i = X_i - \Psi$ (2)

   This set of very large vectors is then subject to principal component analysis, which seeks a set of M orthogonal vectors, U_n, which best describes the distribution of the data. The k^th vector, U_K, is chosen such that the eigenvalues $\lambda_k = \sum_{n=1}^{m}(U_K^T\ \Phi_n)^2$ (3)
which is also subject to eigenvector $U_I^T U_K$, where the vectors $\mathbf{U}_K$ and scalars λ_K are the eigenvectors and eigenvalues, respectively of the covariance matrix C of the training images depicted as
$$C = \frac{1}{m}\sum_{n=1}^{m}\Phi_n\ \Phi_n^T = AA^T. \qquad (4)$$
In essence, we are calculating the covariance matrix C.

3. The matrix A= [ϕ_1, ϕ_2, ϕ_3 …… ϕ_m]. The covariance matrix C, however is N^2 x N^2 real symmetric matrix, and determining the N^2 eigenvectors and eigenvalues is an intractable task for typical image sizes. We need a computationally feasible method to find these eigenvectors.

   Following these analyses, we construct the M x M matrix $L = A^TA$ where $L_{mn} = \phi^t_m\,\phi_n$ and then find the M eigenvectors, V_i of L. These vectors determine linear combinations of the M training set of face images to form the eigenfaces U_i, which we represent as
$$U_I = \sum_{k=1}^{m} V_{IK}\ \Phi_K \qquad (5)$$
Where I = 1……M.

The associated eigenvalues allow us to rank the eigenvectors based on how useful they are in characterizing the variation among the images. This algorithm works well because of the evaluation of the eigenvalues and eigenvectors of the real symmetric matrix L that is composed from the training set of images.

B. *Two Dimensional Principal Component Analysis (2DPCA)*

2DPCA unlike the conventional PCA method uses the 2D features obtained directly from original vector space of a face image rather than from a vectorized 1D space. This novel system was first proposed in [15]. Its advantages over PCA are detailed in [4]. Also, unlike the PCA where Euclidean or Mahalanobis distance is used as the classifier, 2DPCA uses Volume measure (VM) to classify the distance of the probe from the face space and its calculated using the formula $V(A) = \sqrt{AA^T}$ where A is the matrix of full column rank and A^T is its transpose. Below, we briefly describe the process involved in classification and recognition.

Considering a training face set {X1,X2, . . . ,XN}, 2DPCA uses all training samples to build the total sample covariance matrix C as seen in the equation below.

C = E [X-E(X)) T (X-E(X))] = $\frac{1}{N}\sum_{i=1}^{N}(X_i - \Psi)^T$ (6)

Where Xi is the i^{th} training sample, which is [h x w] matrix and Ψ denotes the mean sample matrix of all training sample matrix, and N is the number of training samples. The novel idea in 2DPCA is to select some good projection vectors by using the total scatter of the projected samples, this is denoted by the trace of the covariance matrix of the projected feature vectors. The algorithm continues as follows

J (w) = tr (Sw) where Sw is the covariance matrix of the projected feature vectors of the training images, and tr(Sw) stands for the trace of Sw. In order to maximize the criteria in the equation above, Sw is equal to find a projection direction w, onto which the total scatter of the projected samples is maximized. We can represent the covariance matrix Sw by the equation 7 below

SW=E[y–E(y)]T [y-E(y)] = E[X-E(X))w]T [(X-E(X))w] (7)

Based on the first equation, we can derive J(w) as shown in equation 8 thus J(w) = $W^T C_w$ (8)

The optimal projection axes, w1,w2, . . . ,wd, are the orthonormal eigenvectors of C corresponding to the first d largest eigenvalues. It has been proven in previous works that the covariance matrix in 2DPCA can be computed more efficiently and easily than obtainable with PCA [19]. A feature matrix Yi = [yi1, yi2, … yid] for each training face sample (or each sample in gallery set) can be obtained by , $Y_{ik} = X_i, W_k$, where k = 1,2,. . . ,d. In the similar fashion, the 2DPCA model also gets a feature matrix Yt = [Y_{t1}, [, Y_{t2}, … [, Y_{td}] for each testing face sample after the transformation by 2DPCA as briefly described above. Then, a nearest neighbor classifier based on the matrix distance is used for classification.

C- arg min d(Yt ,Yi) - arg min$\sum_{k=1}^{d}$ || y tk yik ||$_2$ (9)

Where

d (Yt, Yi) = $\sum_{k=1}^{d}$ || y tk yik ||$_2$, C Σ [1,2,….,N], (10)

it should be of note that the distance between Y_c and Y_t is minimal. Also, Y_t belongs to the class where Y_c belongs too. The above described procedure is as done in [20] and employed in 2DPCA.

C. *Independent Component Analysis*

While PCA decorrelates the input data using second-order statistics and thereby generates compressed data with minimum mean-squared re-projection error, ICA minimizes both second-order and higher-order dependencies in the input. It is intimately related to the blind source separation (BSS) problem, where the goal is to decompose an observed signal into a linear combination of unknown independent signals. In this paper, we use the FastICA algorithm as against other methods such as InfoMax or Maximum likelihood that can also be employed.

The FastICA method computes independent components by maximizing non-Gaussianity of whitened data distribution using a kurtosis maximization process [48]. The kurtosis measures the non-Gaussianity and the sparseness of the face representations.

Let s be the vector of unknown source signals and x be the vector of observed mixtures. If A is the unknown mixing matrix, then the mixing model is written as

x =As (12)

It is assumed that the source signals are independent of each other and the mixing matrix A is invertible. Based on these assumptions and the observed mixtures, ICA algorithms try to find the mixing matrix A or the separating matrix W [41] such that U = Wx = WAs (13)

is an estimation of the independent source signals.

Independent Component Analysis aims to transform the data as linear combinations of statistically independent data points. Therefore, its goal is to provide an independent rather that uncorrelated image representation. ICA [42] is an alternative to PCA which provides a more powerful data representation. It's a discriminant analysis criterion, which can be used to enhance PCA. The ICA algorithm is briefly detailed below:

Let cx be the covariance matrix of an image sample X. The ICA of X factorizes the covariance matrix C_x into the following form: $C_x = F\Delta F^T$ (14)

where Δ is diagonal real positive and F transforms the original data into Z (X = FZ). The components of Z will be the most independent possible. To derive the ICA transformation F,

$X = \Phi\Lambda^{\frac{1}{2}} U$ (15)

Where X and Λ are derived solving the following Eigen problem:

$C_x = \Phi\Lambda\Phi^T$ (16)

Then, there are rotation operations which derive independent components minimizing mutual information. Finally, normalization is carried out.

We have elected to employ ICA architecture I, the comparison of architecture I and II can be found in [16]. We shall give a brief breakdown of this architecture. If X is taken to be the mixing model, then the input face images in X are considered to be a linear mixture of statistically independent basis images S combined by an unknown mixing matrix A. The ICA algorithm learns the weight matrix W, which is used to recover a set of independent basis images. In this architecture, the face images are variables and the pixel values provide observations for the variables. The source separation, therefore, is performed in face space. Projecting the input images onto the learned weight vectors produces the independent basis images. The compressed representation of a face image is a vector of coefficients used for linearly combining the independent basis images to generate the image. Bartlet et al [21] first apply PCA to project the data into a subspace of dimension m to control the number of independent components produced by ICA. The InfoMax algorithm is then applied to the eigenvectors to minimize the statistical dependence among the resulting basis images. This use of PCA as a pre-processor in a two-step process allows ICA to create subspaces of size m for any m. Liu et al in [22] opined that pre-applying PCA enhances ICA performance by discarding small trailing eigenvalues before whitening and also by reducing computational complexity by minimizing pair-wise dependencies. PCA decorrelates the input data; the remaining higher-order dependencies are separated by ICA.

Mathematical, we can describe this architecture as follows: let R be a p by m matrix containing the first m eigenvectors of a set of n face images.

Let p be the number of pixels in a training image. The rows of the input matrix to ICA are variables and the columns are observations, therefore, ICA is performed on R^T. The m independent basis images in the rows of U are computed as

$U = W * R^T$. (17)

Then, the n by m ICA coefficients matrix B for the linear combination of independent basis images in U is computed as follows:

Let C be the n by m matrix of PCA coefficients. Then,

$C = X * R$ and $X = C * R^T$ (18)

From $U = W * R^T$ and the assumption that W is invertible we get $R^T = W^{-1} * U$.

Therefore, $X = (C * W^{-1}) * U = B * U$ (19)

Each row of B contains the coefficients for linearly combining the basis images to comprise the face image in the corresponding row of X. Also, X is the reconstruction of the original data with minimum squared error as in PCA [16].

IV. Experimental Setup

A. Database Used

We adopted two of the foremost face databases i.e. the ORL and FERET databases. The ORL face database consists of 40 subjects with 10 images per subject. The images were taken under different pose, light intensities and facial expressions. The dimension of each image is 92 by 112 in which the background has been typically removed.

On the other hand, The FERET face recognition database is a set of face images collected by NIST from 1993 to 1997. The FERET database contains images of 1,196 individuals, with up to 5 different images captured for each individual. The images are separated into two sets i.e. the gallery images and the probe/test images.

Each image contains a single face. Prior to processing, the faces are registered to each other, and the backgrounds are eliminated. In this study, only head-on images are used; faces in profile or at other angles are discarded. Of particular interest is the structure of the database. The gallery contains 1,196 face images. For this study, the training images are a randomly selected subset of 500 gallery images. Most importantly, there are four different sets of probe images: using the terminology in, the fafb probe set contains 1,195 images of subjects taken at the same time as the gallery images. The only difference is that the subjects were told to assume a different facial expression than in the gallery image. The duplicate I probe set contains 722 images of subjects taken between one minute and 1,031 days after the gallery image was taken. The duplicate II probe set is a subset of the duplicate I probe set, where the probe image is taken at least 18 months after the gallery image. The duplicate II set has 234 images. Finally, the fafc probe set contains images of subjects under significantly different lighting. This is the hardest probe set, but unfortunately it contains only 194 probe images. **Gallery** images are images with known labels, while **probe** images are matched to gallery images for identification.

The database is summarized as briefed below:

FB: Two images were taken of an individual, one after the other. In one image, the individual has a neutral facial expression, while in the other they have non-neutral expressions. One of the images is placed into the gallery file while the other is used as a probe. In this category, the gallery contains 1,196 images and the probe set has 1,195 images.

Duplicate I: The only restriction of this category is that the gallery and probe images are different. The images could have been taken on the same day or a year apart. In this category, the gallery consists of the same 1,196 images just like the FB gallery while the probe set contains 722 images.

Fc: Images in the probe set are taken with a different camera and under different lighting than the images in the gallery set. The gallery contains the same 1196 images as the FB and Duplicate I galleries, while the probe set contains 194 images.

Duplicate II: Images in the probe set were taken at least 1 year after the images in the gallery. The gallery contains 864 images, while the probe set has 234 images.

For our experiment, we used only the Duplicate II gallery images.

B. Pre-processing

Histogram equalization [18] was applied on all images prior to the beginning of the experiments. All images were rescaled to 60 by 50 dimensions. Also, the background details were removed.

For the FERET face database, images were first spatially transformed (to get eyes at fixed points in imagery) based upon a ground truth file of eye coordinates supplied with the original FERET data. We have developed a program for automatic rotation and cropping of all images without placing any rectangular mask on each face for background elimination. This pre-processing was done for all the databases i.e. both the FERET and the ORL database.

C. Distance Metrics Used

The main objective of similarity measures is to define a value that allows the comparison of feature vectors (reduced vectors in eigenspace frameworks). With this measure the identification of a new feature vector will be possible by searching the most similar vector into the database. This is the well-known nearest-neighbor method. One way to define similarity is to use a measure of distance, d(x,y), in which the similarity between vectors, S(x,y) is inverse to the distance measure. We used three well known distance measurement methods for our experiments, these are the Mahalanobis, Euclidean and the Cosine metrics. They are briefly described below.

Euclidean Distance is given by:

$$d(x, y) = \sqrt{(x-y)^T (X-Y)} \qquad (20)$$

Cosine Distance is given by:

$$S(x, y) = \mathrm{Cos}(x, y) = \frac{x^T y}{\|x\| \|y\|} \qquad (21)$$

Mahalanobis Distance is given by:

$$D(x, y) = (x-y)^T R^{-1} (x-y). \qquad (22)$$

Where R denotes the correlation matrix

Looking at these distance metrics geometrically, this distance has a scaling effect in the image space. Taking into consideration the face image subset, directions in which a greater variance exists are compressed and directions in which a smaller variance exists are expanded. It can be proved that in the PCA space the Mahalanobis distance is equivalent to the Euclidean distance, weighting each component by the inverse corresponding eigenvalue, and it is often called Whitening (PCA) Transformation.

In this section, we present the experiments conducted in comparing the three analyzed algorithms. The algorithms were implemented successfully as done in the reviewed works. The face database used includes the Olliveti Research Lab (ORL) face database and the FERET face database. FERET has many class but fewer images per class while the other has fewer class with m,ore images per class (ORL).

The ORL face database consists of 10 images each of 40 distinct subjects. The images were taken at different time varying the lighting, facial expression (open/closed eyes, smiling/not smiling) and facial details (glasses / no glasses). All the images were taken against a dark homogenous background with the subjects in an upright frontal position with tolerance from some side movement. On the other hand, the FERET face database has more classes than the ORL, as a matter of fact, it has more probe set per class than the ORL as it consists of a very large dataset.

For the FERET database, we trained the PCA algorithm using a subset of classes for which there were exactly three images per class. There exist 225 such classes, however, we randomly sampled 95 which were selected and used in the experiment. giving us 285 images in all trained. 40% of these 675 images were selected as recommended by the FERET standard resulting in 114-dimensional PCA Subspace.

This subspace was used for recognition as PCA face space and as input to ICA and 2DPCA. After deriving the subspaces, all images from data sets were projected onto each subspace and recognition using nearest neighbor classification with various distance measures was conducted. The same procedure was employed for the ORL face database; however, three images were picked as probe set while the rest were trained when working on each of the algorithm.

Histogram equalization was applied on all the images in the databases including images used as probes in order to get efficient results. Comparisons were done based on the False Acceptance Rate (FAR) and False Rejection Rate (FRR) observed from each of the algorithms under different face databases and distance metrics. The next section shows the results obtained during the experiments.

V. Results

We present here the comparative results obtained during our analysis of the implemented algorithms for the recognition accuracy in terms of FAR and FRR. We applied a rescaling algorithm for all the images to be of the same scale. The algorithms were implemented successfully using MATLAB 7.0 and trained and simulated on a Pentium-IV (2.0 GHz), 2GB RAM to provide valuable results.

We computed our recognition accuracy by adding the percentage of false acceptance (in which the system mistakenly recognize the image of another person as the probe person) with that of false rejection (in which the system was unable to truly recognize a probe even though the probe's images are in the training database) and then divides it by two since we added the two false percentage up, this gives us an average error recognition. To get the recognition accuracy, we deducted the average error recognition from 100.

The tables below shows the summary of the results obtained, with each table describing the results obtained under each of the two face databases used. The first table shows the result obtained in the ORL database while the other one shows the result from the FERET database with each table giving the result for each separate gallery set of the FERET used.

TABLE I

Algorithm Used	Face Database	Distance Metric	Accuracy (%)
PCA	ORL	Mahalanobis	72.43
2DPCA	ORL	Mahalanobis	72.72
ICA	ORL	Mahalanobis	74.27
PCA	ORL	Euclidean	83.61
2DPCA	ORL	Euclidean	87.13
ICA	ORL	Euclidean	87.42
PCA	ORL	Cosine	76.37
2DPCA	ORL	Cosine	76.81
ICA	ORL	Cosine	82.13

Figure 1: Table showing the recognition accuracy under ORL face database.

TABLE II

Algorithm Used	Face Database	Distance Metric	Accuracy (%)
PCA	Fb	Mahalanobis	71.27
2DPCA	Fb	Mahalanobis	71.32
ICA	Fb	Mahalanobis	72.40
PCA	Fb	Euclidean	83.45
2DPCA	Fb	Euclidean	83.35
ICA	Fb	Euclidean	67.92
PCA	Fb	Cosine	81.91
2DPCA	Fb	Cosine	80.73
ICA	Fb	Cosine	81.02

Figure 2: Table showing the recognition accuracy under ORL face database.

TABLE III

Algorithm Used	Face Database	Distance Metric	Accuracy (%)
PCA	Fc	Mahalanobis	41.91
2DPCA	Fc	Mahalanobis	41.93
ICA	Fc	Mahalanobis	41.93
PCA	Fc	Euclidean	56.46
2DPCA	Fc	Euclidean	38.22
ICA	Fc	Euclidean	35.37
PCA	Fc	Cosine	52.31
2DPCA	Fc	Cosine	46.32
ICA	Fc	Cosine	51.29

Figure 3: Table showing the recognition accuracy under FERET face database.

TABLE IV

Algorithm Used	Face Database	Distance Metric	Accuracy (%)
PCA	Dup I	Mahalanobis	37.33
2DPCA	Dup I	Mahalanobis	37.34
ICA	Dup I	Mahalanobis	37.36
PCA	Dup I	Euclidean	35.23
2DPCA	Dup I	Euclidean	35.34
ICA	Dup I	Euclidean	35.19
PCA	Dup I	Cosine	36.72
2DPCA	Dup I	Cosine	33.28
ICA	Dup I	Cosine	36.84

Figure 4: Table showing the recognition accuracy under FERET face database

TABLE V

Algorithm Used	Face Database	Distance Metric	Accuracy (%)
PCA	Dup II	Mahalanobis	42.56
2DPCA	Dup II	Mahalanobis	43.40
ICA	Dup II	Mahalanobis	44.01
PCA	Dup II	Euclidean	33.13
2DPCA	Dup II	Euclidean	34.18
ICA	Dup II	Euclidean	35.29
PCA	Dup II	Cosine	21.78
2DPCA	Dup II	Cosine	23.51
ICA	Dup II	Cosine	24.93

Figure 5: Table showing the recognition accuracy under FERET face database

As can be observed from the tables above, the system based on ICA seems to give the best accuracy under the databases used even though it takes more time during training (the issue of training time has been ignored completely and thus not included in this work i.e. our analysis does not cover training and execution time because based on our personal view, we believe the most important factor is still the accuracy irrespective of the training and recognition time) compared to other systems that it was compared with. The 2DPCA also shows a considerable result, this is consistent with some earlier reports that 2DPCA outperforms its 1 Dimensional counterpart, PCA. Apparently, the best result was achieved under FERET when the Fb gallery set is in use while the DUPII gives the worst accuracy across all algorithms.

VI. Conclusion

This work presents a comparative analysis of three different face recognition systems. These include the PCA, a variant of PCA called 2DPCA and ICA which are all appearance based system. The algorithms were analyzed and compared based on their recognition accuracy. We have shown that the ICA gives the best result followed by 2DPCA which uses 2 Dimensional data as against the conventional PCA which uses 1 Dimensional data. This is consistent with some earlier work done similar to this but with slight differences. We have also shown that using different databases for comparison may give a statistically significant variance, as such; a baseline comparison should employ the use of several face databases before conclusion. Issues of distance metrics used have been observed also in this work, generally, we achieved best result when Cosine is used as the similarity calculator, followed by Euclidean and Mahalanobis distance. Since the algorithms analyzed here are based holistic based systems, further comparative work can include more algorithms both from the holistic and feature based approach to allow rich and extensive analysis.

Acknowledgment

The authors acknowledge the help of anonymous reviewers who helped improve this paper.

References

[1] W. Zhao, R. Chellapa, and P. J Phillips, "Face Recognition: A literature survey ," in *Technical Report, University of ,Maryland, 2000.*

[2] M. Kirby., and L. Sirovich., "Application of the Karhunen-Loeve procedure for the characterization of human faces*", IEEE PAMI, Vol. 12, pp. 103-108, (1990).*

[3] M. Turk, and A. Pentland., "*Eigenfaces for recognition", Journal of Cognitive Neuroscience, Vol. 3, pp. 71-86, (1991).*

[4] Jian Yang, David Zhang, Alejandro F. Frangi, and Jing-yu Yang: Two-Dimensional PCA: A New Approach to Appearance-Based Face Representation and Recognition. In *IEEE transactions on pattern analysis and machine intelligence, vol. 26, no. 1, PG 131-137january 2004*

[5] Volker Blanz, Thomas Vetter, "Face Recognition Based on Fitting a 3D Morphable Model*", IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol.25, No.9, pp.1-12, 2003.*

[6] Meng Joo Er, Shiqian Wu, Juwei Lu and Hock Lye Toh : Face Recognition With Radial Basis Function (RBF) Neural Networks, In *IEEE transactions on neural networks, vol. 13, no. 3, may 2002*

[7] V. Matyas and Z. Riha, Biometric authentication- security and usability, Faculty of informatics, masaryic university Brno, Czech Republic. 2008.

[8] Arvind, M. (2009) Biometrics - Finger print, Iris, Retina and Pupil Recognition : Biometrics for beginners, Version 3, Knol, 2009, *(http://knol.google.com/k/arvind-muthukrishnan/biometrics-finger-print-iris-retina-and/3ktp6efifnl1p/7).*

[9] Sanchez-Reillo, R. (2000) Hand geometry pattern recognition through Gaussian mixture modelling, *Proceedings of 15th International Conference on Pattern Recognition, Vol. 2 , Pp. 937 -940*.

[10] Li, S.Z. and Jain, A.K. Eds., (2005) Handbook of face recognition, Springer.

[11] Face Recognition , Edited by Kresimir Delac and Mislav Grgic, published by *I-TECH Education and Publishing ,Vienna, Austria, 2010*

[12] D. Murugan, S. Murugam,K. Rajalakshmi and T.I Manish, "Performance evaluation of face recognition using Gabor filter, Log Gabor filter and Discrete Wavelet Transform. *International Journal of computer science and information technology. Vol 2, no 1, feb 2010*

[13] Shashank N. Mathur, Anil K. Ahlawat, and Virendra P. Vishwakarma, Illumination Invariant Face Recognition using Supervised and Unsupervised Learning Algorithms, In *World Academy of Science, Engineering and Technology, 2008*

[14] Edward Wilson," Back propagation Learning for Systems with Discrete-Valued Functions", *Proceedings of the World Congress on Neural Networks, 1994*.

[15] Meng, J. and Zhang, W. (2007) Volume measure in 2DPCA-based face recognition, *Pattern Recognition Letters, Science Direct, Elsevier, Vol 28, Pp.1203-1208*.

[16] Bruce Draper, Kyungim Baek, Marian Bartlett, and Ross Beveridge: Recognizing faces with PCA and ICA, in *Computer Vision and Image Understanding 91 (2003) 115–137*

[17] KresimirDelac,MislavGrgic and SonjaGrgic: Independent Comparative Study of PCA, ICA, and LDA on the FERET DataSet. In *Wiley Periodicals, Inc.,. Int J Imaging Syst Technol, 15, 252–260, 2005*

[18] Adebayo Kolawole John and Onifade Olufade: "Employing Fuzzy Histogram equalization to combat illumination invariance in face recognition systems". In *international journal of Intelligent Systems and Applications, Singapore. Vol. 4, No 9, Pg. 54-60, June 2012. Available online at www.mec-press.org*

[19] Kong, H., Li, X.C., Eam, K.T., Wang, J.G. and Ronda, V. (2005) Generalized 2D principal component analysis, *Proceedings of IEEE International Joint Conference on Neural Networks, Montreal, Canada.*

[20] Yang, J., Zhang, D., Frangi, A.F., Yang, J.Y. (2004) Two-dimensional PCA: A new approach to appearance-based face representation and recognition, *IEEE Trans. Pattern Anal. Machine Intell, Vol.26, No.1, Pp.131–137.*

[21] M.S. Bartlett, J.R. Movellan, T.J. Sejnowski, Face recognition by independent component analysis, *IEEE Transaction on Neural Networks 13 (2002) 1450–1464.*

[22] C. Liu and H. Wechsler, Comparative Assessment of Independent Component Analysis (ICA) for Face Recognition, presented at *International Conference on Audio and Video Based Biometric Person Authentication, Washington, DC, 1999*.

[23] S.J. Lee, S.B. Yung, J.W. Kwon and S.H. Hong, "Face detection and recognition using PCA". *Pg84-87. IEEE. TENCOM, 1999.*

[24] Kong, H., Li, X.C., Eam, K.T., Wang, J.G. and Ronda, V. (2005) Generalized 2D principal component analysis, *Proceedings of IEEE International Joint Conference on Neural Networks, Montreal, Canada*.

[25] Wang, L.W., Wang, X. and Feng, J.F. (2006) On image matrix based feature extraction algorithms, *IEEE Trans. Systems Man Cybernet. – Part B, Cybern., Vol.36, No.1, Pp.194–197.*

[26] Phillips, P.J., Moon, H., Rizvi, S.A. and Rauss, P.J. (2000) The FERET evaluation methodology for face-recognition algorithms, *IEEE Trans. Pattern Anal. Machine Intell., Vol.22, No.10, Pp.1090–1104.*

[27] Adebayo K.J and Onifade O.F.W , "Biometric authentication with face recognition using principal component analysis and a feature-based technique", In *International Journal of Computer Applications (0975 – 8887) pg. 13 – 20, Volume 41– No.1, March 2012*

[28] Adebayo K.J and Onifade O.F.W, "Framework for a Dynamic Grid-Based Surveillance Face Recognition System". In *Africa journal of computing and ICT (IEEE Nigerian Section), pp. 1 – 10, Vol. 4 N0. 1, June, 2011*

[29] Adebayo Kolawole John and Onifade Olufade Williams: "Comparative analysis of PCA-based and Neural Network Based face recognition systems" In *proc of 12th International Conference on intelligent systems design and applications, ISDA 2012 and published by IEEE, held at Cochin India, Dec 2012.*

[30] Adebayo Kolawole John and Onifade Olufade Williams: "Performance evaluation of image compression on PCA based face recognition systems" In proc of *12th International Conference on Hybrid intelligent systems, HIS 2012 and published by IEEE, held at Pune India, Dec 2012*

APPENDIX

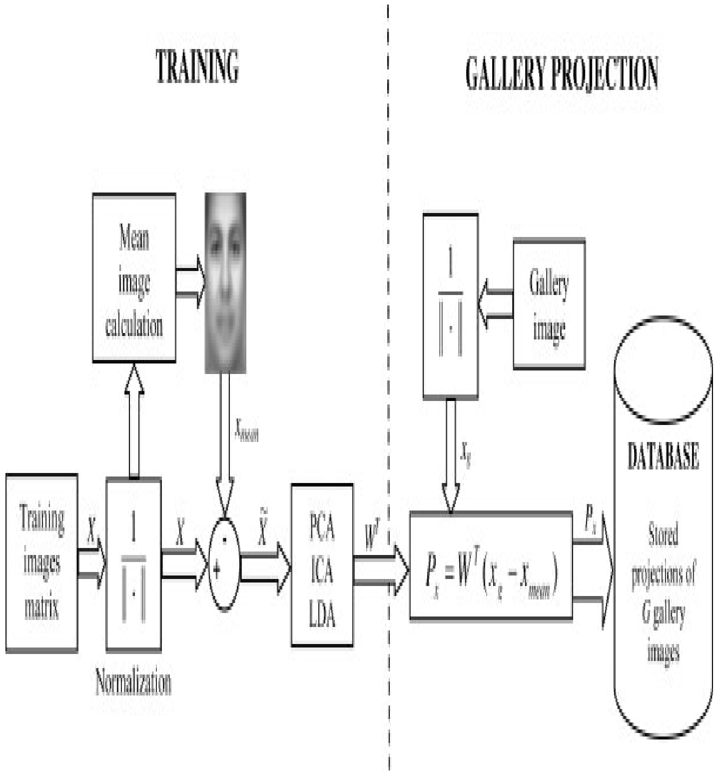

Figure 6: An illustration of general subspace appearance-based face recognition system [17]

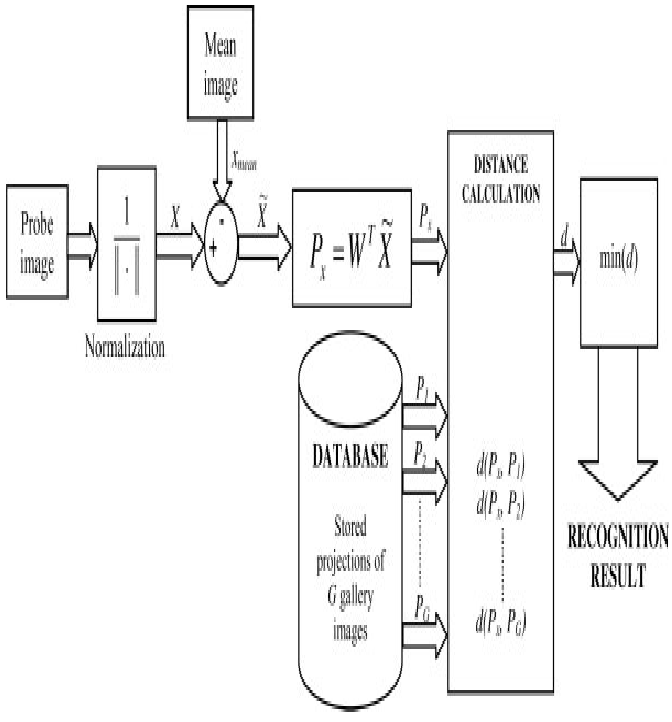

Figure 7: The matching phase of a general subspace face recognition system [17]

Author Biographies

Adebayo Kolawole John holds a B.Sc. and M.Sc. in Computer Science from University of Ilorin and University of Ibadan, Nigeria respectively. He is currently pursuing his Ph.D. in computer Science at the University of Ilorin, Nigeria. His research interest include Intelligent systems, Computer Vision, Biometrics and security systems and Image processing.. He can be reached at collawolley3@yahoo.com and +2348066078619

Dr. Olufade F. W. Onifade is a lecturer at the Department of Computer Science, University of Ibadan, Nigeria. He can be reached at fadowilly@yahoo.com or olufadefwilliams.onifade@loria.fr; Phone: + (234) 807-401-0558; + (234) 803-284-1817..

Fatai idowu yisa is a lecturer at the department of computer science, Oduduwa University, Ipetumodu, Osun state, Nigeria. He received M.Sc. in Computer Science from University of Ibadan, Ibadan (2010), Postgraduate Diploma in Education (PGDE), from Bayero University, Kano (2006) and B.Sc. Computer Science with Economics (Combined Honours) from Obafemi Awolowo University, Ile-Ife (1998). He is a member of Teachers' Registration Council of Nigeria (TRCN). His areas of research interest are m- & e-learning, cloud computing and metadata extraction. He can be reached by phone on +2348139621273 and through email: al_fattah2003@yahoo.com