Statistical Evaluation of Emerging Feature Extraction Techniques for Aging-Invariant Face Recognition Systems

[1]*Ayodele Oloyede, [2]*Temitayo M. Fagbola, [1]Stephen O. Olabiyisi [1]Elijah O. Omidiora and [1]John B. Oladosu

[1]Department of Computer Science and Engineering, Ladoke Akintola University of Technology, Ogbomoso, Nigeria

[2]Department of Computer Science, Federal University, Oye-Ekiti, Nigeria

deledeee@yahoo.com | temitayo.fagbola@fuoye.edu.ng | {soolabiyisi | eoomidiora | jboladosu}@lautech.edu.ng

Abstract— Large variation in facial appearances of the same individual makes most baseline Aging-Invariant Face Recognition Systems (AI-FRS) suffer from high automatic misclassification of faces. However, some Aging-Invariant Feature Extraction Techniques (AI-FET) for AI-FRS are emerging to help achieve good recognition results when compared to some baseline transformations in conditions with large amount of variations in facial texture and shape. However, the performance results of these AI-FET need to be further investigated statistically to avoid being misled. Statistical significance test can be used to logically justify such performance claims. The statistical significance test would serve as a decision rule to determine the degree of acceptability of the probability to make a wrong decision should such performance claims be found faulty. In this paper, the means between the quantitative results of emerging AI-FET (Histogram of Gradient (HoG), Principal Component Analysis-Linear Discriminant Analysis (PCA-LDA) and Local Binary Pattern-Gabor Wavelet Transform (LBP-GWT)) and the baseline aging-invariant techniques (Local Binary Pattern (LBP) and Gabor Wavelet Transform (GWT)) were computed and compared to determine if those means are statistically significantly different from each other using one-way Analysis of Variance (ANOVA). The ANOVA results obtained at 0.05 critical significance level indicate that the results of the emerging AI-FET techniques are not statistically significantly different from those of baseline techniques because the *F*-critical value was found to be greater than the value of the calculated F-statistics in all the evaluations conducted.

Keywords— Statistical Evaluation, Emerging Feature Extraction Technique, Aging-Invariant Face Recognition

1 Introduction

Face recognition across ages is an important problem and has many applications such as passport photo verification, image retrieval and surveillance (Narayanan and Rama, 2006). While facial aging is largely continuous in younger age groups, it is also represented by relatively large texture changes and minor shape changes due to the change of weight, presence of wrinkles or stiffness of skin in older age groups above 18 years (Oloyede *et al.*, 2016). Therefore, Aging-Invariant Face Recognition Systems (AI-FRS) need to be able to manage recognition in both. However, robustness of AI-FETs to variations across illumination, pose, facial expressions and aging is a phenomenal factor to the effectiveness of AI-FRS. Recently, several emerging aging-invariant feature extraction techniques (AI-FET) claimed to realize robust aging-invariant recognition of faces have been proposed.

However, the performance results of these AI-FETs need to be further investigated statistically to avoid being misled. Statistical significance test can be used to logically justify such performance claims. The test would serve as a decision rule to determine the degree of acceptability of the probability to make a wrong decision should such performance claims be found faulty. In this paper, the means of the quantitative results of emerging AI-FET (Histogram of Gradient (HoG), Principal Component Analysis-Linear Discriminant Analysis (PCA-LDA) and Local Binary Pattern-Gabor Wavelet Transform (LBP-GWT)) and the baseline aging-invariant techniques (Local Binary Pattern (LBP) and Gabor Wavelet Transform (GWT)) were computed and compared to determine if those means are statistically significantly different from each other using one-way Analysis of Variance (ANOVA). To this end, the significant difference in the results of the emerging AI-FET and the baseline FET for AI-FRS was determined to further support or refute claims of improved performance of the emerging AI-FET.

*Corresponding Authors

2 Baseline Feature Extraction Techniques

These are early approaches to person recognition with large variations in the face of same individual due to aging. Most of these literal approaches rely on the geometry of fiducial points from the facial features (eyes lids, lips, and nose) and their spatial relationships. In this paper, Local Binary Pattern (LBP) and Gabor Wavelet Transform (GWT), which are the two (2) most commonly adopted baseline techniques were benchmarked with some emerging techniques.

2.1 Gabor Wavelet Transform

A Gabor wavelet can be described as a Gaussian kernel function modulated by a sinusoidal plane wave that has an optimal location in both the frequency domain and the space domain (Kepenekci, 2001). Due to the useful characteristics of Gabor functions, they have been widely and successfully applied to texture segmentation, handwritten numerals recognition, fingerprint recognition and face recognition (Zhang *et al.*, 2005). Gabor features are used to represent the features extracted by a set of Gabor wavelets; they are usually called jets when the wavelet family is applied at a certain facial feature point. Gabor wavelets reveal the directional features of an image while providing a fine adjustment for frequency properties. The decomposition of an image I into these states is called the wavelet transform of the image. A wavelet transform is created by passing the image through a series of filter bank stages. The Gabor wavelet transform uses a set of Gaussian enveloped basis functions that are orthogonal-like basis functions. Shen and Bai (2006) asserted that Gabor wavelet transform seems to be the optimal basis to extract local facial features.

Given an image I(x,y), its Gabor wavelet transform is defined as:

$$W_{mn}(x, y) = \int I(x_1, y_1) g^*_{mn}(x - x_1,\ y - y_1) d_{x1} d_{y1} \qquad (1)$$

where * indicates the complex conjugate. Since the local texture regions are spatially homogeneous (Anila et al., 2011), the mean μ_{mn} and standard deviation σ_{mn} of the magnitude of transform coefficients are used to represent the regions for classification such that:

$$\mu_{mn} = \iint W_{mn}(x,y)| \, dxdy \quad (2)$$

and

$$\sigma_{mn} = \sqrt{\iint (|W_{mn}(x,y)| - \mu_{mn})^2} \; dxdy \quad (3)$$

A feature vector is then constructed using μ_{mn} and σ_{mn} as feature components. Let f_i and f_j represent the feature vector of test and train image respectively; then, the distance between two images in the feature space can be defined to be:

$$d\,(i, j) = \sum_i \sum_j d_{mn}(i,j) \quad (4)$$

where

$$dmn\,(i, j) = \left| \frac{\mu_{mn}^{(i)} - \mu_{mn}^{(j)}}{\alpha(\mu_{mn})} \right| + \left| \frac{\sigma_{mn}^{(i)} - \sigma_{mn}^{(j)}}{\alpha(\sigma_{mn})} \right| \quad (5)$$

The test image will then be referred to class k if dk is the minimum value of di for test image.

2.2 Local Binary Pattern

According to Ahonen and Pietikäinen (2007), the original LBP algorithm is a grayscale irrelevant texture analysis algorithm with powerful discrimination. In more details, Ojala Pietikäinen and Mäenpaa (2002) described LBP as a gray-scale texture operator characterized by the spatial structure of the local image texture. The authors claimed that LBP can provide a unified description including both statistical and structural characteristics of a texture patch, which makes it very effective for texture analysis. The flowchart of the LBP process for face recognition is presented in Figure 1. Given a central pixel in the image, a pattern number is computed by comparing its value with those of its neighborhoods. With the neighborhood set P and a circle of radius R, and the difference between the central pixel "gc" and its neighborhood {g0, g1,…,gp-1}, the value of LBP operator can be obtained as (Ojala et al., 2002):

$$\text{LBP P,R} = \sum_{i=0}^{p-1} s\,(g_i - g_c)2^p \quad (6)$$

$$s = \begin{cases} 1 & g_i - g_c > 0 \\ 0 & g_i - g_c \le 0 \end{cases} \quad (7)$$

The original LBP labels the pixels of an image by thresholding the local area, neighborhood of each pixel with the center value and considering the result as a binary number. Equation 6 means pixels greater than the central pixel are mapped to 1, otherwise. Equations 6 and 7 give the computation of LBPP,R. After identifying the LBP pattern of each pixel (i, j), the whole texture image is represented by building a histogram which is used as a texture descriptor. The LBP histogram contains information about the distribution of the local micro-patterns, such as edges, spots and flat areas, over the whole image, which can be used to statistically describe image characteristics. Figure 2 shows the sample histogram extracted from an image with LBP operator.

The histogram of labeled image $f_i(x,y)$ is defined as (Ahonen and Pietikäinen, 2007):

$$H\,(i) = \sum_{x,y} I\,\{f_i\,(x,y) = i\},\ i = 0,\dots,n-1 \quad (8)$$

where n is the number of different labels produced by LBP operator and

$$I\{x\} = \begin{cases} 1, & x \text{ is true} \\ 0, & x \text{ is false} \end{cases} \quad (9)$$

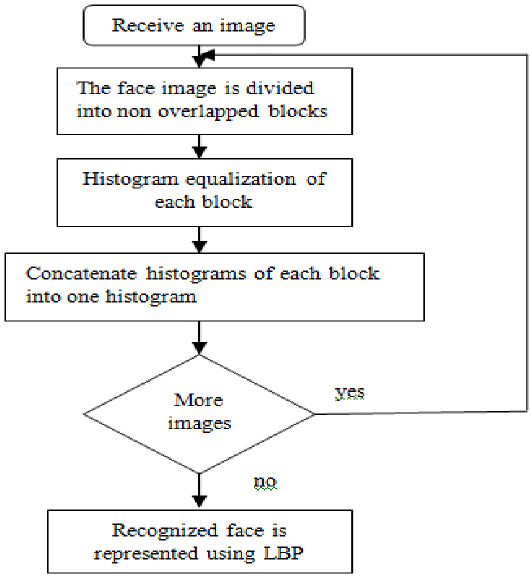

Fig. 1: Control Flow of the LBP Process for face recognition (Ojala et al., 2002)

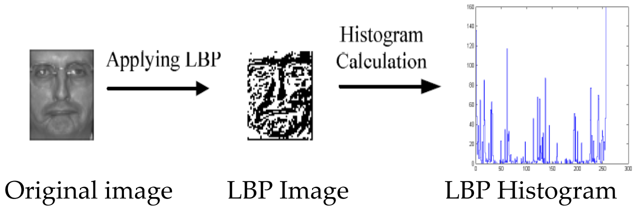

Fig. 2: LBP histogram for a facial image (Ojala et al, 2002)

3 Emerging Aging-Invariant Feature Extraction Techniques

In this section, a concise description of some emerging aging-invariant feature extraction techniques is presented. The techniques considered include Histogram of Gradient (HoG), Principal Component Analysis-Linear Discriminant Analysis (PCA-LDA) and hybrid LBP-GWT.

3.1 Hybrid LBP-GWT for Aging-Invariant Face Recognition

The hybrid LBP-GWT AI-FET was developed by Oloyede et al. (2016). The algorithm is described as follows: Given the coordinates of the centre pixel of an image I(x,y) defined as (xc, yc), the coordinates of the P neighbors (xp, yp) on the edge of the circle with radius R was calculated with the cosine rule:

$$Xp = Xc + R\cos\left(\frac{2\pi p}{P}\right) \quad (10)$$

The algorithm is as follows:

Input a: Training and Test Image set

i. Initialize temp = 0
ii. FOR each image I in the training image set
iii. Initialize the pattern histogram, H = 0
iv. FOR each center pixel tc ε I
v. Compute the pattern label of tc, LBP using Equation (10)
vi. Increase the corresponding bin by 1.
vii. END FOR
viii. Find the highest LBP feature for each face image
ix. Apply particle swarm optimization for feature subset selection

Intermediate Output A: Reduced LBP features of face image

In the same vein, the GWT was implemented as a process depicted in the algorithm as follows:
Input b: Training and Test Image set

i. Convolve Image I (x, y) using Gabor wavelets to extract local features at these feature points
ii. Calculate the mean deviation, μmn, of the Gabor wavelet coefficients for each point
iii. Calculate the standard deviation, σmn, of the Gabor wavelet coefficients for each point
iv. Construct Gabor feature vector using μmn and σmn.
v. Apply particle swarm optimization for feature subset selection

Intermediate Output B: Reduced GWT features of face image

Repeat for all features
For each feature in LBP, choose a corresponding feature in GWT
Take average of each matching features in LBP and GWT
Apply sum rule fusion strategy
End Repeat

3.2 Histogram of Gradient for Aging-Invariant Face Recognition

Dihong *et al.* (2013) developed a new method called Histogram of Gradient (HoG) via Hidden Factor Analysis (HFA). This approach is motivated by the belief that the facial image of a person can be expressed as combination of two components: an identity-specific component that is stable over the aging process, and the other component that reflects the aging effect. Two latent factors were introduced: an identity factor and an age factor, which respectively govern the generation of these two components. Intuitively, each person is associated with a distinct identity factor, which is largely invariant over the aging process and thus can be used as a stable feature for face recognition; while the age factor changes as the person grows.

In this process, the latent factors and the model parameters are iteratively updated to maximize a unified objective (Dihong et al., 2013). In the testing, given a pair of face images with unknown ages, the match score between them were computed by inferring and comparing the posterior mean of their identity factors. Every face image is divided into a set of overlapping patches, and then applied the HOG descriptor on each patch to extract the HOG features. The extracted HOG features from all the patches were concatenated together to form a long feature vector for further analysis. Prior to applying the HOG feature extractor, the face images were pre-processed through the following steps:

i. Rotate the face images to align them to the vertical orientation;
ii. Scale the face images so that the distances between the two eyes are the same for all images;
iii. Crop the face images to remove the background and hair region;
iv. Apply histogram equalization to the cropped face images for photometric normalization.

At the training stage, the training faces were first grouped according to their identities and ages, followed by feature extraction on each image. With each training face represented by HOG feature, the dimension of these features was reduced with slicing using PCA and LDA. Finally, HFA models were adapted independently on each of the sliced features of the dataset, obtaining a set of model parameters for each slice. At the testing stage, the matching score of the given face pair (one from probe and the other one from gallery) was computed by first going through feature extraction and dimension reduction steps the same as training, then estimating the identity latent variables for each slice of the two face features. The final matching score was given by the cosine distance of the concatenated identity features.

3.3 Aging-Invariant Principal Component Analysis – Linear Discriminant Analysis

Based on the assertion by Shinde and Gunjal (2012) that holistic approaches based on PCA and LDA suffer from high curse of dimensionality. That is, the time required for an algorithm grows exponentially with the number of features involved, rendering the algorithm intractable in extremely high-dimensional problems. Huseyin and Osen (2012) in an attempt to develop a more robust AI-FET, used PCA and subspace LDA methods for feature extraction of the face images. PCA projects images into a subspace such that the first orthogonal dimension of this subspace captures the greatest amount of variance among the images and the last dimension of this subspace captures the least amount of variance among the images. In this respect, the eigenvectors of the covariance matrix are computed which correspond to the directions of the principal components of the original data and their statistical significance is given by their corresponding eigenvalues. PCA was used for the purpose of dimension reduction by generalizing the data while Support Vector Machine (SVM) was used for the final classification.

4 METHODOLOGY

In this section, the implementation of the emerging and baseline AI-FET and the statistical significance test are discussed.

4.1 Implementation of Aging-Invariant Feature Extraction Techniques

The FG-NET (Face and Gesture Recognition Research Network) aging database was used and is composed of 1002 images of 82 subjects (6 - 18 images per subject) in the age range 0 – 69 years. The database also provides 68 landmark features that were identified manually, on all the face images. In addition, the following meta-information is available for all the images in the dataset namely: image size, age, gender, spectacles, hat, mustache, beard, horizontal pose and vertical pose. Since the images were retrieved from real-life albums of different subjects, aspects such as illumination, head pose and facial expressions are uncontrolled in this dataset. Nevertheless, this database is the only publicly available resource that provides quite a few age separated face images of individuals in the age range 0 - 18 years. Figure 3 presents some examples of images from FG-NET while figure 4 presents some cropped faces in FG-NET.

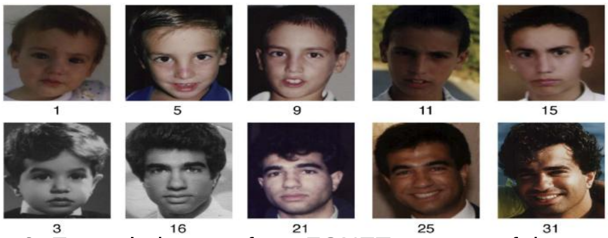

Fig. 3: Example images from FGNET. Images of the same row are of the same subject. The number at the bottom shows the age of the image.

Fig 4: Examples of cropped faces in the FG-NET Aging Dataset

The images in the FG-Net aging dataset were geometrically normalized manually and the illumination was normalized using rgb2gray function in MATLAB environment for training and testing purposes. At the training stage, the training faces were first grouped according to their identities and ages, followed by feature extraction on each image. With each training face represented by LBP, GWT, HoG, PCA-LDA and LBP-GWT features, the dimension of these features was reduced using particle swarm optimization and the cosine distance determined. At the testing stage, the matching score of the probe was computed by first going through feature extraction and dimension reduction steps the same as training, then estimating the cosine distance the face features. The final matching score is given by the trained face with the closest cosine distance.

The choice of cosine distance was informed by the fact that it is widely effective in high-dimensional positive spaces (Kuldeep and Madan, 2013). All the algorithms were implemented using MATLAB 7.7.0 (R2008b) on Windows 7 Ultimate 32-bit operating system, AMD Athlon (tm) X2 Dual Core QL-66 central processing unit with a speed of 2.2GHZ, 2GB random access memory and 320GB hard disk drive. The performance evaluation metrics that were used to evaluate the feature extraction techniques include the False Accept (FA), the False Reject (FR), Recognition Accuracy (RA) and Recognition Time (RT).

i. The False Accept Rate (FAR): This is the percentage of probes a system falsely accepts even though their claimed identities are incorrect (Raghavender, 2008).

$$\text{FAR} = \frac{Number of false accepts}{Number of impostor scores} \qquad (11)$$

ii. The False Reject Rate (FRR): This is the percentage of probes a system falsely rejects despite the fact that their claimed identities are correct.

A false accept occurs when the recognition system decides a false claim is true and a false reject occurs when the system decides a true claim is false (Raghavender, 2008).

$$FRR = \frac{Number of false rejects}{Number of genuine scores} \qquad (12)$$

iii. Recognition Accuracy: This is the main measurement to describe the accuracy of a recognition system. It represents the number of faces that are correctly recognized from the total number of faces tested (Jeremiah *et al.*, 2012).

Recognition Accuracy =

$$\frac{\text{Number of correctly recognized persons}}{\text{Total number of persons tested}} \times 100\% \qquad (13)$$

iv. Recognition Time: This represents the time required to process and recognize all faces in the testing set.

4.2 Statistical Significance Analysis

The statistical significance analysis of the quantitative results of emerging AI-FET (Histogram of Gradient (HoG), Principal Component Analysis-Linear Discriminant Analysis (PCA-LDA) and Local Binary Pattern-Gabor Wavelet Transform (LBP-GWT)) and the baseline aging-invariant techniques (Local Binary Pattern (LBP) and Gabor Wavelet Transform (GWT)) were computed and compared using one-way Analysis of Variance (ANOVA) to determine if those means are statistically significantly different from each other. Analysis of Variance (ANOVA) is a statistical method used to test differences between two or more means. Specifically, the null hypothesis:

$$H_O: \mu_1 = \mu_2 = \mu_3 = \cdots = \mu_k \qquad (14)$$

is tested where μ = group mean and k = number of groups. ANOVA uses F-test to determine whether the variability between group means is larger than the variability of the observations within the groups. Fisher-statistics is a ratio based on mean squares and used to assess the equality of variances, which is an estimate of population variance that accounts for the degrees of freedom used to calculate the estimate. However, to determine which specific methods or metrics differed from each other, a Least Significance Difference (LSD) Post Hoc test is proposed. Least Significance Difference (LSD) Post Hoc test is conducted in situations the results are found to be statistically significant to further determine the groups with the actual significant differences.

$$\text{LSD PostHoc Test} = t\sqrt{MSW}\left(\frac{1}{N_1} + \frac{1}{N_2}\right) \qquad (15)$$

where t = critical value of the tail, N is sample size of each method and MSW is the Mean Square Within.

5 RESULTS AND DISCUSSION

In this section, the quantitative results of all the implementations of the AI-FET and statistical significance tests conducted are discussed.

5.1 Results and Discussion on Aging-Invariant Feature Extraction Techniques

LBP-GWT feature extraction technique produced FA of 6, FR of 15, RA of 93.6% and RT of 81.667s. LBP yielded FA of 18, FR of 32, RA of 84.75% and RT of 101.221s. Furthermore, GWT produced FA of 12, FR of 26, RA of 88.41% and RT of 112.692s. PCA-LDA produced FA of 22, FR of 38, RA of 81.71% and RT of 151.421s. However, HoG yielded FA of 21, FR of 27, RA of 86.92 and RT of 124.533s. The summary of the implementation results is presented in Table 1.

Table 1: Evaluation Results of the Aging-Invariant Feature Extraction Techniques

FET	FA	FR	RA (%)	RT (s)
Baseline Techniques				
LBP	18	32	84.75	101.221
GWT	12	26	88.41	112.692
Emerging Techniques				
LBP-GWT	6	15	93.6	81.667
PCA-LDA	22	38	81.71	151.421
HoG	21	27	86.92	124.533

5.2 Results and Discussion on Statistical Significance Analysis

The results of one-way ANOVA obtained from the quantitative results of the AI-FET are presented in Figs, 3-6. In Fig. 3, it is shown that the F-statistics and F-critical values for baseline and emerging techniques are 0.0358 and 10.128 respectively with FAR values. Similarly, F-statistics and F-critical values for baseline and emerging techniques are 0.2126 and 10.128 respectively with RT values. While analysing RA values, 0.0319 and 10.128 were obtained as the F-statistics and F-critical values for baseline and emerging techniques respectively.

Furthermore, F-statistics and F-critical values for baseline and emerging techniques obtained using FRR values are 0.06934 and 10.128 respectively. In all the statistical evaluations conducted at 0.05 critical significance level, the F-critical values were found to be greater than the value of the calculated F-statistics. Hence, since the one-way ANOVA did not return a statistically significant result (f > f crit), the alternative hypothesis (HA) that there are at least two group means that are statistically significantly different from each other is rejected. This implies that the results of the emerging AI-FET techniques are not statistically significantly different from those of the baseline techniques.

Anova: Single Factor — FAR

SUMMARY

Groups	Count	Sum	Average	Variance
BT	2	30	15	18
ET	3	49	16.33333333	80.33333

ANOVA

Source of Variation	SS	df	MS	F	P-value	F crit
Between Groups	2.133333	1	2.133333333	0.035821	0.861967	10.12796
Within Groups	178.6667	3	59.55555556			
Total	180.8	4				

Fig. 3: One-way ANOVA result for Benchmark and Emerging Techniques on FAR

Anova: Single Factor — RT

SUMMARY

Groups	Count	Sum	Average	Variance
BT	2	213.913	106.9565	65.79192
ET	3	357.621	119.207	1237.68

ANOVA

ce of Varic	SS	df	MS	F	P-value	F crit
Between (	180.0897	1	180.0897	0.212608	0.676104	10.12796
Within Gr	2541.152	3	847.0505			
Total	2721.241	4				

Fig. 4: One-way ANOVA result for Benchmark and Emerging Techniques on RT

Anova: Single Factor — RA

SUMMARY

Groups	Count	Sum	Average	Variance
BT	2	173.16	86.58	6.6978
ET	3	262.23	87.41	35.5231

ANOVA

ource of Variatio	SS	df	MS	F	P-value	F crit
Between Group	0.82668	1	0.82668	0.0319	0.869628	10.12796
Within Groups	77.744	3	25.91467			
Total	78.57068	4				

Fig. 5: One-way ANOVA result for Benchmark and Emerging Techniques on RA

Anova: Single Factor — FRR

SUMMARY

Groups	Count	Sum	Average	Variance
BT	2	58	29	18
ET	3	80	26.66667	132.3333

ANOVA

ce of Varic	SS	df	MS	F	P-value	F crit
Between (	6.533333	1	6.533333	0.06934	0.809351	10.12796
Within Gr	282.6667	3	94.22222			
Total	289.2	4				

Fig. 6: One-way ANOVA result for Benchmark and Emerging Techniques on FRR

6 CONCLUSION

In this paper, the statistical evaluation of LBP-GWT, HOG, PCA-LDA and LBP-GWT was conducted and benchmarked with baseline AI-FET which are LBP and GWT. The results of the emerging techniques are only quantitatively but not statistically significantly different from those of the baseline algorithms. However, further work could be directed towards the statistical evaluation of the complexity of these AI-FET using Halstead measure, Lines of Code (LOC), cyclomatic complexity measures among others.

REFERENCES

Ahonen T. and Pietikäinen M. (2007): "Soft Histograms for Local Binary Patterns", Finnish Signal Processing Symposium, Oulu, Finland. pp.107-117.

Anila S., Bekios C., Buenaposada J. and Baumela L. (2011): "Linear Discriminant Techniques in Gender Recognition", In Proceedings of the IEEE Transactions on Pattern Analysis and Machine Intelligence, 33 (4): pp. 858 – 864.

Dihong G., Zhifeng L., Dahua L., Jianzhuang L. and Xiaoou T. (2013): "Hidden Factor Analysis for Age Invariant Face Recognition", IEEE ICCV Proceedings, pp. 2872-2879.

Kepenekci B. (2001): "Face recognition using Gabor Wavelet Transformation", PhD Thesis, The Middle East Technical University.

Kuldeep Singh Sodhi and Madan Lal (2013): "Face Recognition Using PCA, LDA and Various Distance Classifiers", Journal of Global Research in Computer Science, 4(3): 84-92.

Narayanan R. and Rama C. (2006): "Face Verification across Age Progression", IEEE Transactions on Image Processing, 15 (11), pp. 3349–3361.

Oloyede Ayodele, Fagbola Temitayo, Olabiyisi Stephen, Omidiora Elijah and Oladosu John (2016): Development of a Modified Local Binary Pattern-Gabor Wavelet Transform Aging Invariant Face Recognition System. In Proceedings of ACM International Conference on Computing Research & Innovations, University of Ibadan, Nigeria, pp. 108-114, 7-9 September 2016.

Ojala T., Pietikainen M. and Maenpaa T. (2002): "Multiresolution Gray-Scale and Rotation Invariant Texture Classification with Local Binary Patterns", IEEE Transaction on Pattern Analysis and Machine Intelligence, 24 (7): pp. 971–987.

Shen C. and Bai S. (2006): "Dynamic Facial Expression Recognition Using a Bayesian Temporal Manifold Model", In British Machine Vision Conference (BMVC), Edinburgh, UK, 1: pp. 297–306.

Shinde P.V. and Gunjal B.L. (2012): "Particle Swarm Optimization - Best Feature Selection method for Face Images", International Journal of Scientific & Engineering Research, 3(8): pp. 1-5.

Zhang W., Shan S., Chen X. and Gao W. (2005): "Local Gabor Binary Patterns Based on Kullback-Leibler Divergence for Partially Occluded Face", IEEE Signal Processing Letter, 14(11): 875–878.