Development of an Efficient Face Recognition System based on Linear and Nonlinear Algorithms

Filani Araoluwa S.
Department of Computer Science,
The Federal University of Technology,
P.M.B.704, Akure, Ondo State, Nigeria.

Adetunmbi Adebayo O.
Department of Computer Science,
The Federal University of Technology,
P.M.B.704, Akure, Ondo State, Nigeria.

ABSTRACT
This paper presents appearance based methods for face recognition using linear and nonlinear techniques. The linear algorithms used are Principal Component Analysis (PCA) and Linear Discriminant Analysis (LDA). The two nonlinear methods used are the Kernel Principal Components Analysis (KPCA) and Kernel Fisher Analysis (KFA). The linear dimensional reduction projection methods encode pattern information based on second order dependencies. The nonlinear methods are used to handle relationships among three or more pixels. In the final stage, Mahalinobis Cosine (MAHCOS) metric is used to define the similarity measure between two images. The experiment showed that LDA and KFA have the highest performance of 93.33 % from the CMC and ROC results when used with Gabor wavelets. The overall result using 400 images of AT&T database showed that the performance of the linear and nonlinear algorithms can be affected by the number of classes of the images, preprocessing of images, and the number of face images of the test sets used for recognition.

General Terms
Pattern Recognition, Biometrics.

Keywords
Face Recognition, Gabor Wavelets, Linear Face Recognition algorithms, Nonlinear Face Recognition Algorithms, Appearance Based Methods

1. INTRODUCTION
Face biometrics recognition does not need to make direct contact with an individual during verification and identification. This why it is usually used with other biometics when developing a multibiometric system. Although a number of face recognition algorithms appear to be robust in constrained environments, face recognition does not achieve its optimum performance in real applications [1][2]. In face recognition, appearance-based approach has been mostly used [3][4]. Compared to holistic approaches, feature-based methods are less sensitive to variations in illumination and viewpoint [1]. The methods that are used for appearance face recognition can be classified as linear and nonlinear subspaces [5][6]. Linear methods transform data from high dimensional subspace into low dimensional subspace by linear mapping. In nonlinear techniques, low dimensional data matrix is obtained directly from high dimensional data matrix. Common linear algorithms are PCA, LDA and ICA [1] [7], [2]. Nonlinear techniques tend to addresses hidden nonlinear submanifold not addressed by linear methods [8], [9], [2]. Common examples of nonlinear techniques are the kernel methods [3] [10] and Support Vector Machine (SVM) [11]. Many of variation of different face recognition algorithms have been developed to improve performance [1], [6]. Apart from combination or fusion, some newer algorithms were robust using preprocessing techniques, [12] [2]. However, it has been shown that nonlinear vector like do not always performed better than linear methods in real-world data sets having more complicated distributions, though they easily demonstrate their virtue on artificial nonlinear data [4]. This paper study the result of the using two basics types of face recognition algorithms: the linear and nonlinear algorithms. This was done in view to show which robust algorithm will be useful in certain circumstances. A noted powerful algorithm preprocessing algorithm: Gabor Wavelet [13][14], is used in combination with the algorithms.

2. METHODOLOGY
Figure 1 show the Frame work used. The preprocessing was achieved by the integration of Gabor wavelet for extraction process. The Gabor wavelets (kernels, filters) used is defined as:

$$\varphi_{n,b}(\mathrm{z}) = \frac{\|k_{n,b}\|^2}{\sigma^2} e^{-\frac{\|k_{n,b}\|^2\|z\|^2}{2\sigma^2}} \left[e^{ik_{n,b}z} - e^{-\frac{\sigma^2}{2}} \right] \quad (1)$$

where n and b define the orientation and scale of the Gabor kernels, $z = (x, y)$, $\|.\|$ denotes the norm operator, and the wave vector $k_{n,b}$ is defined as follows:

$$k_{n,b} = k_b e^{i\emptyset_n} \quad (2)$$

where $k_b = k_{max} / f^b$ and $\emptyset_n = \pi n/ 8$. k_{max} is the maximum frequency, and f is the spacing factor between kernels in the frequency domain. The effect of the difference of convex functions term becomes negligible when the parameter σ, has sufficiently large values. Gabor wavelets used is of five different scales, $b \in \{0, \dots, 4\}$, and eight orientations, $n \in \{0, \dots, 7\}$ [13]. The Gabor wavelet representation of an

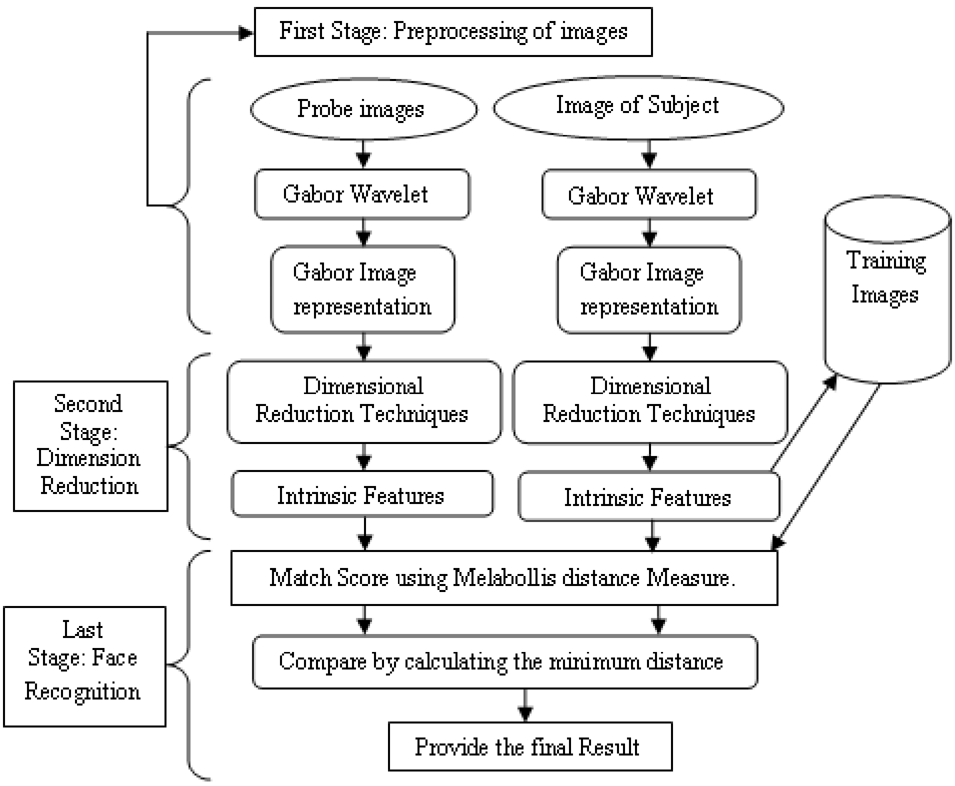

Fig.1. The System Frame Work

image is the convolution of the image as defined by (1). Let $I(x,y)$ be the gray-level distribution of an image, the convolution of image I is defined as:

$$O_{n,b}(z) = I(z) * \varphi_{n,b}(z) \qquad (3)$$

where the following counts $z = (x, y)$, $*$ denotes the convolution operator, and $O_{n,b}(z)$ is the convolution result corresponding to the Gabor kernel at orientation n and scale b. Which produce the set

$$S = \left\{O_{n,b}(z): n \in \{0, \dots, 7\}, b \in \{0, \dots, 4\}\right\} \qquad (4)$$

Equation (4) forms the Gabor wavelet representation of the image $I(z)$. For better result, all the representation results are concatenated and an augmented feature vector X is derived by downsampling [12], [14]. Before the matching, the preprocessed probe images are projected onto the same subspace as the preprocessed gallery images using a similar algorithms. The test image projection is then compared to stored gallery projections by using Mahalanobis Cosine distance metric. The dataset are from the AT&T face database. The first three samples are selected for training, the next four samples were preserved as the test set, the remaining samples were used as the evaluating set. The algorithms used are described in the next section.

2.1 Principal Component Analysis (PCA*)*

PCA is used to find a *t*-dimensional subspace whose basis vector correspond to the maximum variance, where (t <v), The basic vectors are defined as eigenvectors of the scatter matrix V_T is defined as

$$\sum_{i=1}^{M}(x_i - \mu).(x_i - \mu)^T \qquad (5)$$

where μ is the mean of all M images in the training set or the mean face, T is the transpose of its properties and x_i is the *i*th image with its columns concatenated in a vector. The Principal components of t eigenvectors are t largest eigenvalues [7].

2.2 Linear Discriminant Analysis (LDA)

It considers for all samples of all classes, the between-class scatter matrix S_B and the within-class scatter matrix S_W which are defined by

$$S_B = \sum_{i=1}^{c} M_i.(x_i - \mu).(x_i - \mu)^T \qquad (6)$$

where M_i is the number of training samples in class i, c is the number of distinct classes, μ_i is the mean vector of samples belonging to class i and X_i represents the set of samples belonging to class i with x_k being the k–th image of that class. T is the transpose of its properties. S_B represents the scatter of features around the overall mean for all face classes and S_W represents the scatter of features around the mean of each face class. The goal is to maximize the ratio det$|S_B|/|S_w|$[10]. This ratio is maximized when the column vectors of the projection matric (W_{LDA}) are the eigenvectors of $Sw^{-1}.S_B$. In order to prevent S_w to become

$$S_W = \sum_{i=1}^{c}\sum_{x_k \in X_i}(x_i - \mu_i).(x_i - \mu_i)^T$$

(7)

singular, PCA is used as a preprocessing step to get: $W_{opt}^T = {W_{LDA}}^T {W_{PCA}}^T$ [7].

2.3 Kernel Principal Component Analysis (KPCA)

By considering the set of image samples X_k, $x_k = [x_{k_1}, \dots, x_{k_n}]^T \in R^n$ (8)

Kernel PCA projects each vector x from the input space, R^n, to a high dimensional feature space, R^f, by a nonlinear mapping function: Φ: $R^n \rightarrow R^f$, f > n. PCA process is then carried out on the kernel subspaces by solving the corresponding eigenvalue problem:

$$\lambda w^{\Phi} = C^{\Phi} w^{\Phi} \qquad (9)$$

where C^{Φ} is a covariance matrix. All solution w^{Φ} with $\lambda \neq 0$ lie in the span of $\Phi(x_1),\ldots,\Phi(x_m)$ [10].

2.4 Kernel Fisher Analysis (KFA)

KFA is performed using the similar procedure of KPCA except that Fisher Linear Discriminant (FLD) is considered instead of PCA after the transformation of the subspace to higher dimension. If x_k has the same value of equation (8), the same projection is performed on the vector x. to get the function Φ: $R^n \rightarrow R^f$, f > n. Let the projected samples Φ(x) be centred in R^f and let the equations that use dot products be formulated for Fisher linear Discrimate Analysis (FLD) only. Assume the within-class and between-class scatter matrices be S_W^{Φ} and S_B^{Φ}, to apply FLD in kernel space, the solution to eigenvalues λ and eigenvectors w^{Φ} of

$$S_W^{\Phi} w^{\Phi} = S_B^{\Phi} w^{\Phi} \qquad (10)$$

are derived by finding the eigenvectors corresponding to largest generalized eigenvalue. The kernel function is introduce defined by

$$(k_{rs})_{tu} = k(x_{tr}, x_{us}) = \Phi(x_{tr}).\Phi(x_{us}) \qquad (11)$$

where there exists a *c*-class problem and a *r*-th sample of class t and the s-th sample of class u be x_{tr} and x_{us} respectively (where class t has l_t samples and class u has l_u samples). Then finally project $\Phi(x)$ to a lower dimensional space spanned by the eigenvectors w^{Φ} in a way similar to Kernel PCA [14].

3. MATCHING

For the matching task, the Mahalinobis Cosine (MAHCOS) distance metric is used. This is because it is the most accurate and efficient in terms of verification, identification and robustness [15]. After transformations are completed, for two images *u* and *v* with corresponding projections *m* and *n* in Mahalinobis space, where *m* and *n* are two feature vectors transformed into Mahalinobis space, the Mahalinobis Cosine is [16]:

$$S_{MahCosine}(u, v) = \cos(\Theta_{mn}) = \frac{|m||n|\cos(\Theta_{mn})}{|m||n|} = \frac{m.n}{|m||n|} \qquad (12)$$

with an angle Θ defined as the angle between the images after they have been projected into the recognition space as distance between projected images.

4. EVALUATION AND RESULTS.

4.1 Interpretation of Results

Table 1 gives the detailed summary of the results. There are 400 images containing 10 different images of each person. 120 images are used for training, 160 images are used for testing, the remaining images serve as the evaluation sets. From the results, LDA outperformed other methods in most cases and have a very close recognition rate with KFA. Both LDA and KFA have the highest performance of 93.33 % from the CMC and ROC results when the evaluating set is used with Gabor. Their performance using CMC and ROC with Gabor with more test images is 91.88% and 93.13% respectively (i.e. when the test set is used). It is obvious that the KFA and LDA high performance is due to high number of classes of the images of the system (the database has a total of 40 between-class (matrices) and 10 within-class images). The results of the experiment show that incorporating Gabor image representation increases the face recognition performance of all the algorithms. When more number of test images are used with the Gabor wavelets, there is a general reduction in performance of all the algorithms and there is more reduction when PCA based algorithms are used (PCA and KPCA) than the LDA based algorithms (LDA and KFA). For example with Gabor Wavelets, the ROC performance of both PCA and KPCA recognition rate for the evaluating set(which contain 120 probe images) are both 92.50% and they decrease to 63.13% and 56.88% when the test set (which contains 160 images) is used. With the use of Gabor Wavelets, LDA and KFA ROC performance are both 93.33% when the evaluating set is used and they just decrease to 91.88% and 93.13% when the test set is used. This shows that LDA based algorithms still perform better when the number of test/probe is increased. It can also be seen that KPCA perform worst having the highest error rates (2.68% with Gabor and 8.80% without Gabor). It also has the lowest recognition rate. It performs worse than PCA but only perform better than PCA (from the CMC results with Gabor on the evaluating set) when the Gabor filters is used. Overall the linear based algorithm still performs better than the nonlinear ones. The next section contain the conclusions drawn from the results obtained from the experiment.

4.1.1 The performance of the linear and nonlinear algorithms depends on some conditions. These are explained bellow:

4.1.1.1 The number of classes of a facial recognition system can affects the performance of the type of linear and nonlinear algorithm used. LDA (a linear algorithm) and KFA (a nonlinear algorithm) expressly provides best discrimination among classes.

4.1.1.2 The preprocessing using Gabor filters increases the recognition rate of both the linear and nonlinear algorithms.

4.1.1.3 When more test images are used after preprocessing with Gabor wavelets, there is reduction in recognition rate of all algorithms however the reduction is more for the PCA based algorithms than the LDA based ones. This shows that the increase in number of test images can affects recognition rate (of all the algorithms) negatively but the LDA based (classed based) algorithms are less affected than the PCA based ones.

4.1.2 From the overall results the linear algorithms is better than the nonlinear ones.

5. CONCLUSION

The results show that the number of classes and test images of a facial recognition system can have an effect on the recognition rate of a particular algorithm used. Incorporating Gabor image representation with linear and nonlinear algorithms increases their recognition rate. Linear subspace techniques tend to perform better than the nonlinear linear ones from the result of the work carried out. The research will be of outmost importance to any organization that wishes to develop a facial recognition system and know which of the face recognition algorithms have a better recognition rate. This study will also be of immense benefit to prospective researchers who would like to undertake similar studies.

This work is able to compare linear and nonlinear face recognition algorithms produced. The research is only concern about 2D holistic face recognition algorithm. A new development can make use of 2D local based appearance face recognition algorithms using linear and nonlinear algorithms.

Other areas can involve analysis of different 3D linear and nonlinear face recognition algorithms.

Table 1. Recognition rates using different Face Recognition performance metrices.

	Face Recognition Performance Metrics						
	CMC		ROC			EPC	
	Without Gabor Filters (Using Evaluating Set) in (%)	With Gabor Filters (Using the Evaluating set) in (%)	Without Gabor Filters (Using Evaluating Set) in VR (%)	With Gabor Filters (Using the Evaluating set) in VR (%)	With Gabor Filters (Using the test set) in VR (%)	With Gabor FiltersH TER (%)	Without Gabor FiltersHT ER (%)
PCA	66.07	74.17	66.79	92.50	63.13	1.61	4.72
LDA	86.07	93.33	76.43	93.33	91.88	1.56	4.09
KPCA	49.29	80.00	51.43	92.50	56.88	2.68	8.80
KFA	85.71	93.33	60.71	93.33	93.13	1.85	6.66

Keys: VR = Verification Rate; HTER = Half Total Error Rate; CMC = Cumulative Match Curve; ROC = Receiver Operating Characteristics. CMC results are at Rank One Recognition Rate (in percentage).EPC = Expected Performance Curve.

6. REFERENCES

[1] Zhao, W., Chellappa, R., Phillips, P. J., and Rosenfeld, A. 2003. Face recognition: A literature survey.

[2] Abate, A. F., Nappi, M., Ricco, D. and Sabatino G. 2007. 2D and 3D face recognition: A survey.

[3] Beham, M. P. and Roomi, S. S. M. 2012. Face recognition using appearance based approach: a literature survey.

[4] Huang, W. and Yin, H. 2012. On nonlinear dimensionality reduction for face recognition.

[5] Shah, J. H., Sharif, M. and Azeem, A. 2013. A survey: linear and nonlinear pca based face recognition techniques .

[6] Struc, V. Milhelic, M. and Pavesic, N. 2008. Combining experts for improved face verification performance.

[7] Delac, K. Grgic, M. and Grgic, S. 2006. Independent comparative study of PCA, ICA, and LDA on FERET data set.

[8] Sadykhov R. and Frolov, I. 2010.The development features of the face recognition system.

[9] Zhang, B., Chen, X., Shan, .S. and Gao, W. 2005. Nonlinear face recognition based on maximum average margin criterion.

[10] Yang, M. 2002. Kernel eigenfaces vs. kernel fisherfaces: Face recognition using kernel methods. In proceedings of the Fifth IEEE International Conference on Automatic Face and Gesture Recognition

[11] Igor, F.and Rauf, S. 2009. The techniques for face recognition with support vector machines. In Proceedings of the IMCSIT.

[12] Thomas, L. L. , Gopakumar, C. and Thomas, A. A. 2013. Face recognition based on gabor wavelet and backpropagation neural network.

[13] Shen, L., Bai, L. and Fairhurst, M. 2007. Gabor wavelets and General Discriminant Analysis for face identification and verification.

[14] Liu, C.and Wechsler, H. 2002.Gabor feature based classification using the enhanced fisher linear discriminant model for face recognition.

[15] Ibrahim, R. M. Abou-Chadi, F. E. Z. and Samra, A. S. 2013. Plastic Surgery Face Recognition: A comparative Study of Performance.

[16] Mar, N. S. S. Fookes, C. and Yarlagadda, P. K. D. V. 2012. Solder joint defects classification using the Log-Gabor Filter, the Discrete Wavelet Transform and the Discrete Cosine Transform.