International Journal of Engineering Research and Technology. ISSN 0974-3154, Volume 12, Number 12 (2019), pp. 3114-3126

Resolution-aware Ensemble of Pose and Illumination-Invariant Feature Descriptors for Face Identification in Unconstrained Videos

Temitayo Matthew Fagbola[1] and Colin Surendra Thakur[2]

[1,2]*Department of Computer Science, Federal University Oye-Ekiti, Nigeria.*
[2]*KZN e-Skills coLab, Durban University of Technology, South Africa.*
ORCID: 0000-0001-6631-1002 (Temitayo Fagbola)

Abstract

The recurring global security insurgence has posed new opportunities for massive deployment of Video Surveillance Systems (VSS). However, videos captured by such systems suffer characterized Low Resolution (LR), varying illumination and pose challenges of subjects present in the videos. Consequently to manage these limitations, most existing Feature Extraction Techniques (FET) lack support for the limitations inherent in most VSS videos, which accounts for the high computational overhead and low accuracy of most video-based Face Recognition Systems. In this paper, Iterative Back Projection-Maximum A Posteriori (IBP-MAP) resolution reconstruction technique and an ensemble of local and global feature descriptors based on Linear Discriminant Analysis (LDA), Local Binary Pattern (LBP) and Gabor Wavelet Transform (GWT) were used to realize an improved pose and illumination invariant FET suitable for LR videos. 2900 LR frames were obtained from YouTube Celebrities corpus and a Locally Acquired Video Dataset (LAViD). These frames in the range of 30 pixels to 65 pixels were reconstructed using IBP-MAP. LDA-LBP-GWT ensemble was developed by fusing the facial features of Linear LDA, LBP and GWT into a LDA-LBP-GWT Single Feature Set (SFS). The SFS was dimensionally reduced using particle swarm optimization algorithm. The LR and the reconstructed frames were used as testing sets while locally acquired pose-oriented mugshots constituted the training set. Features of each frame in the testing sets were compared with those in the training set for recognition using Euclidean distance. The developed techniques were implemented in MATLAB 2019. The performance of the developed LDA-LBP-GWT ensemble was compared with the baseline techniques by using False Acceptance (FA), Recognition Accuracy (RA), Recognition Time (RT) and False Rejection (FR) as evaluation metrics. Results obtained indicate that the developed LDA-LBP-GWT ensemble serves as improvement over the baseline techniques in terms of FA, RA and RT.

Keywords: Ensemble-features, Face identification, Gabor Wavelet Transform, Illumination, Iterative Back Projection, Linear Binary Pattern, Linear Discriminant Analysis, Low-Resolution, Maximum-A-Posteriori, Particle Swarm Optimization, Pose, Unconstrained videos, YouTube Celebrities' dataset.

I. INTRODUCTION

The demand for emerging Video Surveillance Systems (VSS) is increasing in various fields of security applications and beyond as large publicly accessible facilities and urban sites, (including airports, train stations, power plants, banking halls, shopping malls, museums, parking garages and hotels), now deploy comprehensive VSS [1-2]. Notably, the sporadic influx of these security hardwares has posed new opportunities for the development and proliferation of video-based Face Recognition Systems (FRS) [3-4]. FRS tends to identify, verify or recognize one or more persons in a video frame from a video source by using a stored database of faces [5]-[8]. Surveillance applications based on face recognition are gaining increasing attention after the United States' "9/11" events and with the ongoing security threats [9-10]. Areas aiming at increased individual safety relative to terrorist threats further extend the integration of face recognition into VSS development [11-12]. However, videos captured by surveillance cameras are typically of low quality due to nuisance factors (including varying illumination, expression, pose and occlusion) and Low-Resolution (LR) [13]-[18]. All these challenges strongly affect recognition performance and have led to the failure of most existing video-based FRS especially when used in real-time mode [19]-[23].

This is further corroborated by the result of the evaluation of commercial video-based FRS conducted by Aryaz, Jonathan and Majid [24] revealing increased difficulties in recognizing faces due to variations that intensify the differences in appearance between images of the same individual. Resolution dependent performance differences, off-frontal poses and changing illuminations were observed to play a significant role in the poor verification performance of these systems. Hence, maintaining high recognition accuracy in a computationally-efficient manner continues to present challenges in building truly reliable video based FRS that operate well in less-controlled, low resolution imaging conditions. However, accuracy, scalability, fast performance and robustness to nuisance factors such as pose variations, illumination variation and LR are sought-after capabilities of automated surveillance FRSs [25-26]. Unfortunately, most implementations of the currently existing video-based FRSs are computationally very expensive and cannot cope with a lot of nuisance factors in surveillance videos wholly and these make them less effective for integration into VSSs for practical usage [20-21][27][28]. More often than not, it is almost impossible to control the imaging direction when capturing videos containing human faces in real time which makes a computationally efficient pose and illumination invariant recognition capability very crucial for optimal face recognition in video sequences. Thus, the development of an efficient video-based FRS becomes considerably a more challenging problem which by implication requires an equally wide range of accurate and computationally-efficient video enhancement method and video-based Feature Extraction Technique (FET) for video-based FRS to be usable. Video enhancement methods help enhance the biometric content of videos. To achieve this, Super Resolution (SR) can be used to consolidate the information in successive low resolution frames to generate the details of facial features of potential high resolution highly crucial for human recognition and further analysis [29-30]. In this process, a low resolution frame is being upsampled by recovering the missing high frequency details and degradations in the frame with the objective of constructing a high resolution frame.

One of the major challenges of video-based FRS is to obtain a feature extraction method that is insensitive to pose and illumination variations of videos captured in unconstrained environments [31]. Feature extraction is the most important stage of face recognition because if poor features are used, even the best classifier will fail to achieve an accurate result [17][32]. In the feature extraction stage, optimal discriminant features must be chosen to make the FRS not only computationally efficient but also robust to possible intrinsic and extrinsic facial variations [33]-[35]. Intrinsic factors are due purely to the physical nature of the face and are independent of the observer including aging while the extrinsic factors cause the appearance of the face to alter via the interaction of light with the face and the observer [36]. These extrinsic factors include illumination, pose, scale and imaging parameters like resolution, focus, imaging and noise [26]. Since most existing feature extraction methods are highly sensitive to these performance degrading factors [37], efficient pre-processing and feature extraction methods for the video frame can help realize improved recognition accuracy and minimize overall complexities possibly incurred during the video-based face recognition process [38]. Existing feature extraction methods are either global-based or local-based. The global feature extraction methods are the widely adopted techniques for face recognition task due to their good performance and high accuracy [35-36]; more often than not, they are computationally very expensive and do not perform effectively well under varying pose and illumination conditions [39-40]. In contrast, the local feature extraction methods are more computationally efficient, more robust to pose, facial expression and illumination variations but lack discrimination ability and can fail when local image information is insufficient especially when the target is very small or highly occluded as characterized by surveillance video frames in which using only the context of the image as a whole can help [21]. Consequently, both global and local features are crucial for efficient and accurate face recognition in low resolution, pose and illumination oriented videos. Hence, there arises a need for a feature extraction technique that can combine the strengths of both techniques.

In this paper, a resolution-aware ensemble of pose and illumination-invariant feature descriptors for low resolution video feeds is developed. IBP-MAP [41], a Super Resolution Reconstruction (SRR) technique based on the combination of Iterative Back Projection (IBP) and Bayesian Maximum *A* Posteriori (MAP) was employed to address the low resolution problem while an ensemble of local and global feature descriptors based on Linear Discriminant Analysis (LDA), Local Binary Pattern (LBP) and Gabor Wavelet Transform (GWT) was developed to address the pose and illumination challenges. Incorporation of the two (2) solutions into a video-based face recognition system framework is expected to help realize the sought-after capabilities of automated surveillance FRSs; and produce more accurate and efficient recognition output for real-time practical implementation in VSS [42]. In this paper, the statement of the problem is formulated as follows: given a low resolution, pose and illumination variant video $\overline{\mathrm{V}}$ containing n frames $\{f_1, f_2, f_3, \ldots, f_n\}$, with a subset of frames $\{h_1, h_2, h_3,\ldots, h_k\}$, $k \leq n$, which have favourable facial information such that $n, k \in N$ where N is a set of natural numbers $\{1, 2, 3,\ldots\}$; it is required to develop an accurate and computationally-efficient feature extraction technique combining the strengths of Local Binary Pattern (LBP), Gabor Wavelet Transform (GWT) and Linear Discriminant Analysis (LDA) for video $\overline{\mathrm{V}}$. The two major sub-problems are to reconstruct the LR frames $\{h_1, h_2, h_3,\ldots, h_k\}$ to realize frames with enhanced resolutions and develop an efficient pose and illumination-invariant feature extraction method for these super-resolved subset of frames in video $\overline{\mathrm{V}}$. The rest of this paper is organized as follows: Section 2 presents the materials and method; the results and some discussions are presented in section 3 while the conclusion is presented in section 4.

II. MATERIALS AND METHOD

The proposed resolution-aware ensemble of pose and illumination-invariant feature descriptors for face identification in unconstrained videos is made up of four (4) distinct phases which include:

i. Video dataset acquisition
ii. Video frame grabbing, resolution reconstruction and registration
iii. Pose and Illumination Invariant feature descriptors ensemble development
iv. Training, testing and performance evaluation.

These phases are illustrated in the system flow diagram of the developed ensemble feature extraction technique in Figure 1.

II.I Video Dataset Acquisition Phase

The first phase of this research involves the acquisition of video datasets. These include the YouTube celebrities dataset and the Ladoke Akintola University of Technology (LAUTECH) Video Dataset (LAViD).

II.I.I YouTube Celebrities Video Dataset

YouTube celebrities' video, a publicly available CCTV video dataset, was acquired through a direct download link http://seqam.rutgers.edu/site/media/data_files/ytcelebrity_init.tar provided by seqam educational resource institute. This dataset is challenging as the majority of the videos are low-resoluted with pose, illumination and expression largely uncontrolled [43-44].

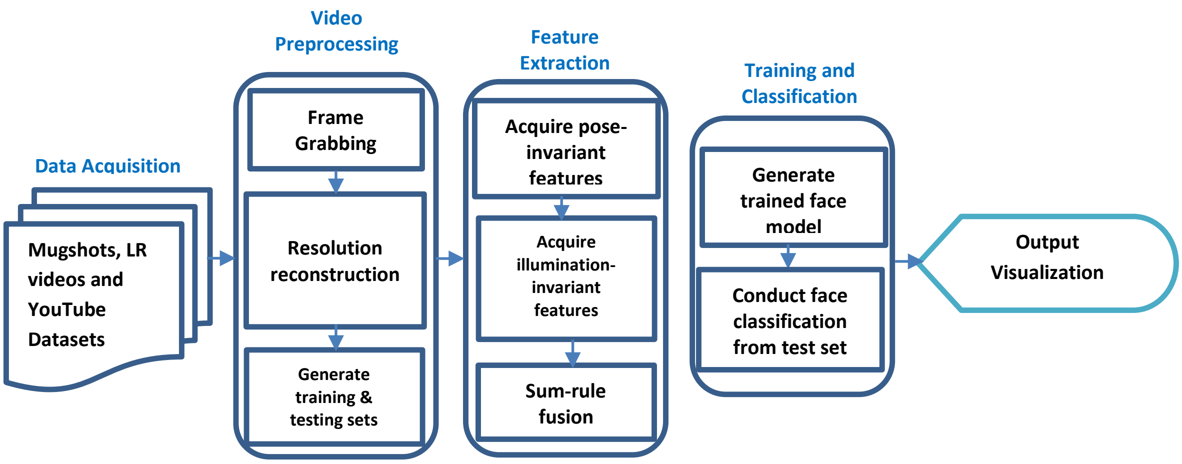

Fig. 1. Architecture of the developed resolution-aware ensemble of pose and illumination-invariant feature descriptors

II.I.II The LAUTECH Video Dataset and Mug Shots

LAUTECH Video Dataset (LAViD) was collected and the mug shots of subjects present in it were also captured. This is a black subject video dataset suitable for research purpose. The problem specifically addressed with this database is the law enforcement person identification from low quality surveillance videos or any other identification scenario where the subject's cooperation is not expected. Two (2) J&S JCC-915D surveillance video cameras with an active pixel of 597×537 and a minimum illumination of 0.3 lux; a digital camera, digital video surveillance recorder and a personal computer (PC) were used to collect the LAViD videos and the mug shots. Individual surveillance camera served indoor and outdoor purposes respectively. The Digital video surveillance recorder was Digital Sprite-2 manufactured by Dedicated Micros. The Digital Sprite-2 has 16 input video channels, two monitor outputs and LAN connector. It has 600 GB internal hard disk for storage of video streams from cameras and CD writer for writing data directly on CDs.

Photographer's camera used to capture the frontal facial mugshots was Canon EOS 10D model with 22.7×15.1 mm CMOS sensor, with 16.1 mega pixels, equipped with Sigma 18–50 mm F3.5–5.6 DC lenses and Sigma EF 500 DG Super flash. The facial mug shots are high quality static color images, taken in controlled indoor illumination conditions. Mug shot imaging conditions are exactly the same as would be expected for any law enforcement or national security use (passport images or any other personal identification document). Participants were photographed with digital camera at close range in controlled conditions (standard indoor lighting, adequate use of flash to avoid shades, high resolution of images). The capturing was conducted over a period of three (3) months at LAUTECH, Ogbomoso, Oyo State. From the total of 100 volunteers, 67 were males and 33 females. However, the participants enrolled in the datasets are students and casual staff of LAUTECH.

II.II Frame Grabbing, Resolution Reconstruction and Dataset Registration

This is the second phase of the development in which frames were grabbed from YouTube and LAViD videos, reconstructed and registered.

II.II.I Video Frame Grabbing

Video2Photo application was used to grab frames from the short video clips. It is an end-user software that can be used to convert video stream to a set of frames. It offers a number of flexibilities including the choice of video, either from online (for example, camcorder or USB camera) or offline video sources.

II.II.II Video Frame Resolution Reconstruction

IBP-MAP [41], a super resolution reconstruction technique, based on the combination of Bayesian Maximum-A-Posteriori (MAP) restoration technique and Iterative Back Projection (IBP), was used for reconstruction of the video resolution. MAP provided the edge restoration properties of the video frame via the $l1$ norm, non-edge restoration properties and computational efficiencies via the SAR prior, while a simple and widely used Tikhonov L_2 regularization technique provided the noise-removing and additional edge smoothing properties. IBP was introduced to minimize the reconstruction error produced by MAP in an iterative manner significantly and helped realize better and more reliable super-resolved frames in real-time mode.

II.II.III Video Frame Dataset and Mugshots' Registration

By registration, disjoint gallery, training and the testing sets were prepared from LR and super-resolved YouTube video frame sets and from LAViD and the Mug Shots for the purpose of evaluating the developed feature descriptor ensemble. The purpose of the training set is for the recognition algorithm to learn a projection matrix '*P*'. The gallery and probe sets are used in the testing stage. The gallery set contains images with known identities and the probe set with unknown identities. The algorithm associates descriptive features with the images in the gallery and probe sets and determines the identities of the probe images by comparing their associated features with those features associated with gallery images.

1) Gallery Set

For the LAViD, there is one frontal mug shot per subject. There are in total 100 frontal facial mug shot images in the dataset, which is enough to eliminate performance results obtained by pure coincidence (chances of recognition by pure coincidence is less than 1/100 ≈ 0.01%) which is far lesser than '<=0.08%' recommended by the surveillance security and identification system developers [45]. Images are in lossless 24 bit color JPEG format with the original size of 3,072×2,048 pixels, cropped to 1,600×1,200 pixels which is a recommended size for gallery images [46]. Cropping was done following the ANSI 385-2004 standard [47] recommendations so that the face occupies approximately 80% of the image. These mug shot images are those expected to be found in a law enforcement database or when registering to a security system. No mug shot exists for the YouTube video dataset. As a result, a video-to-video recognition was performed in which the gallery and the probe are both video frames. The YouTube gallery is made up of forty-five (45) grabbed and IBP-MAP super-resolved facial frames of different celebrities. However, videos having high to considerably normal visual quality level with near-frontal facial information only were grabbed. The pure coincidence level obtained is 0.02% which is far lesser than the recommended ≤ 0.08% and thus highly suitable for the required purpose.

2) Training Set

Using multiple pose samples of the same individual helps improve the quality of the verification system. For the LAViD, this set of mug shots provides eight (8) discrete views of each face per class, ranging from left to right profile in equal steps of 22.5 degrees except for the mug shot at 0 degree (frontal face). To assure comparable views for each subject, numbered markers were used as fixation points. As a final result, dataset contains eight (8) different pose images per subject, which represents a class, from which prototype was built for head pose estimation with views from −90 to +90 degrees. In total, there are 800 images in the dataset containing 100 different classes. For the YouTube videos, the training set is made up of images consisting of 8 distinct poses per subject. In total, there are 360 super-resolved images for 45 different celebrities.

3) Testing Set

For the LAViD, twenty (20) low resolution, different pose and illumination varying frames per subject were grabbed and prepared into a probe set A. These frames were super-resolved using IBP-MAP and prepared into another probe set B, making two (2) distinct probe sets one containing LR probes while the other contains super-resolved probes. In all, each probe set contains 2000 pose and illumination varying frames of 100 different subjects. For the YouTube Video, twenty (20) low resolution, different pose and illumination varying frames per subject were grabbed and prepared into a probe set C. These frames were super-resolved using IBP-MAP and prepared into another probe set D, making two (2) distinct probe sets (the first set containing LR probes while the other set contains the super-resolved probes). In all, each probe set contains 900 pose and illumination varying frames of 45 different celebrities. Each probe is matched with each of the gallery images by using the Euclidean distance. Probe is assigned the identity of the gallery subject for which it has the shortest distance. Comparing the probe image to one gallery image is the most logical real-world law enforcement scenario.

II.III Pose and Illumination-Invariant Feature Descriptors' Ensemble Development

The goal of feature extraction is to create a low-dimensional representation of faces with good discriminatory power for classification [48]. LDA, a global FET and two (2) local FETs (LBP and GWT) were combined by consolidating their respective feature sets into a single feature set after normalization and feature selection schemes to realize an improved feature extraction method referred to as LDA–LBP-GWT technique. A feature-level fusion strategy using sum rule was adopted to fuse the extracted features. Sum rule is defined as [49]:

$$s\,(U,V) = \frac{1}{ab}\sum_{i=1}^{a}\sum_{j=1}^{b} s\,(u_i, v_j) \tag{1}$$

where U and V are two quantities to be added which are input arguments to the sum function s ( ), a and b are the total number of features in U and V respectively while u_i *and* v_j represent specific feature in U and V respectively. Discriminant feature selection plays the central role in illumination and pose invariant recognition and classification [43][50]. To this end, Particle Swarm Optimization (PSO) was used to manage the curse of dimensionality drawback of the FETs to realize a fewer optimal feature subset suitable for recognition. As shown in the developed feature extraction technique presented in Figure 2, GDFV is the global discriminant feature vector, LPFV is the local pattern feature vector and LGFV is the local gabor feature vector. C_{LPFV} is the combined LBP feature vectors, C_{LGFV} is the combined GWT feature vectors, w_G denotes the weight of the global features extracted using LDA and 1-w_G denotes the weight of the local features, a combination of features extracted using LBP and GWT.

The fused local feature vectors for LBP and GWT using sum rule is formulated as:

$$C_{L_i} = (\, C_{LGFV_i} + C_{LPFV_i}) \,/\, 2 \tag{2}$$

such that

$$C_L = \sum_{i=1}^{N} w_{L_i} \,.\, C_{L_i} \tag{3}$$

where C_L is the combined local feature vectors for LBP and GWT, w_{L_i} is the weight of the local feature vector and N is the lowest dimension of the feature vectors of LBP and GWT. However, the fused feature set using sum rule fusion strategy is formulated as:

$$F_n = w_G C_G + (1 - w_G)\, C_L \tag{4}$$

where F_n is the fused set of corresponding optimized low-dimensional LDA, LBP and GWT features and C_G is the optimized LDA global feature vector.

The pseudocode for the fusion of the feature extraction techniques is shown as follows:

Initialize *gallery set*
    For *frame = 1,2,…N*
        Project *probe image into LDA*
        Project *probe image into LBP*
        Project *probe image into GWT*
        Compute *feature vectors FV_{LDA} using processed LDA*
        Compute *feature vectors FV_{LBP} using LBP*

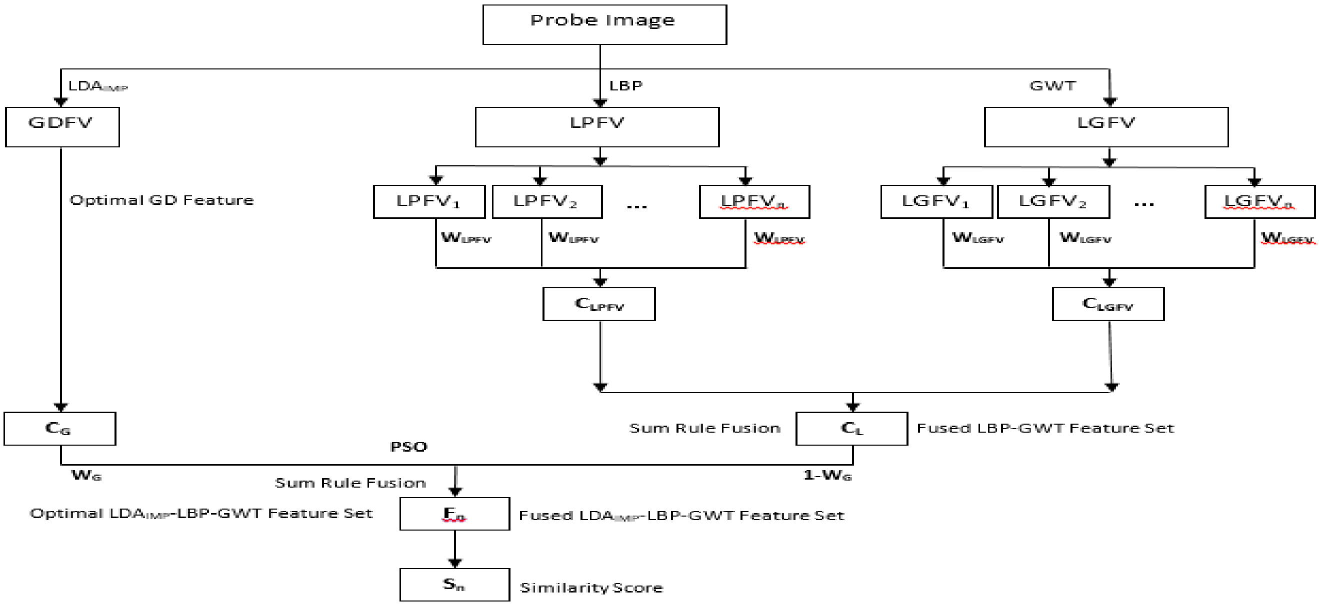

Fig. 2. The Block Diagram of the Developed Ensemble of LDA-LBP-GWT Feature Descriptors

Compute *feature vectors* FV_{GWT} *using GWT*
Sum $\Phi total\text{-}FV(i) = (FV(i)_{LDA} + (FV(i)_{LBP} + FV(i)_{GWT}) / 2))$
Compute *fitness selection measure* $\Phi total\text{-}FV$ *using PSO*
Accumulate *selected feature subsets* $\Phi subset\text{-}FV$
End
Recognize *using the optimal feature subsets obtained from the total feature vector* $\Phi total\text{-}FV$

Within each class, the distance vectors d^{LBP}, d^{GWT} and d^{LDA} were normalised in order to reduce the range of these distances in the interval [0,1] following the FRVT recommendation [45]. A combined distance vector d that must contain LBP, GWT and LDA information was obtained by computing their mean vector using sum rule fusion strategy as follows:

$$d = \left\{ \frac{d_1^{LDA} + d_1^{LBP} + d_1^{GWT}}{3}, \dots, \frac{d_N^{LDA} + d_N^{LBP} + d_N^{GWT}}{3} \right\} \tag{5}$$

where N is the lowest ordered number of discriminating vectors among the three (3) feature descriptors.

II.III.I Training Phase

$Ti = \{t_1, \dots, t_M\}$ is formulated and defined as M observations of class i in the training set $T = \{T_1, \dots T_K\}$ with multiple images of each of the K individuals, where $i \leq K$, and each Ti with observations of a N by N dimension. All color images were pre-processed by converting them to grayscale using the predefined rgb2gray() function in MATLAB and the developed IBP-MAP technique. A vector of dimension N^2 was produced and the summary image Ti was determined as the mean of $\{t_1, \dots, t_M\}$. The gallery images were projected onto a projection matrix W^T. X is a matrix containing the images expressed as vectors in its columns, x_{mean} is the mean image vector, X^{ϕ} is the matrix containing mean subtracted images in its columns (mean deviation) and x_g is the gallery image vector. During the training phase, the projection matrix, P_x, containing the basis vectors of the subspace was calculated and then the gallery images were projected onto that subspace and their projections are stored in a database. The block diagram of the training phase for the development ensemble of LDA-LBP-GWT feature descriptors is presented in Figure 3.

II.III.II Testing Phase

At this phase, each probe was enrolled, pre-processed, mean-subtracted and projected onto the same subspace as the gallery image and its projection was compared with the stored gallery projections. The similarity measure was determined by calculating the distances d from a probe image projection to all gallery images projections and then choosing the minimum distance. The identity of the most similar gallery image was chosen to be the result of recognition and the unknown probe image was identified. Euclidean distance was used as the measure for classification because it is computationally efficient and the most commonly used distance-based similarity measure for high-dimensional positive spaces especially in face pattern matching evaluations [51]. The block diagram of the testing phase for the development ensemble of LDA-LBP-GWT feature descriptors is presented in Figure 4.

II.IV The Performance Evaluation Metrics

The metrics for evaluating the performance of the feature descriptors include:

1) ***Recognition Accuracy (RA):*** 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 extracted from video frames [43].

$$RA = \frac{\text{Number of correctly recognized persons in video frames}}{\text{Total number of persons tested}} \times 100\% \tag{6}$$

2) ***False Accept Rate (FAR):*** This is the percentage of probes a system falsely accepts even though their claimed identities are incorrect [52].

$$FAR = \frac{\textit{Number of false accepts}}{\textit{Number of impostor scores}} \times 100\% \tag{7}$$

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

International Journal of Engineering Research and Technology. ISSN 0974-3154, Volume 12, Number 12 (2019), pp. 3114-3126

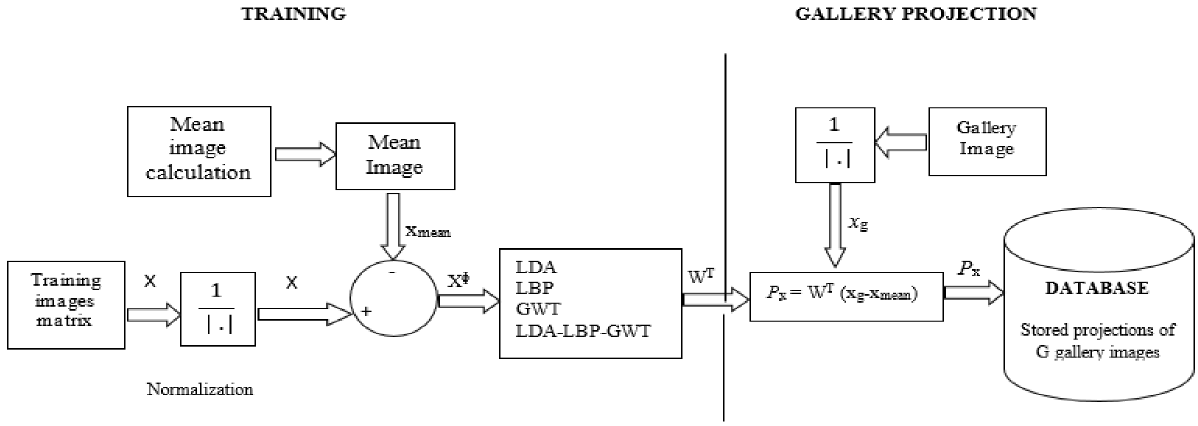

Fig. 3. The Block Diagram of the Training and Gallery Projection Process for the Developed LDA-LBP-GWT Ensemble

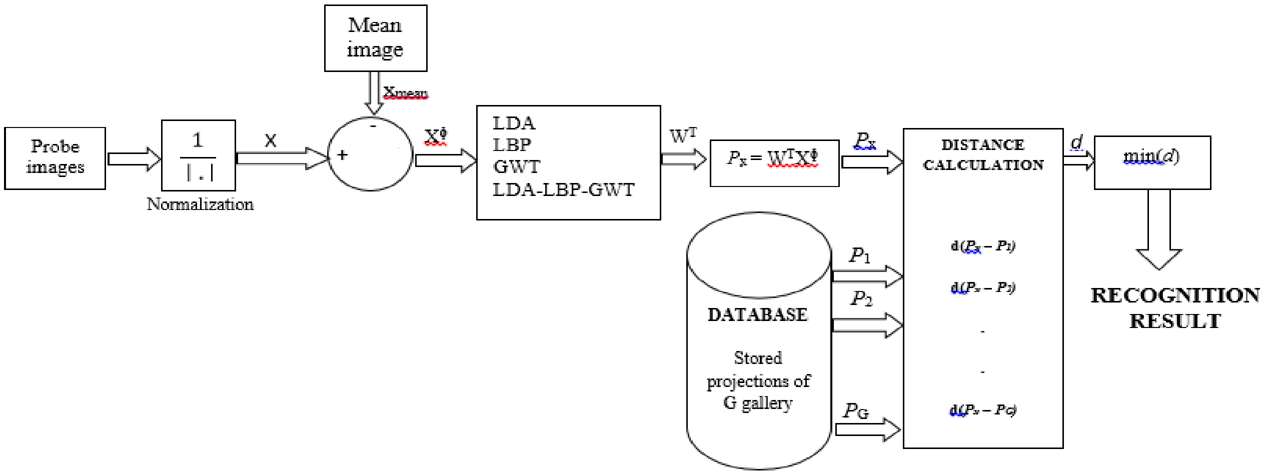

Fig. 4. The Block Diagram of the Matching Phase between the Probe and Gallery Images

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 [52].

$$FRR: \frac{Number\ of\ false\ rejects}{Number\ of\ genuine\ scores} \text{ x } 100\% \qquad (8)$$

III. RESULTS AND DISCUSSION

In this section, the results of several experiments on the LAViD Frontal Mugshots and low resolution LAViD probes, LAViD mugshots and the super-resolved LAViD probes, super-resolved YouTube dataset and Low Resolution YouTube probes and the super-resolved YouTube gallery and super-resolved YouTube probes.

III.I Results Obtained using the LAViD Frontal Mug shots and Low Resolution LAViD Probes

The evaluation results of the FETs obtained using the LAViD frontal mug shots and LR LAViD probes are summarized and presented in Table 1. The architecture is such that the gallery comprises of frontal mug shots and the test probes are low resolution video frames. An identical match of the probe is determined among the frontal mug shots. The developed ensemble produced the least false acceptance of 725 out of a total of 2,000 test probes and as such the most reliable. On the other hand, LDA, LBP and GWT yielded false acceptance of 900, 1160 and 1780 respectively. The considerably high rate of false acceptance is due to the fact that the identification was carried using unsupervised learning approach in which there exists no overlapping among the images in the gallery, training and the testing sets. All the FETs evaluated produced zero (0) false rejections. This could be due to the fact that distance-based measure was used. The justification for this is borne out of the research outputs by Kuldeep and Madan [51] which ascertain that identification using machine learning techniques may incur additional computational overheads during training and testing sessions than the distance-based measures and as well yield higher values of false rejection especially when the hyperplane is fooled as is the case of support vector machine. GWT, LBP, LDA and the developed LDA-LBP-GWT ensemble produced recognition accuracies of 11, 42, 55 and 63.75, respectively. Hence, the developed technique showed remarkable improvement over others following identification of low resolution frames in still image dataset. Followed closely is the LDA with higher recognition rate than GWT and LBP. The high recognition rate produced by LDA confirms the assertion by Yu *et al.* [35] that global feature extraction methods are highly accurate. In the same vein, the low recognition accuracies observed in LBP and GWT corroborate with the findings of Christophe *et al*. [21] which states that local feature descriptors lack discrimination ability and can fail when local image information is insufficient especially when the target is very small or highly occluded as characterized by surveillance video frames in which using only the context of the image as a whole can help. However, in ascending order of computational efficiency, LDA produced a training time of 993.5, followed by the developed technique with 436.321, GWT with 164.531 and LBP with 51.123. These results agree with the report of Wang *et al.* [40] that global techniques including LDA are computationally very expensive. However, the low computational overhead obtained by LBP and GWT is confirmed by Rabia and Hamid [36] who asserted that the local feature extraction methods are more computationally efficient. The testing time of the FETs are 103.651, 140.469, 253.125 and 260.391 for the developed LDA-LBP-GWT ensemble, LBP, LDA and GWT, respectively in descending order of real-time computational efficiency. Thus, the developed technique is the most applicable and computationally efficient for real time face recognition in surveillance videos.

III.II Results Obtained using the Locally Acquired Mugshots and Super-Resolved LAViD Probes

The evaluation results of the FETs obtained using the LAViD frontal mug shots and super-resolved LAViD probes are summarized and presented in Table 2. The architecture is such that the gallery comprises of frontal mug shots and the test probes are super-resolved video frames. An identical match of the probe is determined among the frontal mug shots. The developed feature descriptor ensemble technique produced the least false acceptance of 340. Furthermore, LDA, LBP and GWT produced false acceptance of 440, 900 and 1440 respectively. As a result, the developed feature descriptor ensemble technique is the most reliable of all. All the FETs yielded zero (0) false rejections. This is uncompromisingly an encouraging result as it implies that none of the probes that has an identity in the gallery was falsely rejected. The accuracies of recognition obtained by GWT, LBP, LDA and the developed LDA-LBP-GWT ensemble technique are 28, 55, 78 and 83 respectively.

Combination of local and global feature characteristics make the developed technique surpasses its variants in terms of accuracy. The high recognition rate noticed in LDA confirms the assertion by Yu *et al.* [35] that global feature extraction methods are highly accurate. Similarly, the low recognition accuracies observed in LBP and GWT agrees with the findings of Christophe *et al*. [21] which states that local feature descriptors lack discrimination ability and can fail when local image information is insufficient especially when the target is very small or highly occluded as characterized by surveillance video frames. The training time for LDA is 931.406. This delay can be accounted for by the fact that global feature descriptors suffer from the curse of dimensionality which can in turn affect their performance and computational efficiencies [22].

In the same vein, the developed technique completed the training phase in 387.95s. This is also a lot of time compared to using local feature descriptors only; comparatively, it is approximately three (3) times computationally more efficient than LDA. This improvement explains the reason Nisar [49] suggested the combination of different classes of feature descriptors for efficient and more reliable outcomes. LBP and GWT completed the training phase in 26s and 164.531s respectively. The trade-off between computational efficiency and accuracy of feature descriptors accounted for the poor performance of these two (2) techniques in terms of accuracy; though they are the most computationally efficient out of all. In descending order of real time computational efficiency, the developed LDA-LBP-GWT, LBP, LDA and GWT completed the testing stage in 83.9627s, 93.0938s, 95.3125s and 257.109s respectively. Thus, the developed technique is the most computationally efficient for real time face recognition in surveillance videos.

III.III Results obtained using the Super-Resolved YouTube Dataset and Low Resolution YouTube Probes

This is a video-to-video identification process in which the gallery and the probe (s) used for matching are both video frames. In this case, the gallery is composed of super-resolved, near-frontal YouTube frames while the probes are low resolution, pose and illumination challenged YouTube frames. The evaluation results of the FETs obtained following this arrangement are summarized and presented in Table 3. The developed FET produced the least false acceptance of 318, strictly followed by LDA with 432, LBP with 504 and GWT with 765. The high rate of false acceptance by all the FETs is due to the low resolution, pose and illumination challenged probes used for matching. Secondly, the learning process was unsupervised. Despite all these challenges, the developed ensemble technique emerged as the most reliable of all. All the FETs produced zero (0) false rejection. This means that none of the probes with an identity in the gallery was falsely rejected. The recognition accuracy obtained for GWT, LBP, LDA and the developed LDA-LBP-GWT techniques in percentage are 15, 44, 52 and 64.7 respectively. It is evident that the developed technique is the most accurate over others following identification of low resolution frames in video image dataset. Followed closely is the LDA with higher recognition rate than GWT and LBP.

The high recognition rate produced by LDA confirms the assertion by Yu *et al.* [35] that global feature extraction methods are highly accurate. In the same vein, the low recognition accuracies observed in LBP and GWT corroborate with the findings of Christophe *et al*. [21] which states that local feature descriptors lack discrimination ability and can fail when local image information is insufficient especially when the target is very small or highly occluded as characterized by surveillance video frames in which using only the context of the image as a whole can help. LDA has the highest training time of 444.310s, followed by the developed technique with 92.643s, GWT with 64.317s and LBP with 41.276s. Though the developed technique is approximately five (5) times computationally more efficient than LDA, it is still a fact that it spends a lot of time during the training phase compared to using local feature descriptors only.

Table 1. Evaluation results of the feature extraction methods for LR LAViD probes

FET	False Acceptance	False Rejection	Recognition Accuracy (%)	Training Time (*s*)	Testing Time (*s*)
LDA	900	0	55	993.500	253.125
LBP	1160	0	42	51.123	140.469
GWT	1780	0	11	164.531	260.391
LDA-LBP-GWT	725	0	63.75	436.321	103.651

Table 2. Evaluation results of the feature extraction methods for SR LAViD probes

FET	False Acceptance	False Rejection	Recognition Accuracy (%)	Training Time (*s*)	Testing Time (*s*)
LDA	440	0	78	931.406	95.3125
LBP	900	0	55	26	93.0938
GWT	1440	0	28	164.531	257.109
LDA-LBP-GWT	340	0	83	387.95	83.9627

Table 3. Evaluation Results of the feature extraction methods for LR YouTube probes

FET	False Acceptance	False Rejection	Recognition Accuracy (%)	Training Time (*s*)	Testing Time (*s*)
LDA	432	0	52	444.310	275.814
LBP	504	0	44	41.276	84.992
GWT	765	0	15	64.317	92.635
LDA-LBP-GWT	318	0	64.7	92.643	75.929

More often than not, LBP and GWT are the most computationally efficient; yet, their performances are not encouraging. In descending order of real time computational efficiency, the developed LDA-LBP-GWT, LBP, GWT and LDA completed the testing stage in 75.929s, 84.992s, 92.635s and 275.814s respectively. By implication, the developed technique is the most computationally efficient for real time face recognition in surveillance videos while LDA suffers from a large computational efficiency drawback due to curse of dimensionality menace.

III.IV Results obtained using the super-resolved YouTube Gallery and super-resolved YouTube Probes

In this case, the gallery is composed of super-resolved, near-frontal YouTube frames while the probes are super-resolved, and pose and illumination challenged YouTube frames. The evaluation results of the FETs obtained are summarized and presented in Table 4. The developed technique produced the least false acceptance of 216, followed by LDA with 369, LBP with 428 and GWT with 666. The reduced rate of false acceptance by all the FETs is due to the use of super-resolved, pose and illumination challenged probes for matching. However, the developed technique emerged as the most reliable. All the FETs produced zero (0) false rejection. This means that none of the probes with an identity in the gallery was falsely rejected. The recognition accuracy obtained for GWT, LBP, LDA and the developed LDA-LBP-GWT ensemble techniques in percentage are 26, 52.5, 59 and 71, respectively. This shows that the developed technique thrives well and highly suitable for video-to-video face identification. LDA also yielded higher recognition rate than GWT and LBP. The high recognition rate produced by LDA confirms the assertion by Yu *et al.* [35] that global feature extraction methods are highly accurate. However, the low discrimination ability of LBP and GWT accounted for their poor performances with disjoint datasets. LDA has the highest training time of 391.065s, followed by the developed technique with 76.483s, GWT with 61.705s and LBP with 36.517s. It is noticeable that the developed technique is over five (5) times computationally more efficient than LDA but over two (2) times less efficient than LBP, the most computationally efficient of all. In addition, at testing, the time obtained in seconds (s) for the developed LDA-LBP-GWT ensemble, LBP, GWT and LDA is 71.763, 77.658, 86.332 and 216.416 respectively.

Table 4. Evaluation Results of the Feature Extraction Techniques for SR YouTube Probes

FET	False Acceptance	False Rejection	Recognition Accuracy (%)	Training Time (*s*)	Testing Time (*s*)
LDA	369	0	59	391.065	216.416
LBP	428	0	52.5	36.517	77.658
GWT	666	0	26	61.705	86.332
LDA-LBP-GWT	261	0	71	76.483	71.763

By implication, the developed technique is the most computationally efficient for real time face recognition in surveillance videos while LDA is the most computationally expensive among all the techniques evaluated though its performance is near optimal compared to the more efficient GWT and LBP techniques.

III.V The Effect of Super Resolution Reconstruction on the Performance of the Feature Extraction Techniques

The effect of SRR on the recognition accuracy, false acceptance and processing time of the feature extraction techniques was determined for the LAViD and the YouTube video dataset. The results obtained by the feature extraction techniques using low resolution and super–resolved probes were compared for each dataset. The plot of recognition accuracy of the feature extraction techniques on low resolution and super-resolved LAViD probes is presented in Figure 5. Specifically, LDA gained additional 23% recognition accuracy, LBP gained 11%, GWT gained 17% and the developed technique gained approximately 19% increase. It was observed that all the techniques gained significant improvement in recognition accuracy using the super-resolved probes. This is a clear indication that low resolution challenge is a strong determinant of the recognition accuracy of feature extraction techniques. Similarly, the plot of recognition accuracy for low resolution and super-resolved YouTube probes is presented in Figure 6. Clearly, LDA improved with additional 7% increase, LBP with 8.5%, GWT with 11% and the developed technique with 6.3% when the super-resolved probes were used for identification.

This result confirms that there is a significant relationship between image resolution and recognition accuracy. That is, the higher the resolution, the better the recognition accuracy as hypothesized by Jeremiah *et al.* [43]. The effect of SRR on false acceptance by FETs was determined using the LR and super-resolved LAViD and YouTube probes. The plots of false acceptance of the feature extraction techniques on low resolution and super-resolved LAViD and YouTube video frame probes are presented in Figures (7 and 8) respectively. With LAViD super-resolved probes, false acceptance reduced by 460 for LDA, by 260 for LBP, by 340 for GWT and by 385 for the developed technique. In the same vein, with YouTube super-resolved probes, false acceptance reduced by 63 for LDA, by 76 for LBP, by 99 for GWT and by 57 for the developed technique. This improvement is an indication that there is a strong relationship between resolution of probes and false acceptance rate especially in a video to video matching scenario. The processing time is composed of the training and the recognition time. The plots of the training time of the FETs on LR and super-resolved LAViD and YouTube probes are presented in Figures (9a and 9b) respectively.

The training time in seconds (s) of LDA, LBP, GWT and the developed technique got reduced by 62.1, 25.1, 0 and 48.4 respectively using LAViD probes. However, with the YouTube probes, the training time in seconds (s) of LDA, LBP, GWT and the developed technique got reduced by 53.2, 4.76, 2,612 and 16.16 respectively. It was noticed that LDA is very sensitive to SRR as well as the developed technique with higher changes in training time. However, GWT has remained less sensitive but yet maintains a zero-to-positive noticeable increase. The plots of the recognition time of the FETs on LR and super-resolved LAViD and YouTube probes are presented in Figures (10a and 10b) respectively. With the LAViD probes, the recognition time in seconds (s) by LDA, LBP, GWT and the developed technique got reduced by 157.8, 47.38, 3.28 and 19.69 respectively. In the same vein, with the YouTube probes, the recognition time in seconds (s) by LDA, LBP, GWT and the developed technique got reduced by 59.4, 7.33, 6.3 and 4.17 respectively.

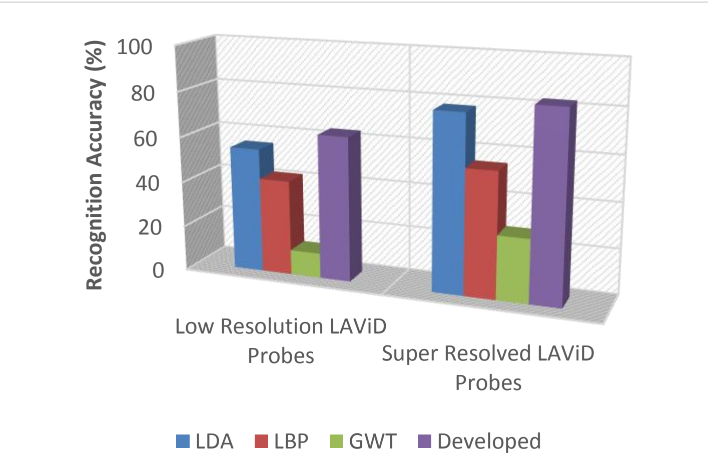

Fig. 5. Recognition Accuracy of the Feature Descriptors with LAViD

International Journal of Engineering Research and Technology. ISSN 0974-3154, Volume 12, Number 12 (2019), pp. 3114-3126
© International Research Publication House. http://www.irphouse.com

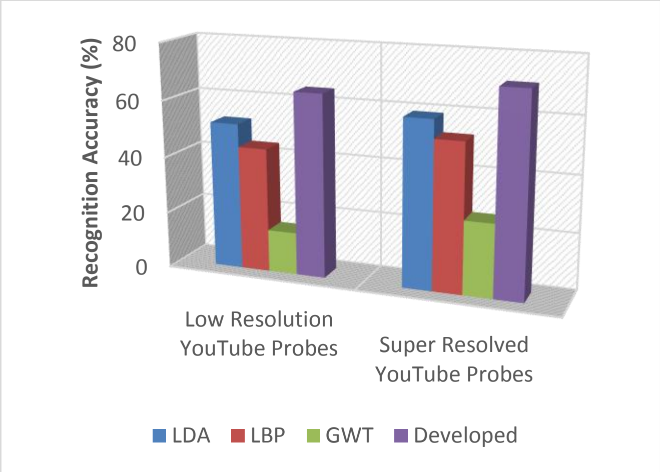

Fig. 6. Recognition Accuracy of the Feature Extraction Techniques with YouTube Frames

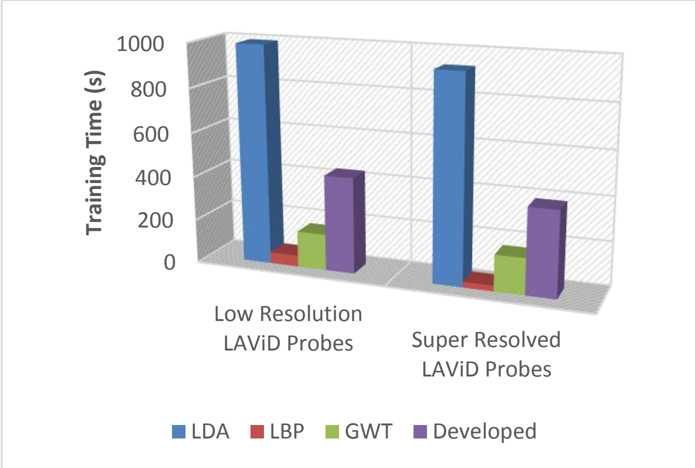

Fig. 9a. Training Time of the Feature Extraction Techniques using LAViD

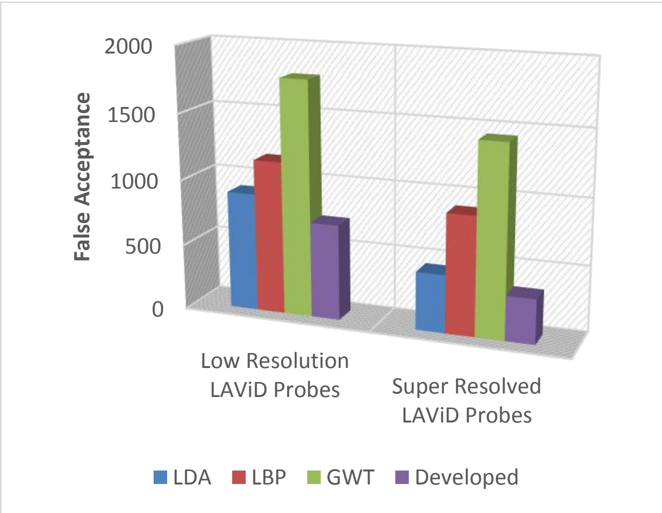

Fig. 7. False Acceptance of the Feature Extraction Techniques using LAViD

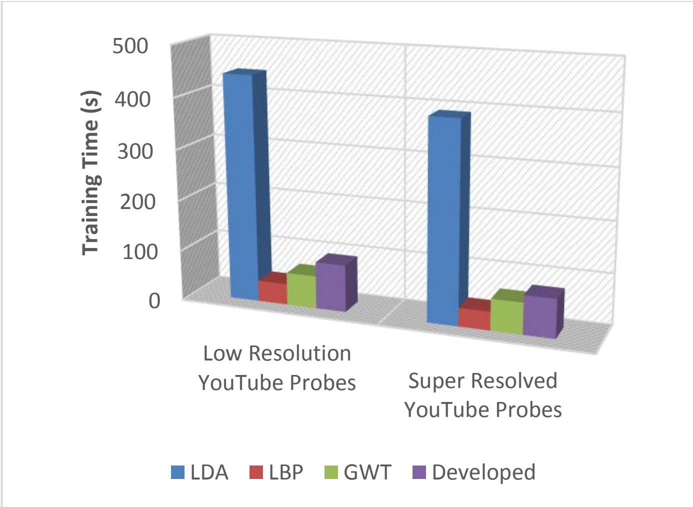

Fig. 9b. Training Time of the Feature Extraction Techniques using YouTube Frames

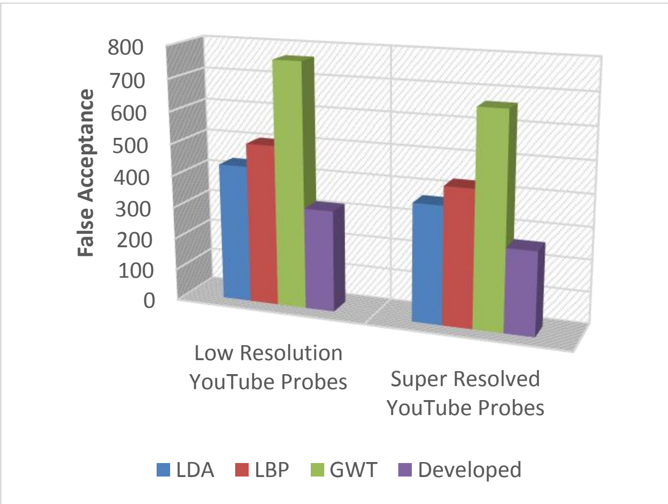

Fig. 8. False Acceptance of the Feature Extraction Techniques using YouTube Frames

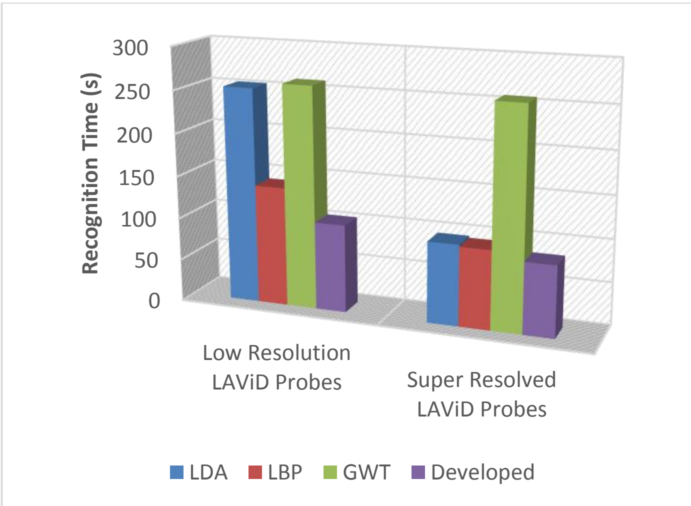

Fig. 10a. Recognition Time of the Feature Extraction Techniques using LAViD Frames

International Journal of Engineering Research and Technology. ISSN 0974-3154, Volume 12, Number 12 (2019), pp. 3114-3126
© International Research Publication House. http://www.irphouse.com

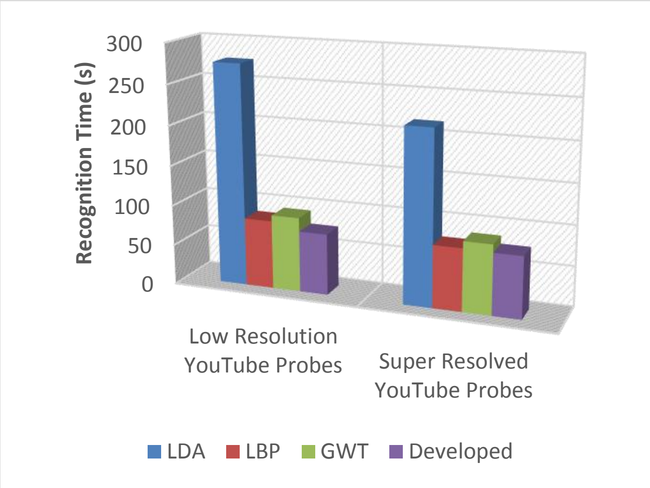

Fig. 10b. Recognition Time of the Feature Extraction Techniques using YouTube Frames

V. CONCLUSION

A resolution-aware ensemble of pose and illumination invariant feature descriptors for face identification in low resolution video feeds was developed. This is in a bid to address low resolution, pose-and-illumination challenges inherent in surveillance videos for improved performance of video-based face recognition systems. The developed technique employs an IBP-MAP super resolution reconstruction technique and a LDA-LBP-GWT ensemble of feature descriptors. The performance of the developed ensemble was evaluated using processing time, False Acceptance and False Rejection. The summarized result of all evaluations conducted for the feature extraction techniques showed that the developed LDA-LBP-GWT ensemble performed better than LDA, LBP and GWT as it produced the highest recognition accuracy, least false acceptance and least testing time. There was noticeably a general significant improvement in performance and computational efficiency of the feature descriptors due to frame resolution reconstruction. LDA, LBP, GWT and the developed LDA-LBP-GWT gained 23%, 11%, 17% and 19% increase in recognition accuracy. Similarly, false acceptance of LDA, LBP, GWT and the developed LDA-LBP-GWT got reduced by 460, 260, 340 and 385 respectively while the recognition time got reduced by 157.8s, 47.38s, 3.28s and 19.69s respectively. These results establish the fact that there is a direct significant relationship among image resolution, recognition accuracy, processing time and false acceptance. However, performance of the developed technique is dependent on the nature and characteristics of the dataset used. Conclusively, this work has helped to manage LR, pose and illumination challenges by surveillance videos limiting the performance and optimality of FRS via the development of a more accurate and computationally-efficient pose and illumination invariant feature extraction technique suitable for low resolution surveillance video feeds. This in turn will improve the performance of video-based FRS adopting the solution in practice. Future works could investigate how super resolution reconstruction systems could be integrated directly into the feature extraction process to manage the time complexities inherent in merging and maintaining separate solutions for resolution reconstruction and face recognition in videos. Software complexity metrics like the Halstead software complexity measure could be adopted to evaluate the soft complexities and performance of the developed system. Other distance metrics can be implemented to test the performance of the feature descriptors on other available pose and illumination video datasets. A linear model establishing a direct relationship between resolution of videos and the rate of recognition can be developed to help guide the resolution reconstruction and recognition process of faces in a more direct manner.

REFERENCES

[1] Alaa E., Hüseyin O. and Hasan D. "Adaptive Fitness Approach - An Application for Video-Based Face Recognition", New Approaches to Characterization and Recognition of Faces, InTech, 2012:153-170.

[2] Sven S. and Wolfgang B. "Adjustable Module Isolation for Distributed Computing Infrastructures", GRID, 2010:98-105.

[3] Jason T., Jeanette BG, Daniel B., Michael C. and Heather Z. "Person Attribute Search for Large-Area Video Surveillance", Homeland Security Affairs, 2012, 5(1).

[4] Laura SL. "Local Binary Patterns applied to Face Detection and Recognition", A thesis submitted to the Department of Signal Theory & Communication in partial fulfillment of the requirement for the award of Masters of Science degree in Electrical & Electronics of the Universitat Poltecnica De Catalunya, 2010.

[5] Soltanpour S., Boufama B. and Wu Q.J. A survey of local feature methods for 3D face recognition. Pattern Recognition. 2017:72:391.

[6] Saroj D., Shilpa T., Niraj W., Kailash D. and Singh K.R. "Face Detection from Pose Varying Facial Images", International Journal of Advances in Engineering & Technology, 2012:3(2):286-292.

[7] Omidiora E.O., Fakolujo O.A., Ayeni R.O. and Ajila T.M. "A Survey of Face Recognition Techniques", Journal of Applied Science, Engineering and Technology, 2007:7(1): 57-65.

[8] Zhao W., Chellappa R., Phillips P. J. and Rosenfeld A. "Face Recognition: A Literature Survey", ACM Computing Surveys, 2003:35(4): 399-458.

[9] Motorola "Video Surveillance Trade-Offs: A Question of Balance: Finding the Right Combination of Image Quality, Frame Rate and Bandwidth" Motorola Solutions, Inc. 1301 E. Algonquin Road, Schaumburg, Illinois 60196 U.S.A, 1-8, 2012.

[10] Gabriel C., Sathish M., Alexandru D., Petronel B. and Peter C. "Real-Time Video Face Recognition for Embedded Devices", A Book of Proceedings on New Approaches to Characterization and Recognition of Faces, InTech, 2012:153-170.

[11] Vhalos J. "Surveillance Society: New High-Tech Cameras Are Watching You", Popular Mechanics, 2008.

[12] Abbas B., Colin S., Morteza BA and Brian L. "Face Detection on Embedded Systems", Retrieved from www.viisage.com, 2006:1-12.

[13] Hu X., Peng S., Li W., Yang Z. and Li Z. Surveillance video face recognition with single sample per person based on 3D modeling and blurring. Neurocomputing. 2017:**235**:46.

[14] Nestor A., Plaut D.C., Behrmann M., Feature-based face representations and image reconstruction from behavioral and neural data. Proceedings of the National Academy of Sciences. 2016:**113**, 416.

[15] Jean-François C., Eric G. and Robert S. "An Adaptive Classification System for Video-Based Face Recognition" Information Sciences, 2012:192: 50–70.

[16] Arandjelovic O. and Cipolla R. "Achieving Robust Face Recognition from Video by Combining a Weak Photometric Model and a Learnt Generic Face Invariant", Pattern Recognition, 2012: http://dx.doi.org/10.1016/j.patcog.06.024.

[17] Huafeng W., Yunhong W. and Yuan C. "Video-based Face Recognition: A Survey", World Academy of Science, Engineering and Technology, 2009:60: 293-302.

[18] Hannan M.A., Hussain A., Mohammed A. and Samad S.A. "TPMS Data Analysis for Enhancing Intelligent Vehicle Performance", Journal of Applied Science, 2008:8:1926-1931.

[19] Chihaoui M., Elkefi A., Bellil W. and Ben A.C., A survey of 2D face recogvnition techniques. Computers **5**, 21 (2016).

[20] Priyadarsini M., Jasmine P. and Murugesan K. "Efficient Face Recognition in Video by Bit Planes Slicing", Journal of Computer Science, 2012:8 (1): 26-30.

[21] Christophe CP., Eric G., Robert S. and Dmitry OG. "Detector Ensembles for Face Recognition in Video Surveillance", IJCNN,http://www.informatik.uni-trier.de/~ley/db/conf/ijcnn/ijcnn2012.html - PaganoGSG12 2012:1-8.

[22] Zhao Q., Bin L. and Fu D. "Combination of Improved PCA and LDA for Video-based Face Recognition", Journal of Computational Information Systems, 2013: 9(1): 273–280.

[23] MageshKumar C., Ragul G. and Thiyagarajan R. "Gabor Featured Statistical Modeling In Face Recognition with Chaotic Database", International Journal of Emerging Technology and Advanced Engineering, 2013:3(1).

[24] Aryaz B., Jonathan W and Majid A. "An Efficient Illumination Invariant Face Recognition Framework via Illumination Enhancement and DD-DTCWT Filtering", Pattern Recognition, 2013:46: 57–72.

[25] Fagbola TM, Olabiyisi SO, Egbetola FI and Oloyede A. Review of Technical Approaches to Face Recognition in Unconstrained Scenes with Varying Pose and Illumination. Federal University, Oye-Ekiti (FUOYE) Journal of Engineering and Technology, Nigeria, 2017:2(1):1-8.

[26] Conrad S., Abbas B., Ting S., Shaokang C., Erik B. and Brian L. "Intelligent CCTV for Mass Transport Security: Challenges and Opportunities for Video and Face Processing", Electronic Letters on Computer Vision and Image Analysis 2007: 6(3):30-41.

[27] Shi L., Song X., Zhang T. and Zhu Y. Histogram-based CRC for 3D-aided pose-invariant face recognition. Sensors. 2019;19:759. doi: 10.3390/s19040759.

[28] Jason T., Jeanette BG, Daniel B, Michael C. and Heather Z. "Person Attribute Search for Large-Area Video Surveillance", Homeland Security Affairs, 2012:5(1).

[29] Anil P. and Jyoti S. "Learning Based Single Frame Image Super-resolution Using Fast Discrete Curvelet Coefficients", International Journal of Image Processing (IJIP), 2012:6 (5): 283-296.

[30] Budi S., Mochamad H. and Mauridhi HP. "Investigation of Super-Resolution using Phase based Image Matching with Function Fitting", Research Journal of Engineering Sciences, 2012:1(3): 38-44.

[31] Hemank L., Ankit S., Mayank V. and Richa S."Face Recognition for Look-Alikes: A Preliminary Study", In Proceedings of International Joint Conference on Biometrics, 2011:87-96.

[32] Oloyede A., Fagbola T., Olabiyisi O., Omidiora E. and Oladosu J. Statistical Evaluation of Emerging Feature Extraction Techniques for Aging-Invariant Face Recognition Systems. Federal University, Oye-Ekiti (FUOYE) Journal of Engineering and Technology, Nigeria, 2017: 2(1): 129-134.

[33] Li L., Peng Y., Qiu G., Sun Z. and Liu S. A survey of virtual sample generation technology for face recognition. Artificial Intelligence Review. **50**, 1 (2018).

[34] Olaleye O., Olabiyisi S., Olaniyan A. and Fagbola T. An Optimized Feature Selection Technique for Email Classification. International Journal of Scientific and Technology Research, France, 2014:3(10): 286-293.

[35] Yu S., Shiguang S., Xilin C. and Wen G. "Hierarchical Ensemble of Global and Local Classifiers for Face Recognition", IEEE Transactions on Image Processing, 2009:18(8):1885-1896.

[36] Rabia J. and Hamid RA "A Survey of Face Recognition Techniques", Journal of Information Processing Systems, 2009:5 (2): 41-67.

[37] Haritha D, Srinivasa RK. and Satyanarayana Ch. "Face Recognition System Using Doubly Truncated Multivariate Gaussian Mixture Model and DCT Coefficients Under Logarithm Domain", International Journal of Image, Graphics and Signal Processing, 2012:10:8-17.

[38] Raghu CN: "Illumination Insensitive Face Recognition Using Gradient Faces": International Journal of Image

Processing and Vision Sciences (IJIPVS) 2012:1(1): 38-44.

[39] Guang YC, Tien DB, Adam K. Illumination invariant face recognition using dual-tree complex wavelet transform in logarithm domain. Sciendo, Journal of Electrical Engineering, 2019:70(2):113–121.

[40] Wang YY, Zheng-Ming L. and Long W. "A Scale Invariant Feature Transform Based Method", Journal of Information Hiding and Multimedia Signal Processing, Ubiquitous International, 2013:4(2).

[41] Fagbola TM., Olabiyisi SO. and Omidiora EO. Combined Iterative Back Projection – Maximum A Posteriori Technique for Reconstructing Low Resolution Surveillance Videos. GESJ: Computer Science and Telecommunications, Georgian Technical University, 2018:53(1), 35-45.

[42] Forczmański P. "Human face detection in thermal images using an ensemble of cascading classifiers," in International Multi-Conference on Advanced Computer Systems. Springer, 2016, 205–215.

[43] Jeremiah RB., Kevin WB, Patrick JF and Soma B. "Face Recognition from Video: A Review", International Journal of Pattern Recognition and Artificial Intelligence, World Scientific Publishing Company, 2012:16(2): 1-56.

[44] Orit KG, Tal H. and Lior W. "The Action Similarity Labeling Challenge", IEEE Transactions on Pattern Analysis and Machine Intelligence, 2012: 34:1-7.

[45] Cognitec Systems. The Face Recognition Company September 2011", Cognitec Systems, Dresden Headquarters, Grossenhainer Str. 101, Tower B, D-01127 Dresden, Germany, 2011.

[46] Phillips J. "Video Challenge Problem Multiple Biometric Grand Challenge: Preliminary Results of 2008, Version 1".

[47] ANSI. Face Recognition Format for Data Interchange, INCITS, 2004:1-385.

[48] Fagbola T., Olabiyisi S. and Adigun A. Hybrid GA-SVM for Efficient Feature Selection in E-mail Classification", Computer Engineering and Intelligent Systems, 2012:3(3): 17-28.

[49] Nisar H. Face Recognition Using Combined Global Local Preserving Projections and Compared with Various Methods", International Journal of Scientific & Engineering Research, 2012: 3(3):1-4.

[50] Chen GY., Bui TD. and Krzyzak A. Illumination Invariant Face Recognition using Dual-Tree Complex Wavelet Transform in Logarithm Domain", Proceedings of the International Conference on Pattern Recognition Artificial Intelligence (ICPRAI) Montreal, QC, Canada, May 14–17, 2018.

[51] Kuldeep SS and Madan L. "Face Recognition Using PCA, LDA and Various Distance Classifiers", Journal of Global Research in Computer Science, 2013:4 (3): 84-92.

[52] Raghavender RJ. "Adaptive Frame Selection for Enhanced Face Recognition in Low-Resolution Videos", Thesis Submitted to the College of Engineering and Mineral Resources at West Virginia University in Partial Fulfillment of the Requirements for the Degree of Master of Science in Electrical Engineering, 2008.