Review of Technical Approaches to Face Recognition in Unconstrained Scenes with Varying Pose and Illumination

[1*] Temitayo M. Fagbola, [2]Stephen O. Olabiyisi, [2]Funmilola I. Egbetola and [2]Ayodele Oloyede

[1]Department of Computer Science, Federal University Oye-Ekiti, Ekiti State, Nigeria
[2]Department of Computer Science and Engineering, Ladoke Akintola University Technology, Ogbomoso, Oyo state, Nigeria
temitayo.fagbola@fuoye.edu.ng|soolabiyisi@lautech.edu.ng|efcutieike@gmail.com|deledeee@yahoo.com

Abstract— Difficulties inherent in human face recognition from unconstrained and motion characterized scenes are usually accounted for due to varying illumination and pose of subjects in the scenes. However, a number of technical approaches have been developed to manage these nuisance factors to make recognition possible and optimal. In this paper, a review of some technical approaches to face recognition challenged with varying illumination and pose is conducted. The logical soundness of each approach and limitations are investigated while a basis for a unified, more efficient and technically optimal approach is established.

Keywords— Pose, Illumination, Face Recognition, Unconstrained Scene

1 Introduction

Face recognition from unconstrained scenes has been an active area of research lately due to the massive influx of video surveillance systems (VSS) and other ubiquitous hand-held video capturing devices. This subject is simply regarded as a technique of differentiating faces or confirming one or more individuals in a particular frame or video image using a stored database (Raghu, 2012; GenciCapi, 2009; Majumdar and Ward, 2008; Wu-Jun, Chong-Jun, Dian-Xiang and Shi-Fu, 2004). In unconstrained scenes, there exists no cooperation between the subject being captured and the face recognition system. This makes face recognition in such situation highly difficult and almost unrealistic especially if the resolution of the video is very low or the scene environment is very dark.

Face recognition is such a challenging but yet an interesting problem that is intertwined with pattern recognition, data mining, artificial intelligence, neural networks, computer vision and computer graphics among others (Rabia and Hamid, 2009). It finds numerous applications in surveillance, human-computer interactions, authentication and security (Rehab, Rabab and Rawya, 2008). For example, deployment of VSS has migrated from the conventional use majorly in industrial, governmental and military zones as security systems and now to homes as personal security hardware. This technological revolution resulted in increased processing power, speed, portability, fault tolerance capability, more sophisticated video graphics processing unit (GPU) and other related security hardware breakthroughs which also come at reduced costs. Although, other reliable methods of biometric personal identification exist; for example, fingerprint analysis or iris scans, these methods inherently rely on the cooperation of the participants, whereas a personal identification system based on analysis of frontal or profile images of the face is often effective without the participant's cooperation or intervention (Raghu, 2012). Compared to traditional face recognition in still images, video based face recognition has a great number of advantages (Jeremiah, Kevin, Patrick and Soma, 2012). Firstly, videos contain more abundant information than a single image (Shreekumar and Nagaratna, 2011). As a result, more robust and stable recognition can be achieved by fusing information of multi frames. Secondly, temporal information becomes available to be exploited in videos to improve the accuracy of face recognition.

* Corresponding Author

Finally, multi-poses of faces in videos make it possible to explore shape information of face and combined into the framework of face recognition. A video-based face recognition system typically consists of three modules: one for detecting the face; a second one for tracking it; and a third one for recognizing it (Nitin and Ranjit, 2016). In practice, it is difficult to perform fully-automatic recognition of individuals under surveillance using commercial systems. State-of-the-art systems applied to video-based face recognition often perform poorly in practice due to complex environments that change during the video capturing process and facial models that are designed using limited data and knowledge of individuals during a preliminary enrolment process (Mahesh and Nagaratna, 2013; Jeremiah *et al.*, 2012). Evidently, human recognition abilities have superseded the operations of most public security organizations due to the badvisual quality, low resolution, pose and illumination differences of facial images typically captured by their cameras.

The 2002 Face Recognition Vendor Test (FRVT) report contains an evaluation that compares face recognition from still images to recognition from videos (Phillips, Grother, Micheals, Blackburn and Tabassi, 2003). In an experiment conducted on a face database comprised by still images of 63 subjects, a number of commercial recognition systems for image based face recognition performed worse when tasked with the identification of faces from videos. The still images from the database contained frontal face views while the videos displayed speaking subjects with varying expressions. The faces in the videos appeared significantly different from those in the database images causing a large number of recognition errors.The Multiple Biometric Grand Challenge featured a problem involving face recognition from videos in which illumination, movement and head pose were not controlled (Phillips, 2008). The video dataset included high resolution (1440x1080) and standard resolution (720x480) sequences with subjects walking toward the camera. Out of four state-of-the-art commercial face recognition algorithms, the best performers on the high and standard resolution videos only reached about 70% and 40% verification rates respectively. All algorithms performed significantly better on the high resolution videos. Resolution dependent performance differences notwithstanding, off-frontal poses and changing illuminations were observed to play a large role in the poor verification performance of all of the systems (Jeremiah *et al*., 2012).

In principle, even the best solution to face recognition in surveillance videos at real-time is highly unacceptable. This is due to the ill-posed nature of the video challenges. Usually, there is a trade-off between these requirements as observed from existing solutions. However, there have been active contributions from researchers in related areas including pattern recognition, machine learning, biometrics system, image segmentation, scene detection and analysis, processing, video analysis and processing, mathematical modeling, statistics, computer vision, image resolution reconstruction and video enhancement to address these technical challenges. In this paper, technical approaches adopted by researchers from these inter-related research domains to addressing face recognition from unconstrained videos in the face of pose and illumination challenges are reviewed. Furthermore, some technical limitations of these approaches were also identified. In the concluding part, recommendation on the identified technical limitations as areas for further research was made.

2 CHALLENGES OF VIDEOS FOR FACE RECOGNITION

A number of video-oriented challenges can arise in unconstrained video face recognition applications but the two (2) most challenging ones are pose and illumination variations (Jeremiah *et al.*, 2012). Figure 1 illustrates the major challenges of face recognition in videos. These challenges may cause differences in appearance between distinct shots of the same person to be greater than those between two people viewed under similar conditions (Kavita, Mukesh and Mukesh, 2010). Although, pose and illumination are traditionally regarded as the two (2) most challenging nuisance factors (Jeremiah *et al.*, 2012; Zhao *et al.*, 2003), Phillips, Beveridge, Draper, Givens and Weimer (2011) presented evidence that some factors including expression variation, occlusion and motion blur have nearly as significant of an impact on face recognition performance in uncontrolled contexts.

Fig. 1. Challenges of Face Recognition in Videos (Kavita et al., 2010)

2.1 Pose Variation

Uncontrolled cameras can record non-ideal face shots from a variety of angles, causing the correspondences between pixel locations and points on the face to differ from image to image. Same face appears differently due to changes in viewing condition. This situation is presented in Figure 2 with faces of the same individual under varying pose. However, pose-invariant recognition capability is crucial for accurate face recognition in video sequences because in general, it is difficult or almost impossible to control the imaging direction when capturing videos containing human faces in real time. Broadly speaking, there are three classes of algorithms aimed at achieving pose invariance (Kavita*et al.,* 2010). The first, a model-based approach, uses an explicit 2D or 3D model of the face and attempts to estimate the parameters of the model from the input (Blanz and Vetter, 2003). This is a view-independent representation.

A second class of algorithms consists of global parametric models such as the eigenspace method (Murase and Nayal, 1995) that estimates a single parametric (typically linear) subspace from all the views for all the objects. In comparative face recognition evaluation trials, such methods are usually outperformed by methods from the third class known as view-based techniques.

Fig. 3.Images under Varying Pose (Kavita et al., 2010)

2.2 Illumination Variation

Illumination problem arises due to uneven lightning on faces (Kavita *et al.*, 2010). A typical illustration of the illumination problem is presented in Figure 2. This uneven lightning brings variations in illumination and affects the recognition process greatly since the facial features that will be used for classification gets affected by this variation. However, all the approaches towards illumination problem can be broadly categorized based on transformation of images with variable illumination to a canonical representation; illumination invariant feature extraction; illumination variation modeling and 3D face model utilization.

Fig.2.The Illumination Problem (Kavitaet al., 2010)

The categories are explicitly discussed as follows (Kavita *et al.*, 2010):

a. *Transformation of images with variable illumination to a canonical representation*:

With this method, images with varying illumination are being transformed to their canonical representations. Eigenfaces have been the widely adopted technique for the detection of unknown faces in an image using this method. Arandjelovic, Shakhnarovich and Cipolla (2009) proposed simple image filtering techniques for rapid recognition under varying illumination and pose. This technique has demonstrated a reduction of 50–75% in recognition error rates, with a recognition rate of 98% of the individuals.

b. Features that are invariant to light changes

This is done by extracting only those features that are not affected by variations in lighting conditions. A few of such representation of image are gradient faces (Zhang, Wu, Yang and Zhang, 2009), 2D Gabor Filter (Lades, Vorbruggen, Buhmann, Lange and Konen, 1993), DCT coefficients (Shim, Moon and Han, 2008; Chen, Er and Wu, 2006) and LBP Feature (Tan and Triggs, 2007).

c. *Modeling of illumination variation*:

With this approach, a small number of training images are used to synthesize novel images under changes in lighting and viewpoint conditions (Kavita *et al*, 2010). However, because the space of lighting conditions is infinitely dimensional, sampling this space is no small task. This can be simplified by a convex cone termed as illumination cone formed from the set of images of an object in fixed pose but under all possible illumination conditions (Georghiades *et al.*, 1998). This illumination cone can be well approximated by a low-dimensional linear subspace. Under variable lighting, the set of images is characterized by a family of illumination cones parameterized by the pose. The illumination cones for non-frontal poses can be constructed by applying an image warp on the extreme rays defining the frontal cone.

d. *Utilization of some 3D face models:*

With this method, the facial shapes are obtained in advance. A 3D morphable model is used to generate 3D face models from three input images from each person in the training database. Thus, the 3D models are rendered under varying pose and illumination conditions to build a large set of synthetic images. In all the recent approaches, a 3D model of a face is utilized to transform the input image into the same pose as the stored prototypical faces, and then, direct template matching is used to recognize faces (Zhao *et al.*, 1999; Vetter and Poggio, 1997). Although approaches based on 3D face models are robust to uncontrolled conditions, they still suffer some drawbacks including a high computational cost compared to 2D approaches and the accuracy is much more dependent on the number and quality of the selected features. It is important to mention that a 3D model is difficult to be constructed from low-resolution inputs.

3 TECHNICAL APPROACHES TO POSE AND ILLUMINATION INVARIANT FACE RECOGNITION IN UNCONSTRAINED SCENES

Several technical approaches have been developed to address face recognition along illumination and pose differences of same subject in unconstrained scenes. Passive approaches (model based methods, subspace-based statistical methods, illumination invariant representation and other handling methods) have been widely adopted to address face appearances modified by extreme varying illumination changes (Ranji, 2010). Principal Component Analysis (PCA) (Omidiora, 2006), ICA (Wang, Zheng-Ming and Long, 2013), LDA (Anila, Bekios, Buenaposada and Baumela, 2011) are among the subspace-based statistical methods most commonly used. Some modified version of these subspace algorithms include Gabor Wavelet based Modular PCA (GW-MPCA) (Gudur and Asari, 2006), Adapted PCA (APCA) (Chen, Lovell and Shan, 2009), among others.

A number of spatial frequency techniques, including Fourier transform (Hwang *et al.*, 2006; Lai *et al.*, 2001), Discrete Cosine Transform (DCT) (Aman, Pallavi, Roja, 2011) and Discrete Wavelet Transform (DWT) (Hafed and Levine, 2001), have also been widely adopted and extended. For example, due to the fact that DWT could only extract features in the low frequency region only, Complex Wavelet Packet Transform (CWPT), an extension of DWT, was developed by Adinarayana, Sreenivasa and

Durga (2016) to address this limitation. Spatial frequency techniques transform face images to the frequency domain and only the coefficients in the low-frequency band are reserved for face (Unsang, 2009). Local feature extraction techniques and their variants have also been applied to managing pose and illumination challenges. These techniques include Local Feature Analysis (LFA), Local Preserving Projection (LPP), Local Binary Pattern (LBP) (Timo, Abdenour and Matti, 2004), Elastic Bunch Graph Matching (EBGM), Gabor Wavelet Transform (GWT) (Kepenekci, 2001), Speeded-Up Robust Feature (SURF)algorithm and Color Local Binary Pattern from Mutually Independent Color Channels (Gholamreza, 2013) among others.

An improvement over the fully automatic pose-invariant face recognition via 3D pose normalization (FAP-3D Norm) proposed by Asthana, Marks, Jones, Tieu and Rohith (2011) and a pose robust face recognition system (PRFRS) proposed by Yi, Lei and Li (2013), which both suffer from landmark ambiguity, is a LBP that is localized via deformation components such that a face pose estimation procedure is able to select stable and visible landmarks(Iacopo, Claudio, Alberto and Gerard, 2016). As observed with PRFRS, a 3D Morphable Model (3DMM) was also introduced efficiently to the image using a Ridge regression solution that globally preserves the shape of the face while minimizing the landmark re-projection error locally. As a modification over FAP-3D Norm and PRFRS, LBP histograms were localized on the deformed vertices instead of computing LBP on a uniform grid in order to obtain feature vectors with same dimension irrespective of the size of the image.

This is beneficial for the following reasons (a) the feature vector becomes independent of the image size; (b) the LBP are more efficiently localized since the deformed model has been optimized to fit the face; (c)makes the restriction of the feature vector to those parts which are not visible considering a self-occluding face much easier. Nearest Neighbor (NN) classifier was used to select the closest feature from the counterparts in the gallery. Cumulative Matching Characteristic (CMS) curves and the normalized Area Under the Curve (nAUC) are the evaluation metrics used. For poses at {-30,0,+45} degree, the developed approach performs better than some state-of-the-art approaches including 3DMM. As stated, 3DMM can yield appreciable results for illumination invariant face recognition. However, the high dense nature of 3DMM makes fitting the model incur high computational effort which is not suitable for real-time face recognition systems (Fatih, Binnur and Muhittin, 2010).

Hierarchical ensemble of global and local facial features for recognition (Yu, Shiguang, Xilin and Wen, 2009) was developed to handle pose and illumination challenges in video sequences. In the proposed method, the global features are extracted from the whole face images using Fourier transform and the local features are emphasized on some spatially divided face patches using Gabor wavelets. The position and size of the patches are learned from a training data via greedy search. The hierarchical ensemble classifier is formed by weighted sum of the component classifiers which are all fisher linear discriminants on either global or local features. Experimental results on both FERET and FRGC version 2.0 databases show that the ensemble classifier outperforms other competitors. The solution is limited by the use of still images for testing; it fails when implemented for use in practice with+ online video feeds. As an improvement over hierarchical ensemble of Fourier transform and Gabor wavelet features, Discrete Fourier Transform (DFT) and Gabor Wavelet Transform (GWT) feature vectors were fused using Region-Based Image Fusion by Anila and Devarajan (2011).The composite feature set helped to increase the recognition rate significantly. The classification errors were reduced using Fisher Linear Discriminant (FLD). The method of correlation coefficient was used for matching the test image with the database. Though, it performs fairly well handling illumination challenges, it is still very sensitive to pose variations outside the region of the image fused.

Combined Global Local Preserving Projections (CGLPP) was proposed by Nisar (2012) with the aim to integrate the advantages of Principal Component Analysis (PCA) and Local Preserving Projection (LPP), a global and local feature extraction technique, respectively. However, there are a lot of problems associated with this approach. There is no external criterion variable such as group membership against which to test the solution. A second problem is that, after extraction, there is an infinite number of rotations available, all accounting for the same amount of variance in the original data. It tends to be more computationally demanding than PCA and also sensitive to the scaling of the variables. To address this problem, gradient faces method was proposed by Raghu (2012). The gradient faces method is an image preprocessing technique for face recognition under varying lighting. This method transforms image into the gradient domain, then an illumination insensitive measure is extracted for recognition of faces. The high recognition rates achieved by gradient faces show that gradient faces is an effective method for solving illumination problem and robust to different lighting and noise. This method is highly computationally expensive and cannot be implemented for practical use.

Machine learning approaches like Support Vector Machine (SVM) and Nearest Neighbor use features that are insensitive to pose, scale, illumination, rotation, occlusion and expression to address pose and illumination challenges in unconstrained scenes. A local geometric feature and SVM-based 3D face recognition approach was proposed by Yinjie *et al*. (2013). This approach was tested on FRGC v2.0 and BU-3DFE datasets through a number of experiments and a high recognition performance was achieved. Consequently, this method is only suited for small dataset; suffers high computational cost and the accuracy is much more dependent on number and quality of the selected features. A Gabor featured statistical model for feature extraction in chaotic database was developed by Kumar *et al.* (2013). Gabor wavelet was used for the preprocessing of the human face image. LDA was applied on the reduced features from PCA to get more discriminating features. The classification was done using distance measure classifiers and SVM. However, the developed system suffers from curse of dimensionality and is highly computationally expensive.

The Gabor featured statistical model was extended by combining improved PCA and LDA for a more efficient feature extraction process suitable for video sequences (Zhao *et al.*, 2013). Recognition was performed using the Euclidean distance to compute distances between the target face and gallery video sequences. The resultant solution however shows considerable improvement over PCA and LDA, however, it cannot cope with large variations in pose and illumination of faces in videos. The method was tested on controlled datasets that do not address the difficulties in real surveillance video data. A Double-Density Dual-Tree Complex Wavelet Transform (DD-DTCWT) filtering technique is an improvement over the combined Discrete Fourier transform (DFT) and Gabor Wavelet Transform (GWT) technique. It was utilized by Aryaz *et al.* (2013) for mask extraction. After extracting a mask, feature vector is formed and the principal component analysis (PCA) is used for dimensionality reduction which is then fed to the extreme learning machine (ELM) as a classifier to evaluate the performance of the proposed algorithm. Though, the framework yielded better performance when compared with DD-DTCWT; however, it only performs excellently well at some pose angles and illumination levels, it is very complex and expensive to implement into a video-based face recognition system.

The use of Oriented Phase Congruency (OPC) image features was proposed to improve face recognition rate against illumination variations (Hendra, Wirawan and Adi, 2016). It is an extension of the two-dimensional Gabor Phase Congruency Model (2D-GPCM) developed by Kovesi (1999) to achieve a robust edge detection of images. A major advantage of phase congruency is the provision of a measure that does not depend on the overall magnitude of the signal which makes it insensitive to illumination variations. In 2010, Struc and Pavesic modified 2D-GPCM such that the phase deviation measure and phase angle of the Gabor filter are independent of the magnitude of the filter responses. This concept was termed Oriented Gabor phase congruency patterns (OGPCP). As claimed, the independence of the phase changes of the Gabor filter and magnitude of the response accounted for the highly suitable of OGPCP for illumination invariant image representation. As presented in Figure 4, OPC illumination invariant face recognition system employs OGPCP to obtain feature vector sets. The high dimensional OGPCP features were reduced using subspace LDA. The final classification was carried out using nearest neighbor technique and cosine distance measure. Yale-B database with 5760 frontal face images was used for testing. The rank-one recognition rate (R1-RR), determined as change between the number of tested images that are correctly identified and the total number of test images, was used as evaluation metric. R1-RR was determined with varying value of filter OGPC scales.

In another approach, a fuzzy-bat clustering based pose and illumination invariant technique proposed by Kannan

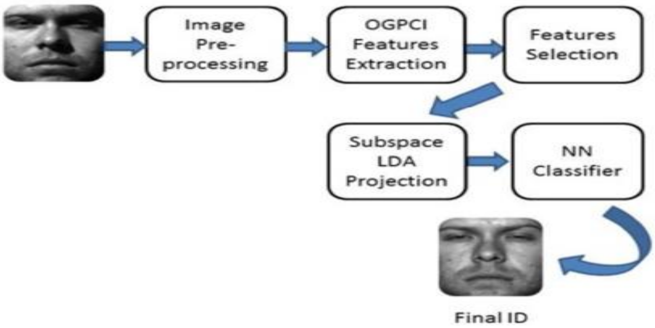

Fig. 4.Oriented phase congruency Features–based Illumination Invariant Face Recognition System (Hendra et al, 2016)

and Jana (2016), feature vectors were extracted using (AAM), Speeded-Up Robust Feature (SURF) algorithm and weighted holoentropy fused together. Adaptive appearance model tracking models like the Active Appearance Model (AAM) uses the coordinates of feature points and the difference between the immediate estimate of an image appearance and its target counterpart for optimizing the process of matching objects based on appearance and shape (Bay, Ess, Tuytelaars and Gool, 2008). SURF algorithm is an excellent interest point detector based on Hessian matrix measure and a Haar wavelet response distribution based descriptor. It is invariant to displacement, noise, rotation, pose and illumination and often believed to be more superior to other descriptors such as SIFT with respect to distinctiveness, repeatability, robustness and computation speed (Raymer, Punch, Goodman, Kuhn and Jain, 2000).Since videos are characterized with additional relevant information than still images, holoentropy is a suitable computationally-efficient technique for the detection of redundant (outlier) frame(s) among a large number of frames with unique features. By definition, given a dataset of n objects $\{x_1, x_2, x_3, \ldots, x_n\}$ with each object representing a $m \times 1$ categorical attribute vector$Y = [y_1, y_2, y_3, \ldots, y_m]^T$,entropy of Y is given by:

$$H_x(Y) = H_x(y_1) + H_x(y_2|y_1) + , \ldots, + H_x(y_m|y_{m-1,\ldots,}y_1) \quad (1)$$

Holoentropy is expressed as the sum of the entropies on all attributes (Cootes, Taylor, Cooper and Graham, 1995) such that:

$$HL_x(Y) = H_x(Y) + C_x(Y) = \sum_{i=1}^{m} H_x(y_i) \quad (2)$$

Where $C_x(Y)$ is the total correlation of random vector Y. With weighted holoentropy, weights which are obtained directly from the data are assigned to each attribute using inverse sigmoid function as given by:

$$W_x(y_i) = 2\left(1 - \frac{1}{1+\exp(-H_x(y_i))}\right) \quad (3)$$

The weighted holoentropy of the random vector Y is obtained by adding up the weighted entropies of the individual attributes of Y stated mathematically as:

$$W_x(Y) = \sum_{i=1}^{m} W_x(y_i)\, H_x(y_i) \quad (4)$$

For each bat, an initial value of frequency between A_{min} and A_{max} is assigned. The minimum and maximum value of the frequency is determined by the size of the domain of the problem in consideration.

$$A_i = A_{min} + (A_{max} - A_{min}) \times \delta \quad (5)$$

$$V_i^t = V_i^{t-1} + (X_i^t - X_{best}^t) \quad (6)$$

$$X_i^t = X_i^{t-1} + V_i^t \quad (7)$$

Equations (5), (6) and (7) are the update equations for the bat's frequencies A_i, velocity V_i and position X_i respectively in a *d*-dimensional search space.

The value of δranges between 0 and 1. The value X_{best}^t represents the current best location at time t after comparison with all locations of all *n* bats. Since velocity is given by *V* = *Aλ*where λis the wavelength at which the waves are emitted, the increment in velocity is based on varying either *A* or *λ*keeping the other value fixed.

Fuzzy-Bat algorithm according to Kannan and Jana (2016) is as follows:

1. *Initialize the parameters including population size, frequencies, number of generations and dimension.*
2. *Create a swarm with d bats.*
3. *Assume an initial best solution.*
4. *Calculate the fitness value of each bat.*
5. *Update the frequency for each bat. (5)*
6. *Update the velocity matrix for each bat. (6)*
7. *Update the location vector for each bat. (7)*
8. *Calculate the (global) best and (local) personal best location for all the bats.*
9. *If terminating condition is not met, go to step 4.*
10. *Update the location using the output of Fuzzy inference system.*

Honda UCSD Video database was used to test the performance of the Fuzzy-Bat clustering algorithm. Parameters for the fuzzy-bat algorithm (population size, number of generations, loudness, pulse rate, minimum and maximum frequencies and dimension of search variables) were set with values (10, 100, 0.5, 0.5, 0, 2 and 629) respectively. Mathews Correlation Coefficient (MCC), Accuracy and Specificity are the performance evaluation metrics used. MCC is a correlation coefficient between the observed and predicted binary classification and ranges between -1 and +1. A value of +1 indicates a perfect prediction and a value of -1 denotes total deviation between the predicted and observed value (Kannan and Jana, 2016).

$$MCC = \frac{TP \times TN - FP \times FN}{\sqrt{(TP+FP)(TP+FN)(TN+FP)(TN+FN)}} \quad (8)$$

Specificity is a measure of proportion of negatives that are correctly identified as negatives.

$$SPEC = \frac{TN}{TN+FP} \quad (9)$$

Accuracy represents the number of faces that are correctly recognized from the total number of faces in the gallery (Jeremiah *et al.*, 2012).

$$ACC = \frac{TP+TN}{TP+TN+FP+FN} \quad (10)$$

where TP = True Positive, TN = True Negative, FP = False Positive, FN = False Negative.

Summarily, the fuzzy-bat algorithm outperforms fuzzy rule based and fuzzy C-Means algorithm. In the same vein, evaluation using squared Euclidean distance measure yields higher accuracy than Euclidean distance.

Some recent approaches to pose and illumination invariant face recognition include Logarithmic Non-Subsampled Contourlet Transform (LNSCT) which is a fully shift-invariant, multi-scale and multi-direction transform (Xiaohua, Jianhuang and Wei-Shi, 2010); gallery scaling (Kavitha and Mirnalinee, 2016); multi-class SVM and affine transformation with DCT (Shilpa and Shreekumar, 2016); use of DCT coefficients with Contrast Limited Adaptive Histogram Equalization (CLAHE)(Thamizharasi and Jayasudha, 2016a) and CLAHE with thresholding (Thamizharasi and Jayasudha, 2016b). Others include template matching and Sequential Karhunen–Loeve (SKL) method with a particle filter-based face tracker (Ali et al, 2016); Overlapping Local Phase Feature (OLPF) with an Adaptive Gaussian Mixture Model (AGMM) (Qiang, Wei, Hongliang and King, 2016) and a logarithmic fractal dimension-based complete eight local directional patterns with adaptive homomorphic filtering (Mohammad and Xiaojun, 2016).

4 TECHNICAL LIMITATIONS OF THE APPROACHES TO POSE AND ILLUMINATION INVARIANT FACE RECOGNITION

Most promising approaches to pose and illumination invariant face recognition reviewed still suffer from a number of major challenges limiting their implementation and use in practice. Thus, face recognition in unconstrained scenes remains an open problem domain. Summarily, the following are the major technical limitations of these approaches.

(i) The pose and illumination invariant models were not integrated directly into the face recognition module in such a way to manage the time complexities inherent in merging and maintaining separate solutions for pose and illumination variations and face recognition in unconstrained scenes.

(ii) Software complexity metrics like the Halstead software complexity measure was not adopted to evaluate the soft complexities and performance of the pose and illumination invariant face recognition systems. Emerging similarity and dissimilarity metrics were notused to test the performance of the feature descriptors on pose and illumination video datasets.

(iii) A linear model that can establish a direct relationship amongpose, illumination and the rate of recognition of the face recognition system was not formulated to help guide the pose and illumination invariant recognition of faces in a more direct manner.

(iv) Most of the approaches are computationally very expensive to implement and use in practice; they suffer from high computational cost and the accuracy is much more dependent on number and quality of the selected facial features.

5 CONCLUSION AND RECOMMENDATION

In this paper, a review of technical approaches to face recognition with varying poses and illumination of subjects in unconstrained scenes was conducted. Extrinsic factors including pose and illumination variation in unconstrained scenes have been identified as detrimental to the performance efficiency and optimality of face recognition systems. Technical limitations of those approaches indicate that face recognition in unconstrained video sequences remains an open problem domain. Therefore, this paper recommends that these set of limitations be further investigated and improved upon in order to realize the sought-after capabilities of pose and illumination invariant face recognition systems. The improved solution slightly anticipated for are expected to serve as more technically optimal approach which will in turn assist to improve the performance of face recognition systems adopting such solutions in practice.

REFERENCES

Adinarayana, Sreenivasa and Durga (2016), Illumination invariance face recognition using complex wavelet packet transform" International Journal of Engineering Development and Research, Vol. 4, No. 1, pp. 525-528.

Ali AkberDewan M., Granger E., Marcialis G.-L., Sabourin R. and Roli F. (2016), Adaptive appearance model tracking for still-to-video face recognition, Pattern Recognition, Vol. 49, pp. 129–151.

Anila S. and Devarajan N. (2011), Global and Local Classifiers for Face Recognition, European Journal of Scientific Research, ISSN 1450-216X, Vol. 57, No. 4, pp. 556-566.

Arandjelovic O, Shakhnarovich G. and Cipolla R (2009), A Pose-Wise Linear Illumination Manifold Model for Face Recognition Using Video, Computer Vision and Image Understanding, Vol 113, No 1, pp. 113-125.

Arandjelovic O. and Cipolla R. (2004), An Illumination Invariant Face Recognition System For Access Control Using Video, in Proceedings of the British Machine Vision Conference (BMVC '04), Kingston, Canada, pp. 537-546.

AryazBaradarani, Jonathan Wu and Majid Ahmadi (2013), An Efficient Illumination Invariant Face Recognition Framework via Illumination Enhancement and DD-DTCWT Filtering, Pattern Recognition, Vol. 46, pp. 57–72.

Asthana A., T. Marks, M. Jones, K. Tieu, and M. Rohith (2011), Fully Automatic Pose-Invariant Face Recognition via 3D Pose Normalization, in International Conference on Computer Vision.

Bay H., Ess A., Tuytelaars T., and Gool L. (2008), Speeded-Up Robust Features (SURF), Computer Vision and Image Understanding, Vol 110, No. 3, pp. 346-359.

Blanz V. and Vetter T. (2003), A Morphable Model for the Syntheses of 3D faces, In Proceedings of International Conference on Computer Graphics, pp. 187-194.

Chen S., B. C. Lovell, and T. Shan (2009), Robust Adapted Principal Component Analysis For Face Recognition," International Journal of Pattern Recognition and Artificial Intelligence, Vol. 23, No. 3, pp. 491-520.

Chen W., Er M. J. and Wu S (2006), Illumination Compensation And Normalization For Robust Face Recognition Using Discrete Cosine Transform In Logarithm Domain, IEEE Transactions on Systems, Man and Cybernetics, Vol 36, No 2,pp. 458-466.

Cootes T. F., C. J. Taylor, D. H. Cooper, and J. Graham (1995), Active Shape Models – Their Training and Application. Computer Vision, Graphics and Image Understanding, Vol 61, No. 1, pp: 38–59.

FatihKahraman, Binnur Kurt and MuhittinGokmen (2010). Robust Face Alignment for Illumination and Pose Invariant Face Recognition, Face Recognition, Milos Oravec (Ed.), ISBN: 978-953-307-060-5, InTech.

GenciCapi (2009), A Robotic System for Intelligent Real Time Face Recognition, ICGST International Journal on Automation, Robotics and Autonomous Systems, Vol. 9, No 1, pp. 25-29.

Georghiades A., Kriegman D. and Belhumeur P. (1998), From Few to Many: Generative Models for Recognition under Variable Pose and Illumination, IEEE Transactions Pattern Analysis and Machine Intelligence, Vol. 40, pp. 643-660.

GholamrezaAnbarjafari (2013), Face Recognition Using Color Local Binary Pattern from Mutually Independent Color Channels, EURASIP Journal on Image and Video Processing, Vol. 6, pp. 1-11.

Gudur N. and V. Asari (2006), Gabor Wavelet Based Modular PCA Approach for Expression and Illumination Invariant Face Recognition, in 35th Applied Imagery and Pattern Recognition Workshop.

Huafeng Wang, Yunhong Wang and Yuan Cao (2009), Video-based Face Recognition: A Survey, World Academy of Science, Engineering and Technology, Vol. 60, pp. 293-302.

IacopoMasi, Claudio Ferrari, Alberto Del Bimbo and Gerard Medioni (2014), Pose Independent Face Recognition by Localizing Local Binary Patterns via Deformation Components, 22nd International Conference on Pattern Recognition, pp. 4477-4482, IEEE Computer Society.

Jeremiah R. Barr, Kevin W. Bowyer, Patrick J. Flynn and Soma Biswas (2012), Face Recognition from Video: A Review, Internatrnal Journal of Pattern Recognition and Artificial Intelligence, World Scientific Publishing Company Vol. 16, No 2, pp. 1-56.

Kannan N. and Jana S. (2016), Analysis of Effect of Optimization Algorithm on Pose and Occlusion Invariant Feature for Face Recognition, Proceedings of 47th IRF International Conference, Pune, India, pp. 47-51.

Kavita R. Singh, Mukesh A. Zaveri and Mukesh M. Raghuwanshi

(2010), Illumination and Pose Invariant Face Recognition: A Technical Review, International Journal of Computer Information Systems and Industrial Management Applications (IJCISIM) http://www.mirlabs.org/ijcisim.

Kavitha. J, Mirnalinee. T.T. (2016), Gallery Scaling for Efficient Pose Invariant Face Recognition, International Journal of Advanced Engineering Technology, Vol. 7, No 2, pp. 1-4.

Lades M., Vorbruggen J. C., Buhmann J., Lange J., Von der Malsburg C., Wurtz R. P. and Konen W. (1993), Distortion invariant Object Recognition In The Dynamic Link Architecture, IEEE Trans. on Computers, Vol. 42, No 3, pp. 300-311.

MageshKumar C., Ragul G. and Thiyagarajan R. (2013), Gabor Featured Statistical Modeling In Face Recognition with Chaotic Database, International Journal of Emerging Technology and Advanced Engineering, Vol. 3, No 1.

Mahesh Prasanna and NagaratnaHegde (2013), A Fast Recognition Method for Pose and Illumination Variant Faces on Video Sequences, IOSR Journal of Computer Engineering (IOSR-JCE), Vol. 10, No. 1, pp. 8-18.

Majumdar and Ward (2008), Pseudo-Fisher face Method for Single Image Per Person Face Recognition, In Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing, Las Vegas, NV, pp. 989-992.

Mohammad Reza Faraji and Xiaojun Qi (2016), Face recognition under varying illuminations using logarithmic fractal dimension-based complete eight local directional patterns, Neurocomputing, Elsevier, Vol. 199, pp. 16-30.

Murase H. and Nayar S. (1995), Visual Learning and Recognition of 3-D Objects from Appearance, International Journal of Computer Vision, Vol. 14, pp. 5-24.

NisarHundewale (2012), Face Recognition Using Combined Global Local Preserving Projections and Compared with Various Methods, International Journal of Scientific & Engineering Research, Vol. 3, No 3, pp. 1-4.

Nitin Sharma and Ranjit Kaur (2016), Review of Face Recognition Techniques, International Journal of Advanced Research in Computer Science and Software Engineering, Vol. 6, No 7, pp 29-37.

Omidiora Elijah Olusayo (2006), A Prototype of Knowledge-based System for a Black Face Recognition System Using Principal Component Analysis and Fisher Discriminant Algorithms, Unpublished Doctoral Dissertation, Department of Computer Science, LadokeAkintola University of Technology, Ogbomoso, Nigeria.

Pentland A., Jebara T., Clarkson B. and Choudury T. (1999), Multimodal Person Recognition Using Unconstrained Audio and Video, IEEE Proceedings on Audio-and Video-based Biometric Person Authentication, pp. 176-181.

Phillips J. (2008), Video Challenge Problem Multiple Biometric Grand Challenge: Preliminary Results of Version 1".

Phillips P. J., Grother P., Micheals R. J., Blackburn D. M., Tabassi E. and Bone J. M. (2003), Face Recognition Vendor Test 2002, Proc. IEEE International Workshop on Analysis and Modeling of Faces and Gestures, pp. 44-48.

Phillips P., Beveridge J., Draper B., Givens G., O'Toole A., Bolme D., Dunlop J., Lui Y. M., Sahibzada H. and Weimer S. (2011), An Introduction to the Good, the Bad, the Ugly Face Recognition Challenge Problem, Proc. of the IEEE Conference on Automatic Face and Gesture Recognition, pp. 346-353.

Qiang Liu, Wei Zhang, Hongliang Li and King NgiNgan (2016), Hybrid human detection and recognition in surveillance, Neurocomputing, Elsevier, Vol. 194, pp. 10–23.

Rabia Jafri and Hamid R. Arabnia (2009), A Survey of Face Recognition Techniques, Journal of Information Processing Systems, Vol. 5, No. 2,pp. 41-67.

Raghu C.N. (2012), Illumination Insensitive Face Recognition Using Gradient Faces": International Journal of Image Processing and Vision Sciences (IJIPVS) Vol 1, No. 1, pp. 38-44.

Raymer M. L., W. F. Punch, E. D. Goodman, L.A. Kuhn, and A. K Jain (2000), Dimensionality Reduction Using Genetic Algorithms, IEEE Trans. Evolutionary Computation, Vol. 4, No. 2, pp. 164-171.

Rehab Abdel-Kader, Rabab Ramadan and RawyaRizk (2008), Rotation Invariant Face Recognition Based on Hybrid LPT/DCT Features, International Journal of Electrical and Computer Engineering, Vol. 3, No. 7,pp. 488-493.

Shim S. Y., Moon J. H. and Han J. K. (2008), PSNR Estimation Scheme Using Coefficient Distribution of Frequency Domain in H.264 Decoder, IET Electronics Letters, Vol. 44, No. 2,pp. 108–109.

Tan X. and Triggs B (2007), Enhanced Local Texture Feature Sets for Face Recognition under Difficult Lighting Conditions, In Proceedings of the IEEE International Workshop on Analysis and Modeling of Faces and Gestures, pp. 168-182.

Thamizharasi A. and Jayasudha J.S. (2016a), An Illumination Invariant Face Recognition by Selection of DCT Coefficients" International Journal of Image Processing (IJIP), Vol. 10, No 1, pp 14-21.

Thamizharasi A. and Jayasudha J.S. (2016b), An Illumination Invariant Face Recognition by Enhanced Contrast Limited Adaptive Histogram Equalization" ICTACT Journal on Image and Video Processing Vol. 6, No 4, pp. 1258-1266.

Vetter T. and Poggio T. (1997), Linear Object Classes and Image Synthesis from A Single Example Image, IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 19, pp. 733-742.

Wang Yu-Yao, Zheng-Ming Li and Long Wang (2013), A Scale Invariant Feature Transform Based Method, Journal of Information Hiding and Multimedia Signal Processing, Ubiquitous International, Vol. 4, No 2.

Wu-Jun Li, Chong-Jun Wang, Dian-Xiang Xu and Shi-Fu Chen (2004), Illumination Invariant Face Recognition Based on Neural Network Ensemble, In Proceedings of 16th IEEE International Conference on Tools with Artificial Intelligence, pp. 486-490.

XiaohuaXie, Jianhuang Lai and Wei-Shi Zheng (2010), Extraction of Illumination Invariant Facial Features from a Single Image using Non-Subsampled Contourlet Transform, Pattern Recognition, Vol. 43, pp. 4177–4189.

Yi D., Z. Lei, and S. Li (2013), Towards Pose Robust Face Recognition, Computer Vision and PatternRecognition, pp. 3539–3545.

Yinjie Lei, Mohammed Bennamoun and Amar A. El-Sallam (2013), An efficient 3D Face Recognition Approach Based on the Fusion of Novel Local Low-Level Features, Pattern Recognition, Vol. 46, pp. 24–37.

Yu Su, Shiguang Shan, Xilin Chen and Wen Gao (2009), Hierarchical Ensemble of Global and Local Classifiers for Face Recognition, IEEE Transactions on Image Processing, Vol. 18, No. 8, pp. 1885-1896.

Zhang F., Wu X., Yang X., Zhang W. and Zhang L. (2009), Robust Color Demosaicking with Adaptation to Varying Spectral Correlations, IEEE Transaction on Image Processing, Vol. 18, No. 12, pp. 2706-2717.

Zhao Qiuyong, Bin Liang and Fu Duan (2013), Combination of Improved PCA and LDA for Video-based Face Recognition, Journal of Computational Information Systems Vol. 9, No. 1, pp. 273–280.

Zhao W., Chellappa R. and Phillips P. J. (1999), Subspace Linear Discriminant Analysis for Face Recognition, Center for Automation Research, University of Maryland, College Park, Technical Report CAR-TR-914.

Zhao W., Chellappa R., Phillips P. J. and Rosenfeld A. (2003), Face Recognition: A Literature Survey, ACM Computing Surveys, Vol. 35, No. 4, pp. 399-458.