IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received November 9, 2018, accepted November 24, 2018, date of publication November 28, 2018, date of current version December 27, 2018.

Digital Object Identifier 10.1109/ACCESS.2018.2883748

Improving Face Recognition Systems Using a New Image Enhancement Technique, Hybrid Features and the Convolutional Neural Network

MUHTAHIR O. OLOYEDE, (Member, IEEE), GERHARD P. HANCKE, (Fellow, IEEE), AND HERMAN C. MYBURGH (Member, IEEE)
Department of Electrical Electronic and Computer Engineering, University of Pretoria, Pretoria 0002, South Africa
Corresponding author: Muhtahir O. Oloyede (tahir.oloyede@gmail.com)

This work was supported by the Council for Scientific and Industrial Research DST-Interbursary support, ICT, Meraka.

ABSTRACT The performance of most face recognition systems (FRSs) in unconstrained environments is widely noted to be sub-optimal. One reason for this poor performance may be the lack of highly effective image pre-processing approaches, which are typically required before the feature extraction and classification stages. Furthermore, it is noted that only minimal face recognition issues are typically considered in most FRSs, thus limiting the wide applicability of most FRSs in real-life scenarios. Therefore, it is envisaged that installing more effective pre-processing techniques, in addition to selecting the right features for classification, will significantly improve the performance of FRSs. Hence, in this paper, we propose an FRS, which comprises an effective image enhancement technique for face image pre-processing, alongside a new set of hybrid features. Our image enhancement technique adopts the use of a metaheuristic optimization algorithm for effective face image enhancement, irrespective of the conditions in the unconstrained environment. This results in adding more features to the face image so that there is an increase in recognition performance as compared with the original image. The new hybrid feature is introduced in our FRS to improve the classification performance of the state-of-the-art convolutional neural network architectures. Experiments on standard face databases have been carried out to confirm the improvement in the performance of the face recognition system that considers all the constraints in the face database.

INDEX TERMS Face recognition, image enhancement, hybrid features, metaheuristic algorithms, unconstrained environments.

I. INTRODUCTION

Face recognition is an essential biometric system that has achieved awareness globally due to its efficient performance in person identification and verification. This has resulted in its being used in various applications such as access and security, retail, healthcare and human-computer interaction [1]. Face recognition systems has achieved satisfactory performance in normal face conditions, where the trained face image is the same as the probe face image. However, face recognition systems fail in unconstrained environments, where the face of the same individual tends to look different due to certain conditions [2], [3]. These conditions in unconstrained environments include occlusion, lighting conditions, expressions, pose variations, aging and plastic surgery [4]–[6]. This has made researchers in both academia and industry come up with different face recognition approaches in such scenarios. Kakadiaris *et al.* [7] proposed a 3D-2D framework for face recognition where the gallery data consisted of 2D texture and 3D shape data, while the probes were arbitrary 2D images. For normalizing images affected by pose and lightning conditions, a 3D-2D system was designed based on a deformable face model. For recognition, the 2D images were depicted in a normalized image space with the use of the gallery 3D models and landmark-based 3D-2D projection estimation. Experiments performed on the FRGC face database showed that their approach could generalize using various illumination conditions. In [8], Ding and Tao proposed a pose normalization approach for pose-invariant face recognition. First, a 3D facial landmark is projected to every 2D face image that allows for extraction of features regardless of the change in pose. Secondly, an optimal warp is calculated based on the homograph to correct the texture distortion that might have been caused by the changes in pose. The corrected frontal view patches are further used for the recognition process. Experiments on four face databases showed the effectiveness of their approach in constrained environments. Basaran *et al.* [9] proposed a face recognition system using the local Zernike moments that can be used for verification and identification. In their approach, local patches surrounding the landmarks are generated from the complex components derived from the local Zernike moments transformations. The phase magnitude histograms are further created by using the patches to create descriptors for face images. Experiments showed that their approach was robust against selected face recognition conditions.

Furthermore, in Pujol *et al.* [10], a face detection system was designed based on the skin colour segmentation using fuzzy entropy. Their fuzzy system is designed in such a way that each colour tone is assumed to be a fuzzy set. The fuzzy three-partition model is used to calculate the parameters needed for the fuzzy system; then a face detection method is developed to confirm the segmentation results. Though some approaches have been made towards face recognition systems in the wild, it is evident that research is still on-going to resolve these issues.

Of all the proposed techniques that have been reported in the literature, the convolutional neural network has been shown to be more effective in handling these highlighted issues of the face recognition system. As a result, researchers have attempted to use this approach in designing face recognition system models. Li *et al.* [11] proposed a recurrent regression neural network that captures different poses adaptively for face recognition. For face recognition on still images, their approach predicts the images with sequential poses and expects to capture unique information from different poses. Experiments carried out on the MultiPIE dataset displayed the effectiveness of their proposed method. Zhang *et al.* [12] proposed an adaptive convolutional neural network (ACNN) for face recognition. Their method functions by first initializing its network architecture with every layer having a single map feature. After that, this first network is accessed to determine if it is convergent or not in ACNN. If convergent, global expansion will be avoided, and the initial network will be trained to satisfy the predefined system average error; otherwise, the system will be extended by global expansion until the necessary system average error is met. ACNN automatically constructs the network without performance comparison which amounts to easier and less training time when compared with the traditional CNN. An experimental result with the ORL database verifies the practicability and efficiency of the proposed ACCN.

In another related study, Li *et al.* [13] introduced a CNN cascade for detection of the face by considering large visual variations such as expression and lighting. The CNN cascade was proposed to solve the high computational expense attributed to CNN when exhaustively scanning full images on multiple scales. It functions by first evaluating the input image at low resolution before rejecting the regions of non-faces and then processes the challenging regions at a higher resolution for accurate detection. Results obtained indicate improved performance. Rikhtegar *et al.* [14] have proposed a hybrid face recognition method that benefits from the advantage of both support vector machine and convolutional neural network. First, a genetic algorithm is used to search the efficient structure of the CNN; then the system is enhanced by replacing the last layer of CNN with an ensemble of SVM. Their approach, when evaluated on the ORL face database, provided a recognition system that functioned with changes in illumination and pose. Li *et al.* [15] proposed a deep CNN method for cross-age face recognition that combined an identity discrimination network with an age discriminative network. Experiments carried out on face databases such as MORPH and CACD-VS showed the effectiveness of their method. In [16], Hu *et al.* presented facial recognition challenges such as occlusion and illumination variances that affected the recognition performance and proposed a robust two four-layer CNN architecture to solve these challenges. In this architecture, the first CNN is designed for frontal images with occlusion, illumination variances, and facial expressions, while the second CNN is designed for various poses and facial expressions. Experimental results displayed satisfactory results on the LFW face database.

Though the highlighted studies have used the convolutional neural network for major issues of the face recognition system in an unconstrained environment, very few have considered a scenario where all the constraints are present in the database. As a result, their approach applies to minimal issues at a time in the face database. Also, the research works ignore approaches at the pre-processing stage, thus not resulting in optimal performance of the proposed system. Face recognition systems requires that face images be effectively pre-processed before extracting features to increase recognition performance. The main idea of this research work is to confirm the effectiveness of image enhancement on face recognition systems using the CNN. Also, ideal features to be extracted from the enhanced face dataset should be investigated. Hence, the proposed face recognition system has put in place an effective image enhancement technique as a pre-processing approach where a new evaluation function (EF) has been developed, to add more features to the face image.

In this research work, a face recognition system is proposed where all the constraints in the face database are considered. Also, a new selection of hybrid features that consists of the pyramid histogram of gradients (PHOG), edge histogram descriptor (EHD) and local binary pattern (LBP) has been proposed for effective extraction of unique features from the enhanced face dataset. This is presumed to influence the increase in performance of the face recognition system positively. In the literature, there have been several image enhancement techniques such as histogram equalization (HE) [17], image intensity adjustment (IIA), adaptive histogram equalization (AHE) [18] and linear contrast stretching (LCS) that have been proposed for different image processing tasks. Enhancement methods using metaheuristic algorithms have also been proposed by inter alia Ye *et al.* [19] and Munteanu and Rosa [20]. These methods have shown good performance however, limited work has been carried out on face recognition applications. Most of the stated image enhancement techniques either produce an over-enhanced or low-enhanced image that can lead to poorer recognition performance as compared to the original input image. The primary contributions of this research work are as follows:

(i) A new enhancement method has been applied to improve the performance of face recognition systems in unconstrained environments by using state-of-the-art convolutional neural networks.

(ii) A set of effective hybrid features that can be extracted from the enhanced images has been presented to improve the recognition performance.

(iii) Detailed performance analysis has been provided to confirm the effectiveness of the face image enhancement approach to increase recognition performance considering all constraints in the face database.

The organization of the rest of the paper is as follows: Section II presents the proposed face recognition system, where the new image enhancement technique and the new hybrid features are presented. Also, the selected state-of-the-art CNN architectures used are described in this section. Section III describes the public dataset and performance evaluation used in our research work. In section IV, results and discussions are presented based on the experiments carried out. Finally, the conclusion is given in section V.

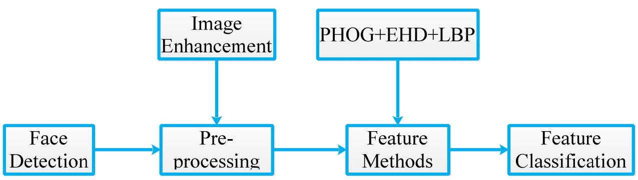


FIGURE 1. The framework of the proposed face recognition system.

II. PROPOSED FACE RECOGNITION SYSTEM

A typical face recognition system is made up of different stages, i.e., face detection, pre-processing, feature extraction and feature classification stages. The face detection stage is the stage where the face is detected from an image. In our research work, datasets used have the face images detected and cropped. Hence, in this section, we propose our face recognition system, where the different components are discussed as shown in Figure 1. First, our proposed face image enhancement method is described, where the various components that it comprises such as the transfer function, our new evaluation function, and the selected metaheuristic algorithm are highlighted. Secondly, we briefly describe our selected hybrid features that are made up of a combination of different features. These methods have been selected based on their unique characteristics which further helps in extracting features from the enhanced images effectively. Finally, the architecture of the convolutional neural network used as the classifier is described.

A. PROPOSED IMAGE ENHANCEMENT TECHNIQUE

The main objective of the image enhancement technique as a pre-processing approach is to come up with an enhanced face image from the original, which is supposed to improve the overall performance of the face recognition system. To achieve this, an effective transformation function should be used that is capable of efficiently mapping the intensity values of the original input image so that it produces an improved output image. Furthermore, the image enhancement technique requires an evaluation function that automatically selects the optimal enhancement parameters, thus producing the most efficient enhanced face image. These components are briefly described in the following sub-sections.

1) TRANSFORMATION FUNCTION

In our research work, we have utilized the transfer function of Munteanu and Rosa [20]. Their function is applied to each pixel of the image at a location (i, j), where a transformation t that uses the grey level intensity of the pixel in the input image $f\ (i, j)$ is converted to the value $g\ (i, j)$, i.e. The grey level intensity in the output face image. The vertical and horizontal dimension of the image is represented as $\mathrm{v}_{\mathrm{size}}$ and $\mathrm{h}_{\mathrm{size}}$ respectively. Hence, the transformation function t is defined in equation (1) below as:

$$g(i,j) = k\left(\frac{M}{\sigma(i,j)+b}\right)[f(i,j) - c.m(i,j)] + m(i,j)^a \quad (1)$$

where, $m(i, j)$ and $\sigma(i, j)$ represent the grey level mean and standard deviation generated for the pixels present in the neighbourhood centred at (i, j). The global mean of the image $M = \sum_{i=0}^{H_{size}-1}\sum_{j=0}^{V_{size}-1} f(i,j)$ Furthermore, a, b, c, and k are the parameters of the enhancement kernel, whose individual values range as follows: $2 \le a \le 2.5$; $0.3 \le b \le 0.5$; $0 \le c \le 3$ and $3 \le k \le 4$.

2) EVALUATION FUNCTION

The primary objective of the EF is to select the appropriate face image automatically without human intervention. A well-designed EF must be able to quantify the qualities of a well-enhanced image [19]. Hence, it is considered a significant component of an effective image enhancement technique. An effective image enhancement technique should be able to quantify the number of edge pixels, the number of pixels in the foreground, entropic measure and the peak-to-signal noise ratio of the enhanced image. The computation of these metrics in relation to our new EF is presented as follows: First, the number of edge pixels, N_g, in the enhanced image is computed. To achieve this, a Sobel threshold, T_f, is automatically computed from the original image, $f(i, j)$, using Sobel's edge detector. This threshold, T_f, is then used in the Sobel edge detector to obtain the edge intensities, $E_g(i, j)$ of the enhanced image. In addition to being invariant, T_f

was considered for computing $E_g(i,j)$ in order to ensure a fair comparison between the original image and the different instances of the enhanced image. Therefore, the number of edge pixels N_g, in the enhanced image is obtained as:

$$N_g = \sum_{i=1}^{H}\sum_{j=1}^{V} E_g(i,j) \tag{2}$$

Secondly, the number of pixels, ϕ_g, belonging to the foreground objects in $g(i,j)$ is computed. To achieve this, the variance, $\vartheta_g(i,j)$, of $g(i,j)$, and the variance, $\vartheta_f(i,j)$, of $f(i,j)$ are computed within a neighbourhood (window) having $n \times n$ pixels. A threshold value, η_f, is automatically computed for $\vartheta_f(i,j)$ using Otsu's threshold algorithm. A representation, $D_g(i,j)$, revealing pixels belonging to the foreground objects in the enhanced image is obtained as:

$$D_g(i,j) = \begin{cases} 1 & \textit{if } \vartheta_g(i,j) \geq \eta_f \\ 0 & \textit{if otherwise} \end{cases}$$
$$\text{for } i = 1, 2, \ldots, H; \quad j = 1, 2, \ldots, V \tag{3}$$

Thus, is obtained as

$$\phi_g = \sum_{i=1}^{H}\sum_{j=1}^{V} D_g(i,j) \tag{4}$$

Thirdly, an entropic measure, β_g, of $g(i,j)$ is computed as

$$\beta_g = \begin{cases} -\sum_m \Omega_m \log(\Omega_m) & \text{for } \Omega_m \neq 0 \\ 0 & \text{for } \Omega_m \neq 0 \end{cases} \tag{5}$$

where Ω_m is the frequency of pixels having grey levels in the histogram bin, $m = 1, \ldots, 256$. The PSNR, ρ_g, of $g(i,j)$ is obtained as

$$\rho_g = 10\log_{10}\left[\frac{(L-1)^2}{MSE}\right] \tag{6}$$

where L is the maximum pixel intensity value in $g(i,j)$ and MSE is given as

$$MSE = \frac{1}{H \times V}\sum_{i=1}^{H}\sum_{j=1}^{V} |f(i,j) - g(i,j)|^2 \tag{7}$$

Based on the parameters computed in equations 2 -7, a new evaluation function, E, is proposed as

$$\mathrm{E} = 1 - \exp\left(-\frac{\rho_g}{100}\right) + \frac{N_g + \phi_g}{H \times V} + \frac{\beta_g}{8} \tag{8}$$

where E is a linear combination of the normalized values of the different metrics described in equations (2 – 7). By normalization, each metric in E is made to have values between 0 and 1. Thus, based on this linear combination, our evaluation function is characterized by a defined scale bounded by a minimum value of zero and the maximum value of four. A minimum value of zero represents an entirely black enhanced image, while a maximum value of four represents an entirely white enhanced image. Furthermore, the

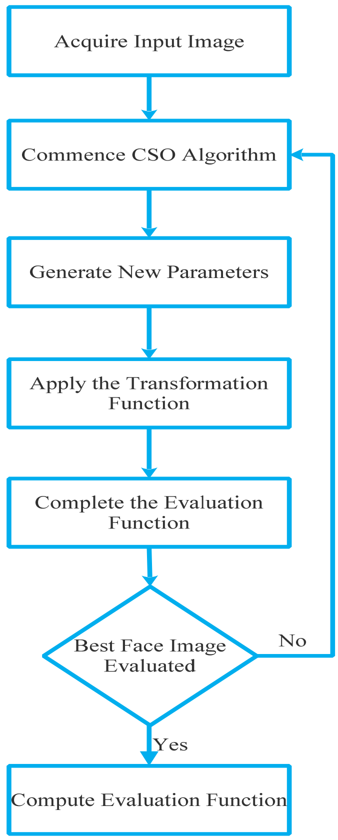


FIGURE 2. Graphical representation of the proposed image enhancement method.

Cuckoo search optimization (CSO) technique has been selected as the ideal metaheuristic algorithm for our research work. The CSO algorithm was introduced by [21], and it is among the most recent metaheuristic algorithms used for global optimization, with a look-alike process of the brood parasitic behavior of certain cuckoo species. There are some optimization algorithms in the literature, however, the CSO has been considered due to its simplicity, fast convergence and its effective capability which has been proven in the literature [22]. Figure 2 furthermore shows a graphical representation of the proposed image enhancement method.

B. SELECTED HYBRID FEATURES

The feature extraction stage of the face recognition system model is crucial as its main objective is to simplify the number of resources that represent a large set of data. It is also the stage where the extraction of unique features present on the face image is carried out. In this research work, our focus on the feature extraction stage should come up with features that will be effective for classification, thus leading to improved recognition. Hence, we present a set of hybrid features that consists of PHOG, EHD, and LBP. The fusing of these three methods which is done in serial form enjoys the benefit of being able to extract unique features from the enhanced face image. The presented selection of features can capture local appearance information, invariance to grey level changes and computational efficiency, and the ability to extract important shape and edge information from the face image. The selected features used in this research work are further briefly described.

1) PYRAMID HISTOGRAM ORIENTATION GRADIENTS

The PHOG is an active feature extraction technique that has been effective in recent years in image classification tasks. It is a spatial pyramid extension of the histogram of gradients, and it is a spatial shape descriptor that represents the image with its local shape and the layout of the shape [23]. The PHOG descriptor is inspired by using pyramid representation; and the histogram of orientation gradients [24]. In this work, the Canny edge detector is first applied to the face image, and it is divided into spatial grids at all pyramid levels. Then a 3×3-Sobel mask is further applied to the edge contours to estimate the orientation gradients. The gradients of each grid are further fused together at each pyramid level.

2) EDGE HISTOGRAM DESCRIPTOR

The Edge Histogram Descriptor (EHD) can describe both texture and shape features that are vital components for content-based image analysis [25]. In our work, the EHD was used to obtain features describing the edges in each image. In this case, each image was converted into its corresponding grey image. The grey image was then divided into 4×4 blocks. The local edge histogram was calculated, and the percentage of pixels that correspond to an edge histogram was obtained. The same procedure was used to obtain the pixels that corresponded to the global edge histogram bin. Both the local and global histogram values were then saved in a feature vector that described each image.

3) LOCAL BINARY PATTERN

Since the initial local binary pattern operator was implemented, it has been regarded as an efficient texture descriptor, used in different applications where it has been shown to be adequately discriminative due to its advantages such as its invariance to monotonic grey-level changes and computational efficiency, thus making it appropriate for challenging image analysis tasks [26]–[28]. In this work, the LBP operator was used to transform each image considered in each respective dataset into an array of integer labels that described the small-scale appearance of each image. The LBP operator typically described the texture associated with each image where a label was given to every pixel. This was achieved by creating a binary image from the greyscale image considering the three by three neighbourhood of each pixel. The LBP operator was used to label centre points, and then the difference between each centre point and the points in the neighbourhood was calculated. For differences greater than zero, a value of one was assigned, while a zero value was assigned for differences of less than zero.

C. FEATURE CLASSIFICATION TECHNIQUE

Feature classification techniques in FR systems are used as classifiers for recognition, which determines the overall performance of the face recognition system. This stage involves both the identification and verification processes where face images of individuals are trained and stored in a database, and a test face image is used to identify or verify an individual. Also, to further carry out the analysis of the different image enhancement methods and the selection of our hybrid features, the CNN method has been selected in our research work. This is due to its recent and effective performance in image processing and computer vision tasks, and its ability to be inspired by biological processes where the connecting patterns amongst neurons look like the arrangement of the human visual cortex [11], [29]. In our work, we adapted the state-of-the-art CNN model architecture in [30], that consists of six layers including the input and the output layers. Other layers include the Convolutional, Rectified Linear Unit (RELU), Pooling, and the Fully-Connected layers.

An array of numbers representing the input features was considered in the input layer while producing another set of arrays of numbers as output. The Input layer holds the $[1 \times Q]$ raw feature values of the face images used in our work. The number of features, Q, typically changes based on the type of features used. The Convolutional layer consists of 12 filters, which results in a volume of $[1 \times Q \times 12]$. The RELU layer uses a max (0, x) thresholding at zero, thus leaving the size of the volume unchanged as $[1 \times Q \times 12]$. Down-sampling operations along spatial dimensions are performed in the Pooling layer resulting in a volume size of $[1 \times Q/2 \times 12]$.

The Fully-Connected layer consists of an $[1 \times 1 \times N]$ volume size, where N denotes the number of class scores associated with the different image databases considered in our work. In this way, our CNN architecture typically transforms the input features in a layer-by-layer process to the final class scores. The feature methods used in our work standardly produce different numbers of features such as: PHOG = 630; EHD = 80; LBP=755; PHOG+EHD=710; PHOG+LBP=1386; and PHOG+LBP+EHD=1467. Also, taking into consideration that not all features produced will be relevant, we use the information gain ranking method to select the most important features produced from the different feature methods. Hence, the following Q values were used in the CNN architectures based on each feature: Q = 26 for PHOG; Q = 70 for LBP; Q = 120; PHOG+EHD= 125; PHOG+LBP=150; and PHOG+LBP+EHD=170.

III. DATA SAMPLES AND PERFORMANCE EVALUATION

In this research work, the focus is on improving the recognition performance of face recognition system models in unconstrained environments by putting in place an effective image enhancement method. Also, effective features to be extracted from the enhanced face image need to be identified. Since we are concerned with unconstrained environments, we use standard and public face databases that represent these facial conditions such as AR face database [31] and Yale face database, and the Life in the wild (LFW) face database.

The Yale face database consists of frontal images with varying facial conditions such as lighting conditions (right-light, center-light, and left-light); facial expression (normal, sad, happy, surprised, sleepy and wink). Also, the AR face database consists of frontal images with facial conditions such as lighting conditions (both lights on, left-light on, right-light on); facial occlusion (sunglass and scarf); and facial expression (smile, scream, and anger). The labelled faces in the wild (LFW) face dataset was also used in our research work. They comprised 13233 images of faces obtained from the web and labeled using each subject's name. There are 5749 different persons with 1680 having two or more images in the dataset. Most images represent the various facial constraints of the real-world scenario [32]. To measure performance, the average recognition rate was used. This represents the percentage of the correctly recognized face image of a subject.

IV. EXPERIMENTAL RESULTS AND DISCUSSION

This section discusses the results obtained from the experiments carried out in our research work. First, a presentation of the enhanced face images by the different image enhancement methods is displayed. This is done to qualitatively analyse the output face images produced by the different image enhancement methods. The AR face database has been used for this. An experiment is carried out to determine the selection of our proposed hybrid features. Thereafter, experiments are carried out on the different enhanced face datasets to confirm the effect of our enhancement method on face images using the state-of-the-art CNN architecture model and the selected hybrid features. Furthermore, experiments are carried out based on the different types of constraint present in the face databases. Experiments are also carried out on the LFW face dataset regardless of the constraints, to further confirm the effect of our enhancement method on face images using an 18-Layer Residual Network (ResNet) state-of-the-art CNN architecture with the selected hybrid features. It is important to note that all the enhanced face datasets produced by the different image enhancement methods have been experimented on using the same selected hybrid features and the CNN architectures. A 10-fold cross validation approach was used in the training/testing process in our experiments. The mean score over the 10-folds is reported as the average recognition rate for each dataset. Our focus in this paper is on improving recognition performance by putting an effective image enhancement technique in place and selecting effective features from the enhanced face images, thus improving the state-of-the-art performance of the selected CNN architectures. The MATLAB 2016b software has been used to design our enhancement method and carry out all the analyses. All experiments were carried out on a computer system running on Windows 7 with the 64-bit operating system. The system consists of random-access memory (RAM) of 16GB and 3.40Hz Intel Core i7-4790 CPU @ 3.60GHz.

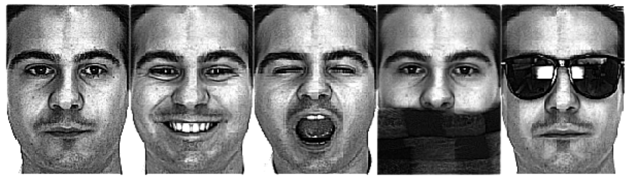


FIGURE 3. Enhanced facial images using the proposed method.

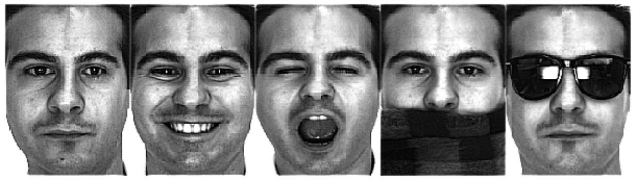


FIGURE 4. Enhanced facial images using the MP method.

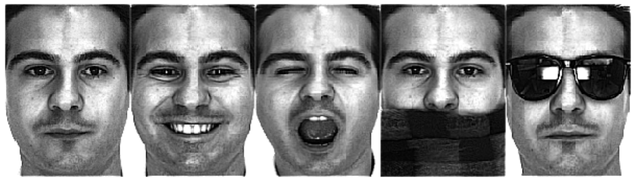


FIGURE 5. Enhanced facial images using the MG method.

FIGURE 6. Enhanced facial images using the IIA method.

FIGURE 7. Enhanced facial images using the AHE method.

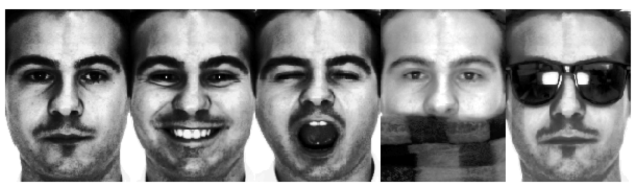


FIGURE 8. Enhanced facial images using the HE method.

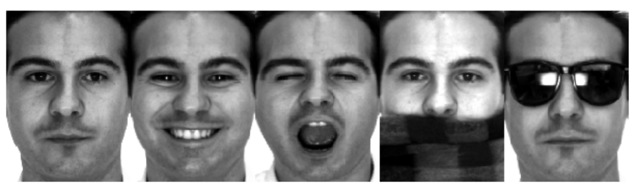


FIGURE 9. Enhanced facial images using the LCS method.

TABLE 1. Performance of proposed selection of features ON AR.

	Feature methods					
Method	Phog	EHD	LBP	Phog+EHD	Phog+LBP	PLE
UnEh	83.1	62.6	45.5	85.6	87.4	89.4
LCS	88.7	65.1	50.1	89.7	90.5	91.8
HE[17]	87.5	65.4	50.5	90.5	91.7	92.4
AHE[18]	88.4	67.6	52.2	92.1	93.3	94.8
IIA	87.9	66.5	45.8	89.1	90.8	92.6
MP[20]	89.4	68.3	57.5	93.3	94.1	94.9
MG[20]	89.7	67.5	56.5	92.3	93.8	94.5
Prop.	92.3	70.6	60.3	95.4	97.6	98.4

UnEh – Unenhanced; LCS – Linear contrast stretching; HE – Histogram equalization; AHE – Adaptive histogram equalization; IIA – Image intensity adjustment; MP – Munteau's method with PSO; MG – Munteau's method with genetic algorithm; Prop – Proposed method, PLE – Phog+EHD+LBP.

TABLE 2. Performance of proposed selection of features on yale.

	Feature methods					
Method	Phog	EHD	LBP	Phog+EHD	Phog+LBP	PLE
UnEh	79.4	56.1	71.3	83.5	80.1	86.9
LCS	81.3	59.8	80.6	86.1	82.4	87.5
HE[17]	82.8	60.1	76.4	87.7	83.6	86.4
AHE[18]	83.2	60.8	80.6	88.3	84.2	87.5
IIA	81.5	59.5	80	86.5	81.3	88.2
MP[20]	85.7	63.4	76.9	90.3	85.9	90.8
MG[20]	85.2	63.5	75.8	90.8	86.4	92.3
Prop.	87.5	65.5	80.4	92.7	88.6	94.6

A. ENHANCED FACE IMAGES FROM THE DIFFERENT ENHANCEMENT METHODS

In this result section, we present the enhanced face images produced using the different image enhancement methods. Figures 3 – 9 show samples of the different enhanced face images of a subject from the AR face dataset using the different enhancement methods.

B. CHOICE OF PROPOSED HYBRID FEATURES

This section presents the results from experiments done to confirm the effectiveness of our selected hybrid features with the different image enhancement methods using the CNN classifier. In achieving this, we use all the data samples in the individual face database, regardless of the facial constraints. The performance of our proposed enhancement method and the different selected features on the AR and Yale face databases are shown in Table 1 and 2 respectively.

Various individual and hybrid features have been selected for our research work. The selected features have been used for our research work after an extensive experiment with other methods which performed poorly in the enhanced face datasets. The features selected include PHOG, LBP, and EHD. In table 1, when the individual features were used, the Phog method performed best on all types of enhanced method. With the unenhanced image, an average recognition rate of 83.1% was achieved, while a recognition rate of 92.3% was achieved by our proposed enhancement method. The LBP feature extraction method achieved least with an average recognition rate of 45.5% on the unenhanced image; 50.5 % on the HE method, 57.5 % on the MP method and 60.3 % on our proposed enhancement method.

When the individual methods were combined, i.e., Phog and EHD, there was a slight increase in recognition performance compared to when the individual feature extraction methods were used. On the unenhanced images, an average recognition rate of 85.6% was achieved; while on the LCS, HE MG and MP methods, an average recognition rate of 89.7%, 90.5%, 92.3%, and 93.3% were achieved respectively. The recognition performance of this selection of features on our proposed enhanced method was 95.4%. Also, the combination of the Phog and LBP feature extraction methods performed even better across all enhanced and the unenhanced image datasets. This led to the proposed selection of hybrid features where all the methods were combined to extract essential features effectively from the enhanced images.

The proposed method, which is a combination of the Phog, EHD, and LBP, interestingly performed better as regards the recognition rate. On the unenhanced image, an average recognition rate of 89.4% was achieved, while a recognition rate of 91.8%, 94.8%, 92.6%, and 94.9% was achieved with the LCS, AHE, IIA and MG enhancement methods. Finally, the proposed feature extraction method performed best with our enhancement method with an average recognition rate of 98.4%.

In table 2 as shown above, a similar experiment has been carried out on the Yale face database. Unlike the AR face database, the LBP outperformed the EHD feature extraction method across all types of enhanced dataset types. For instance, on the LCS enhanced dataset, average recognition performance of 80.6% was achieved with the LBP method as compared to 59.8% with the EHD technique. Also, on the MG enhanced dataset, average recognition performance of 75.8% was achieved with the LBP method, while an average recognition performance of 63.5% was achieved with the EHD technique. This is due to the change in the dataset, where the Yale face database has more background in the images. Also, the combination of Phog and EHD outperformed that of Phog and LBP across all types of image. This shows that there is inconsistency in the performance recognition rate of these methods. However, our proposed feature extraction method performed best on all enhanced and unenhanced datasets. With the unenhanced image, an average recognition rate of 86.9% was achieved, 87.5% was achieved with the AHE enhancement method, 92.3% with the MG enhancement method and finally, 94.6% with our proposed enhancement method. The consistency in performance of our selected hybrid features has shown its ability to extract features effectively from enhanced images to increase recognition performance using the CNN classification method.

C. RECOGNITION PERFORMANCE BASED ON CONSTRAINTS

In this result section, we confirm the performance of our enhancement method on different constraints present in the face database when our proposed selected features and the selected CNN classification method are used. To achieve this, the different constraints present in each database have been grouped, i.e., lighting condition, expression, and occlusion from the AR face database. Also, lightning conditions and expressions have been grouped from the Yale face database. Figures (10-12) and figures (13-14) further detail the results generated based on different constraints from the AR and Yale face databases respectively.

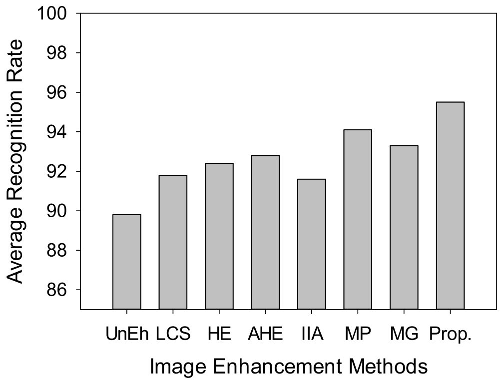


FIGURE 10. **Average recognition performance based on lighting.**

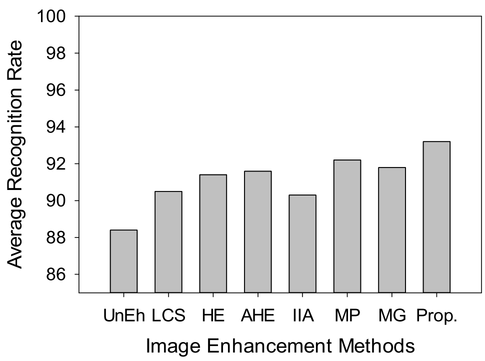


FIGURE 11. **Average recognition performance based on expression.**

1) BASED ON AR DATASET

Figure 10 shows the performance of the different enhancement methods and the selected hybrid features on the CNN classifier, taking into consideration the issue of lighting conditions. It is seen that performance improved when the different enhanced datasets are used as compared to the unenhanced dataset. With the unenhanced dataset, an average recognition rate of 89.8% was achieved while with the LCS, HE, MP and MG average recognition rates of 91.8%, 92.4%, 94.1%, and 93.3% were achieved respectively. With our enhanced method, the highest recognition performance was achieved with 95.5%.

Similarly, for the issue of expression as shown in figure 11, the recognition performance was seen to be the lowest with the unenhanced dataset with an accuracy of 88.4%, while

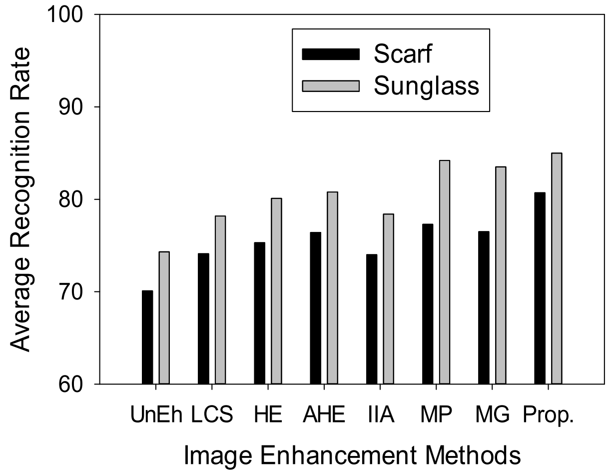


FIGURE 12. **Average recognition performance based on occlusion.**

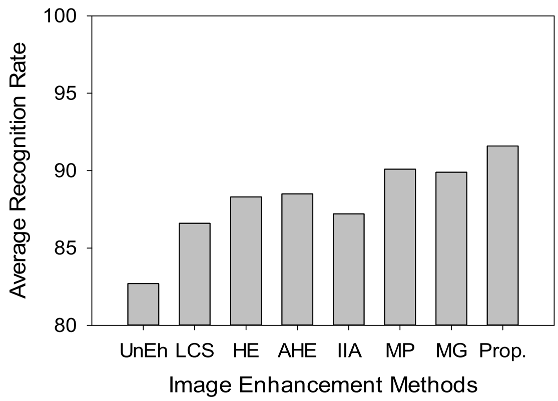


FIGURE 13. **Average recognition performance based on lighting.**

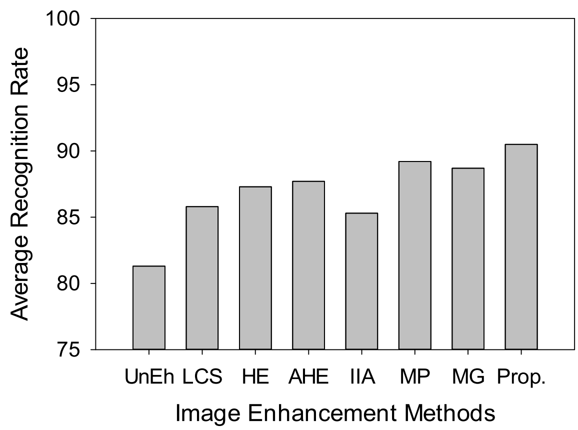


FIGURE 14. **Average recognition performance based on expression.**

performance improved with the different enhancement methods. For instance, average recognition rates of 91.8%, 92.3%, 93.5%, 90.8% were achieved with the LCS, HE, AHE and MP enhanced datasets respectively. Furthermore, our enhanced dataset produced the highest recognition performance with an accuracy of 93.2%.

TABLE 3. Performance of the 18-layer resnet CNN architecture on the LFW dataset.

	Feature methods					
Method	Phog	EHD	LBP	Phog+EHD	Phog+LBP	PLE
UnEh	82.4	65.2	75.6	85.7	82.3	88.9
LCS	84.8	68.8	84.2	88.1	84.6	89.2
HE[17]	85.4	71.3	79.8	89.5	85.8	88.5
AHE[18]	86.2	72.5	84.4	90.3	86.2	89.1
IIA	84.5	68.5	83.7	89.1	83.4	90.5
MP[20]	88.7	75.2	80.2	92.3	87.7	92.5
MG[20]	88.2	75.4	79.6	92.8	88.4	94.3
Prop.	89.8	77.5	84.5	94.5	90.2	**96.7**

Concerning the issue of occlusion, results were based on both the upper and lower types of face occlusion, i.e., sunglasses and scarf respectively as shown in figure 12. Results with the upper face occlusion show that there is a reduction in the performance of the face recognition system. However, the highest recognition accuracy was achieved with our enhanced face dataset. A performance recognition accuracy of 85.0% was achieved as compared to 74.3% produced when the unenhanced face dataset was used. Performance accuracy with the other enhanced methods equally outperformed the unenhanced face dataset but did not reach the level of performance of our enhanced face dataset. Similarly, for the lower face occlusion, the recognition performance went even lower. An average recognition rate of 70.1% was achieved with the unenhanced face dataset. The highest recognition rate achieved was with our enhanced face dataset with 80.66%, which outperformed the other enhanced face dataset. This clearly shows the effectiveness of our enhanced method; however, the issue of the lower face occlusion reduces the performance of our face recognition approach.

2) BASED ON YALE DATASET

Figure 13 shows results generated from our face recognition system approach on the different enhanced images based on lighting conditions in the Yale face database. An average recognition rate of 82.7% was achieved on the unenhanced face dataset, while 91.6% was achieved on our enhanced face dataset. This indicates an increase in recognition accuracy using our enhanced method. Also, other enhanced methods achieved better results as compared to the unenhanced face dataset, for example, average recognition rates of 86.4%, 87.6%, 88.4%, 86% were achieved on the LCS, HE, AHE and IIA enhanced dataset respectively. This shows that none performed better as compared to our enhancement method.

On the issue of expression on the Yale face database, figure 14, shows the experimental results. On the unenhanced face dataset, there is a slight increase as compared to the issue of lighting with an average recognition rate of 84.4%. Also, average recognition rates of 86.8%, 88.3%, 88.5% and 90.1% were achieved on the face datasets of LCS, HE, AHE and MP enhancement methods. Finally, an average recognition rate of 93.8% was achieved on the face datasets when our image enhancement method was used.

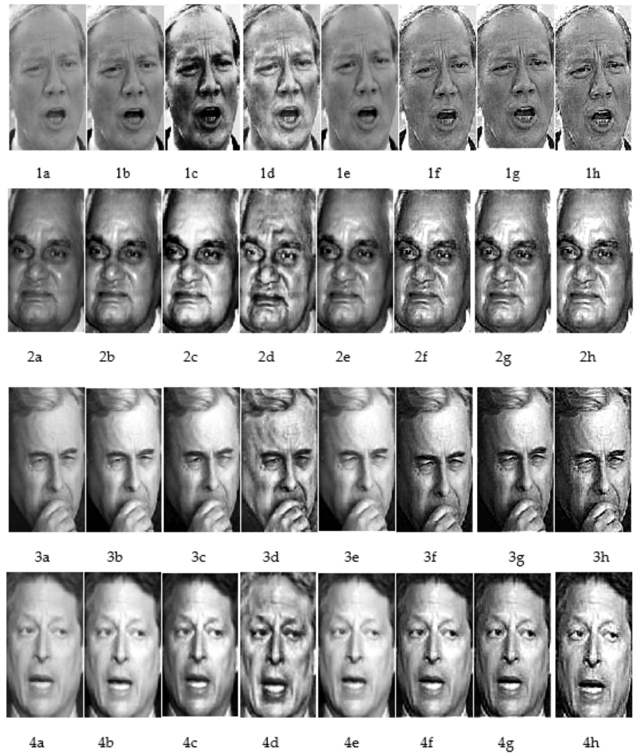


FIGURE 15. Qualitative comparison of the different image enhancement algorithms on the LFW face dataset where figures 1 – 4 represent images of different subjects respectively; and a – h denote the methods labelled (a) original, (b) LCS (c) HE (d) AHE (e) IIA (f) MP (g) MG (h) Proposed.

D. RECOGNITION PERFORMANCE USING THE STATE-OF-THE-ART CNN ARCHITECTURE ON THE LFW DATASET

In this result section, we further confirm the performance of our approach to the LFW face database using our new enhancement method; proposed selected hybrid features, and an 18-Layer Residual Network (ResNet) state-of-the-art CNN architecture described in [33]. ResNet CNN architecture was considered because of its high-level performance in the recent ImageNet Large Scale Visual Recognition Competition (ILSVRC). We considered the LFW face database for this experiment due to the various facial conditions present in the dataset. As in section 4.2, we have made use of all the facial conditions of selected subjects in the face database. We have selected subjects who have four or more facial images, and, which vary amongst the different facial constraints such as occlusion expression, pose and lighting conditions. We used 2110 LFW images in total for the results reported in Table 3. These comprise subjects having five different images with different facial constraints. Hence, 422 subjects were selected with five different images per subject, which resulted in the 2110 images used in our experiments. Figure 15 displays the enhanced face images of four different subjects by using the proposed and different enhancement methods.

As shown in Table 3, the results from the experiments carried out on the LFW face database confirm the effectiveness of our proposed approach using an 18-layer ResNET state-of-the-art CNN architecture. It is evident that our proposed enhancement method with the presented hybrid selection of features outperforms other approaches. With the LCS method, an average recognition performance rate of 65.2% was achieved with the EHD feature, while our enhancement method achieved an average recognition rate of 77.5%. With the MP method, an average recognition performance rate of 92.3% was achieved with the Phog+EHD features, while with our enhancement method there was a slight increase in performance with an average recognition rate of 92.5%. For the proposed enhancement method, with the Phog+EHD features, an average recognition rate of 94.5% was achieved; with the Phog+LBP features, an average recognition rate of 90.2% was obtained. While for the presented hybrid feature, i.e., Phog+LBP+EHD an average performance recognition rate of 96.7% was achieved. This experiment has further confirmed the effectiveness of our enhancement method, and the right selection of features to be extracted from the enhanced face dataset.

V. CONCLUSION

In this research work, we have come up with a face recognition system model in unconstrained environments that comprises a new enhancement method at the pre-processing stage and a new selection of hybrid features capable of efficiently extracting features from the enhanced image. This research work has shown that putting an effective enhancement technique in place as a pre-processing approach, increases performance as compared to using the unenhanced face images. Also, it has been shown that there is a significant increase in the recognition rate when our enhancement method is used as compared to other enhancement methods with two state-of-the-art CNN classification methods. Furthermore, the selection of our hybrid features from the enhanced face images has been shown to have an impact on the increase in recognition performance. The issue of lower face occlusion significantly reduces the performance of the face recognition system model as compared to other facial conditions. Hence, as future work, we intend to investigate approaches to increase the recognition rate in such conditions. Also, we intend to consider the effect of enhancement of other facial conditions in unconstrained environments such as plastic surgery and aging.

REFERENCES

[1] M. O. Oloyede and G. P. Hancke, "Unimodal and multimodal biometric sensing systems: A review," *IEEE Access*, vol. 4, pp. 7532–7555, 2016.

[2] N. Dagnes, E. Vezzetti, F. Marcolin, and S. Tornincasa, "Occlusion detection and restoration techniques for 3D face recognition: A literature review," *Mach. Vis. Appl.*, vol. 29, no. 5, pp. 789–813, 2018.

[3] A. B. Sargano, P. Angelov, and Z. Habib, "A comprehensive review on handcrafted and learning-based action representation approaches for human activity recognition," *Appl. Sci.*, vol. 7, no. 1, p. 110, 2017.

[4] F. D. Guillén-Gámez, I. García-Magariño, J. Bravo-Agapito, R. Lacuesta, and J. Lloret, "A proposal to improve the authentication process in m-health environments," *IEEE Access*, vol. 5, pp. 22530–22544, 2017.

[5] F. Marcolin and E. Vezzetti, "Novel descriptors for geometrical 3D face analysis," *Multimedia Tools Appl.*, vol. 76, no. 12, pp. 13805–13834, 2017.

[6] M. O. Oloyede, G. P. Hancke, and N. Kapileswar, "Evaluating the effect of occlusion in face recognition systems," in *Proc. IEEE AFRICON*, Sep. 2017, pp. 1547–1551.

[7] I. A. Kakadiaris *et al.*, "3D-2D face recognition with pose and illumination normalization," *Comput. Vis. Image Understand.*, vol. 154, pp. 137–151, Jan. 2017.

[8] C. Ding and D. Tao, "Pose-invariant face recognition with homography-based normalization," *Pattern Recognit.*, vol. 66, pp. 144–152, Jun. 2017.

[9] E. Basaran, M. Gökmen, and M. E. Kamasak, "An efficient multiscale scheme using local zernike moments for face recognition," *Appl. Sci.*, vol. 8, no. 5, p. 827, 2018.

[10] F. A. Pujol, M. Pujol, A. Jimeno-Morenilla, and M. J. Pujol, "Face detection based on skin color segmentation using fuzzy entropy," *Entropy*, vol. 19, no. 1, p. 26, 2017.

[11] Y. Li, W. Zheng, Z. Cui, and T. Zhang, "Face recognition based on recurrent regression neural network," *Neurocomputing*, vol. 297, pp. 50–58, Jul. 2018.

[12] Y. Zhang, Y. Lu, H. Wu, C. Wen, and C. Ge, "Face occlusion detection using cascaded convolutional neural network," in *Proc. China Conf. Biometric Recognit.* Chengdu, China: Springer, 2016, pp. 720–727.

[13] H. Li, Z. Lin, X. Shen, J. Brandt, and G. Hua, "A convolutional neural network cascade for face detection," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2015, pp. 5325–5334.

[14] A. Rikhtegar, M. Pooyan, and M. T. Manzuri-Shalmani, "Genetic algorithm-optimised structure of convolutional neural network for face recognition applications," *IET Comput. Vis.*, vol. 10, no. 6, pp. 559–566, Sep. 2016.

[15] H. Li, H. Hu, and C. Yip, "Age-related factor guided joint task modeling convolutional neural network for cross-age face recognition," *IEEE Trans. Inf. Forensics Security*, vol. 13, no. 9, pp. 2383–2392, Sep. 2018.

[16] G. Hu *et al.*, "When face recognition meets with deep learning: An evaluation of convolutional neural networks for face recognition," in *Proc. IEEE Int. Conf. Comput. Vis. Workshops*, Dec. 2015, pp. 384–392.

[17] M. Shakeri, M. Dezfoulian, H. Khotanlou, A. H. Barati, and Y. Masoumi, "Image contrast enhancement using fuzzy clustering with adaptive cluster parameter and sub-histogram equalization," *Digit. Signal Process.*, vol. 62, pp. 224–237, Mar. 2017.

[18] Y. Chang, C. Jung, P. Ke, H. Song, and J. Hwang, "Automatic contrast-limited adaptive histogram equalization with dual gamma correction," *IEEE Access*, vol. 6, pp. 11782–11792, 2018.

[19] Z. Ye, M. Wang, Z. Hu, and W. Liu, "An adaptive image enhancement technique by combining cuckoo search and particle swarm optimization algorithm," *Comput. Intell. Neurosci.*, vol. 13, Jan. 2015, Art. no. 825398.

[20] C. Munteanu and A. Rosa, "Gray-scale image enhancement as an automatic process driven by evolution," *IEEE Trans. Syst., Man, Cybern. B. Cybern.*, vol. 34, no. 2, pp. 1292–1298, Apr. 2004.

[21] X.-S. Yang and S. Deb, "Cuckoo Search via Lévy flights," in *Proc. IEEE Int. Conf. World Congr. Nature Biologically Inspired Comput. (NaBIC)*, Dec. 2009, pp. 210–214.

[22] B. Yang, J. Miao, Z. Fan, J. Long, and X. Liu, "Modified cuckoo search algorithm for the optimal placement of actuators problem," *Appl. Soft Comput.*, vol. 67, pp. 48–60, Jun. 2018.

[23] U. R. Acharya *et al.*, "Automated screening tool for dry and wet age-related macular degeneration (ARMD) using pyramid of histogram of oriented gradients (PHOG) and nonlinear features," *J. Comput. Sci.*, vol. 20, pp. 41–51, May 2017.

[24] A. Dhall, A. Asthana, R. Goecke, and T. Gedeon, "Emotion recognition using PHOG and LPQ features," in *Proc. IEEE Int. Conf. Autom. Face Gesture Recognit. Workshops*, Mar. 2011, pp. 878–883.

[25] C. Turan and K.-M. Lam, "Histogram-based Local descriptors for facial expression recognition (FER): A comprehensive study," *J. Vis. Commun. Image Represent.*, vol. 55, pp. 331–341, Aug. 2018.

[26] L. Liu, P. Fieguth, Y. Guo, X. Wang, and M. Pietikäinen, "Local binary features for texture classification: Taxonomy and experimental study," *Pattern Recognit.*, vol. 62, pp. 135–160, Feb. 2017.

[27] C.-S. Yang and Y.-H. Yang, "Improved local binary pattern for real scene optical character recognition," *Pattern Recognit. Lett.*, vol. 100, pp. 14–21, Dec. 2017.

[28] J. Chen *et al.*, "Robust local features for remote face recognition," *Image Vis. Comput.*, vol. 64, pp. 34–46, Aug. 2017.

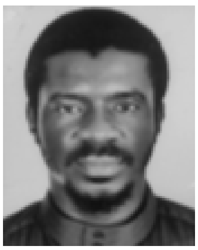
[29] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in *Proc. Adv. Neural Inf. Process. Syst.*, 2012, pp. 1097–1105.

[30] P. Kamencay, M. Benco, T. Mizdos, and R. Radil, ''A new method for face recognition using convolutional neural network,'' *Digit. Image Process. Comput. Graph.*, vol. 15, no. 4, pp. 664–672, 2017.

[31] A. M. Martinez, ''The AR face database,'' CVC Tech. Rep. 24, 1998.

[32] G. B. Huang, M. Mattar, T. Berg, and E. Learned-Miller, ''Labeled faces in the wild: A database forstudying face recognition in unconstrained environments,'' in *Proc. Workshop Faces 'Real-Life'Images, Detection, Alignment, Recognit.*, 2008, pp. 1–15.

[33] K. He, X. Zhang, S. Ren, and J. Sun, ''Deep residual learning for image recognition,'' in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2016, pp. 770–778.



MUHTAHIR O. OLOYEDE (S'15) received the B.Eng. degree in electrical engineering from the University of Ilorin, Nigeria, and the M.Sc. degree in information systems from the University of East London, U.K., in 2011. He is currently pursuing the Ph.D. degree at with the Department of Electrical Electronic and Computer Engineering, University of Pretoria, under the Advanced Sensor Networks Research Group. His research interests include the application of biometric sensing systems, image processing, computer vision, wireless sensors networks, and Internet of Things.

GERHARD P. HANCKE (F'16) received the B.Sc., B.Eng., and M.Eng. degrees from the University of Stellenbosch, South Africa, and the D.Eng. degree from the University of Pretoria, South Africa, in 1983. He is currently with the University of Pretoria, where he is also a Professor. He is recognized internationally as a pioneer and leading Scholar in ASN research, especially aimed at industrial applications. He initiated and co-edited the first Special Section on Industrial Wireless Sensor Networks in the IEEE TRANSACTIONS ON INDUSTRIAL ELECTRONICS in 2009. His papers attracted high citation numbers in high impact journals. He co-edited a textbook *Industrial Wireless Sensor Networks: Applications, Protocols, and Standards*, (2013), the first on the topic. He received the Larry K. Wilson Award (2007), for inspiring membership development and services as a member of several regional and technical conferences worldwide. Apart from serving on several IEEE Committees on Section, Regional, and Board level, some as Chair, he has been very active in the IEEE Technical Activities, in Society administration, and conference organization on many levels, notable as the General Chair of the IEEE ISIE, INDIN, ICIT, and AFRICON.

HERMAN C. MYBURGH graduated in computer engineering from the University of Pretoria in 2008. He received the M.Sc. degree in electronic engineering from the University of Pretoria in 2009 and graduated in 2010 with specialization in wireless communication system. He received the Ph.D. degree in computer engineering from the University of Pretoria in 2013. In his thesis, he designed two novel low complexity joint detector/decoders for the digital mobile receiver. He has presented his work at various international conferences and has published numerous research articles in accredited peer-reviewed journals. His current research interests are in wireless communication systems, sensor fusion, artificial intelligence, and mobile health.