INTERNATIONAL JOURNAL OF COMPUTING, SCIENCE AND NEW TECHNOLOGIES (IJCSNT)

VOL. 1 NO. 2 DECEMBER 2023

ENHANCED CRIME SUSPECT IDENTIFICATION SYSTEM USING DEEP NEURAL VISION PROCESSING TECHNIQUE

by

Odo J.I.[1], E.F. Nworabude[2], Onyia T.C.[3], Obianyo O.R.[4]

1: Dept. of Computer Science, Peter University, Achina-Onneh, Anambra State, Nigeria.

2,3: Dept of Electronic & Computer Engineering, Nnamdi Azikiwe University, Awka, Nigeria.

4: Dept of Computer Engineering, Madonna University of Nigeria, Akpugo Campus, Enugu State, Nigeria.

Abstract

This paper discusses the improved facial recognition system in crime suspect identification system using deep neural vision processing technique. The research was motivated based on the problem of identification of crime suspect which security agents do experience these days. Several models like discrete wavelet transform model, rendering model, convolutional neural network model were used to design the new system which utilizes enhanced facial recognition approach and implemented on a Mathlab environment. The system was tested and validated using tenfold cross validation technique and the accuracy achieved was 99.22% which is very good. The system was later deployed at the Nigerian Police Force and tested for reliability using various facial expressions of volunteered criminals; and the result was excellent.

KEYWORDS/PHRASES: Facial recognition, deep neural vision, biometrics, and rendering model.

I. INTRODUCTION

1.1 Background of Study

Today, in a social domain, the conventional means of verification makes use of data such as identification cards, passwords, user names, etc. to authenticate or verify users. However, according to Khalil et al. (2011), these techniques have demonstrated several disadvantages that biometrics is currently trying to overcome. The conventional solutions mentioned, require the user to store and safely keep the knowledge of their password to prove their identity. But the limitation is that this verification approach is not reliable and can easily be guessed, thus putting the conventional human identification system at risk.

Biometrics as opposed to the aforementioned relies on a physical trait of the user and not on something he/she has, thus avoiding the human risks (Jimenez and Janet 2010). The use of biometric technology has the capacity to help security personal to verify individuals via various traits such as face, iris, finger print, voice, signature among others as discussed in (Neha et al., 2016).

There are number of reasons why it is better to choose face recognition which includes the following among others:

i. It is accurate and allows for high enrolment and verification rates.

ii. It does not require an expert to interpret the recognition result.

iii. Easy to use, cheap to implement and reliable

iv. It requires no physical interaction or contact with the query person.

v. It can use your existing hardware infrastructure; existing cameras and image capturing devices to work without problems.

It is the only biometric that allows you perform passive identification in a one-to-many environments (e.g. identifying a terrorist in a busy Airport terminal).

According to Haider et al. (2020), biometric technology is a system which can verify individual identity based on features like faces, finger print, iris scan, voice, signature, etc. however, among all, face recognition is the holy grail as it has lots of advantages over other biometric traits which includes autonomy, easy to use, accuracy, reliability, versatility, etc. Nevertheless, the technology suffers certain constrain due to the complex nature of human face.

The human face is dynamic as it can be changed with cosmetics, surgery, caricature, aging, etc.

These complexities affect the performance of the various face recognition methods which are geometric based, spatial based, eigen vector based, color based, motion based and artificial intelligence based. However, the use of artificial intelligence provided better results in terms of accuracy compared to the other counterparts (Shang, 2000).

According to (Sharma, 2020) artificial intelligence are system which has the capacity to learn from training dataset or rules and make accurate decisions. The artificial intelligence involves techniques such as expert system, fuzzy logic, genetic algorithm and machine learning with their pros and cons discussed in Mira (2021). From the study it was observed that machine learning techniques with position in artificial neural network provides the best performance compared to others.

Artificial Neural Network (ANN) are set of neurons which has weights, bias and activation function with capacity to learn from training dataset and make accurate decisions (Tejal, 2020). Today this ANN has been advanced into more complex structures called deep learning with configurations such as convolutional neural network, recurrent neural network, etc. These deep learning systems have potentials to make accurate decisions rivaling human intelligence and have been used to develop recent facial recognition system as in (Meenakshi et al., 2019; Michael et al. 2020; Andrew, 2017; Arun and Peter, 2018; Ashish, 2019) and more.

The limitation of these systems was that they were developed using dataset containing already prepared faces such as AT&T, ORL and other, without considering the real nature of face in ideal real live scenario. In practical, faces are not processed and when input for recognition with the conventional system, the results are not reliable. This has undermined the performance of the conventional face recognition system and to the best of the researcher's knowledge today there is no single facial recognition system universally available in the market.

To solve this problem, the researcher proposed the development of novel deep learning algorithm using image processing technique. This when achieved, will be able to process and recognize all kinds of faces irrespective of the image quality and achieve high level of precision.

II. LITERATURE REVIEW

2.1 Review of Past Relevant Literatures

Several researchers have worked on crime identification systems. Praneeth et al. (2020) worked on face recognition system based on deep learning technique. In the study, the features of face were extracted using the binary pattern histogram algorithm and then trained with convolutional neural network. The performance can be improved using image processing technique.

Micheal and Aman (2020) presented a research on face recognition using deep learning technique. In the study, a systematic review was performed, and then convolutional neural network was adopted for the training faces with an accuracy of 97% when deployed and tested on tensor-flow software. The performance was however limited to the training dataset used.

In Saiba and Monjurul (2017), a research on face recognition system was presented using morphological analysis of images. The study employed various techniques such as binarization, morphological dilation, erosion, and then intensity value measurement was used to train data of ORL, Yale and UMIST face dataset. The performance showed accuracy of 80%; but despite the success, the performance can be improved for system reliability.

Marios (2018) researched on a simultaneous low resolution an off-pose angle face matching algorithm as an investigative lead generative tool for law enforcement. The study presented a security system which has the capacity to detect faces of suspected individuals who disguised themselves to commit crime. The system when tested showed good accuracy on multiple datasets, but can be improved using deep learning algorithm.

Nishtha (2019) studied face recognition using deep convolutional neural network. In the research the architecture was developed and used to train ORL dataset which has 40 classes of different peoples. The result when tested showed that the accuracy is 95%; however, the performance when deployed in real time application may not be accurate as the data used to achieve this 95% accuracy is already processed. Hence there is need to incorporate filter to the system to sustain the same level of accuracy when deployed in real life application.

Cmak et al. (2020) researched on face mask detector with Pytorch and Deep learning techniques. The study used OpenCV, artificial neural network (ResNet) and computer vision to develop a system which could detect if a person is wearing face mask. The aim was to prevent the spread of coronavirus. The percentage accuracy when tested was 97%; however, the test was not done with a standard dataset, and hence could not be justified.

Mira (2021) presented a dissertation on face recognition and face mask detection using machine learning technique. In the study, systematic review of various approaches for the recognition of objects was discussed, and then machine learning was singled out and used for the design of the proposed system. The performance was tested using face mask person, and also for face recognition and the result was good. However, there is need for improvement using better dataset and also image processing to achieve system reliability.

Nuruz et al. (2017) presented a research on face recognition for criminal identification. In the study, principle component analysis was used to train a face recognition system and then tested with data collected from CCTV camera. The performance presented an accuracy of 80%; however, despite the success, there is need for improvement.

Andrew (2020) presented a deep learning critical approach for face recognition analysis. In the study, convolutional neural network was developed and used to train a face recognition system which was proposed to solve the problem of occlusion, illumination and pose invariance. The performance when tested showed that the use of deep learning performed better than other shallow solutions, but can be further optimized using image processing.

Arun and Peter (2018) developed a face recognition system using machine leaning. The study used histogram-oriented gradient, multi class support vector machine to develop a face classification system and then deploy for facial recognition. The system was tested with the AT&T dataset, ORL dataset and Yale dataset. The performance showed good recognition result. The work however could still be improved using deep learning.

Bindushree and Rakshitha (2020) presented a research on face recognition using deep learning. In the study, convolutional neural network was developed and used to train a simple dataset of 21 people and then deployed for their face recognition. The system when tested showed that deep learning algorithm was reliable to solve pattern recognition problem.

III. MATERIALS AND METHODS

3.1 Materials

The materials used for the development of the proposed system are laptop, dataset, HP camera, filter, Deep Learning tool, excel software, and among others. The laptop was used for the installation of the software after it was developed. The HP camera was computer vision configured and was used for data collection and face detection, the filter was the tool used for image processing, Deep learning tool was used for the training and recognition result, the excel software was used for data analysis and validation of the system after testing.

3.2 Methods

The methods used for the development of the proposed system are data collection, image acquisition, computer vision, face detection, data pre-processing, data processing, training and face recognition.

3.2.1 Data collection

Artificial intelligence approach remains the most reliable means of solving pattern recognition problem, due to its ability to learn from training dataset and then makes accurate decisions. In the proposed system, the case was the same as data was collected and used to train AI in order to achieve the desired objectives.

The primary source of data collection was from the Nigerian Police Force (NPF). The secondary sources of data collection were from research, oral interviews, emails, investigations, etc; and these were all employed to collect relevant data.

Consultation of magazines, newspapers and library were carried out and data were collected; from which references to relevant documentation which aided the development of the proposed system was identified and used.

3.2.2 Data Description

The dataset provided by the NPF has a sample size of 1850 different suspected identities apprehended for various reasons and awaiting trials. The data were already sized and formatted according to their specification of the NPF as presented in table 1 and stored for official use.

Table 1: Data properties

Image parameters	Values
Dimension	190 x 260
Height	260
Weight	190
Horizontal Resolution	96dpi
Vertical resolution	118dpi
Bit depth	24
Size	5.6kb

The table 1 presents the properties of which all the images were formatted and stored in the station for investigation and other official purposes. Some data samples of the data collected are presented in figure 1.

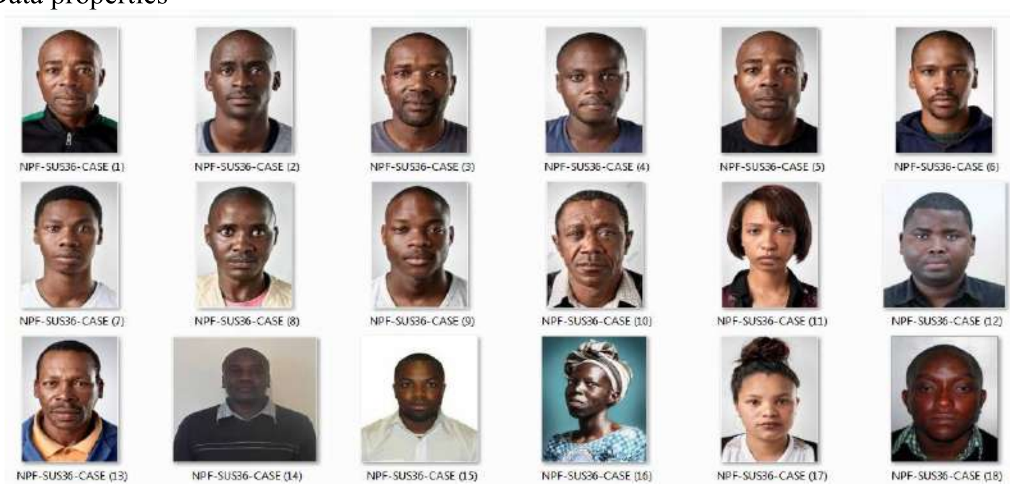

Figure 1: Sampled data from the NPF

3.2.3 Image acquisition

This is an action taken to retrieve or capture an image from a source using a hardware device such as camera. This process of image acquisition can be used for many other applications, but in this case of face detection, it is a high-definition camera that was used and also recommended for high quality image, which enhances the accuracy of the recognition process. However, due to the high cost of HD camera which might not be available at all the Central Police Stations or at the site of crime to capture and collect real time data, computer vision was used to improve the quality of the camera data as discussed in the next method.

3.2.4 Computer vision

A camera can be likened to the eye of a young child who can see without a common sense of the objected in focus. This camera, irrespective of its specifications and qualities, lacks common sense of the data it acquires in time series and as a result, captures unnecessary data which induced more noise and irrelevant features to the main image needed. This problem was addressed using computer vision method. This is a process where the camera is incorporated with intelligence which enables it to derive useful information from a digital image before capturing them. There are various computer vision algorithms in (Girshick et al., 2014) which can achieve this objective today. However, the adoption of any algorithm is dependent on the type of data under investigation, which in this case is a human face.

This study adopted the Viola Jones algorithm which enables camera to not just detect face but also tracks it in real time and collects quality data. According to Vilmal (2015), the algorithm has many advantages, enables via its sophisticated and invariant detector which locates the scales of human facial features. This algorithm was reconfigured to be compatible with the specifications presented in table 1. The aim is to ensure that any image captured in time series are of the same size and properties. The data flow model of the reconfigured Viola Jones algorithm is presented in figure 2 as flowchart.

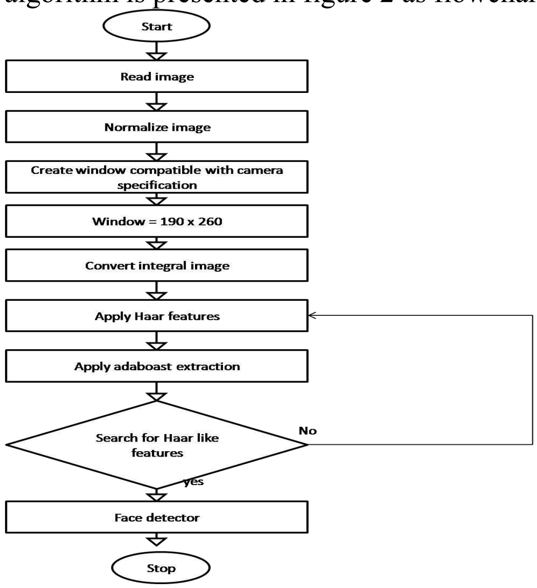

Figure 2: The Configured Viola Jones algorithm

The flowchart in figure 2 presentsd as algorithm shows how it was configured to align with the properties of the recommended camera. The algorithm came into play when the camera reads image data using Haar features and Adaboast algorithms and then recognizes facial judicial points (features) to detect a face.

3.2.5 Face Detection

Face detection simply means the process of identifying and capturing of face using computer vision technology. This was achieved when the camera successfully read a computer vision-based scales image according to Haar features in figure 2 and then acquire the data. In this case the camera, despite its ability to view other objects within its focus, identifies only face like image and captures it.

3.2.6 Data Pre-Processing

Noise is a factor which is inevitable when image is concerned as there are many parameters such as poor camera quality, background, face color, face biers, makeup, among others; which induce noise and affect the quality of data collection via face detection. All these parameters induce various types of corruptions in the query image such as impulse noise, background noise and striped noise as discussed in (Zeenathunisa et al., 2011) with their various characteristics. The background noise which is a reflection of light from the image background and also from camera lens affects the quality of the image and has to be pre-processed first before the other noise types. This is because the features are of unwanted light reflection dominated other types of noise attributed with the image and has to be managed first to make other noise types visible.

The data pre-processing commonly involves removing low-frequency background noise, normalizing the intensity of the individual particles images, removing or enhancing data images prior to computational processing; and this was done using Histogram Equalization Technique (HET) as specified in (Christophe et al., 2007). HET is a contrast adjustment technique developed as a low band pass filter to eliminate unwanted noise attributed with low frequency and was adopted to pre-process the real image.

3.2.7 Data Processing

From the pre-processing stage discussed in the previous section, it was revealed that the histogram equalization technique only mitigated the background noise induced in the image due to adverse illumination from the environment and camera used. The impulse noise and stripped noise which cause severe degradation on the image quality according to Mohd et al. (2012), has to be processed to guarantee quality image result. The aim is to ensure that the most interesting features of the face were made available in good quality for optimized training and recognition result. To achieve this, wavelet filtering and normalization techniques were used.

3.2.8 Wavelet Filtering

The wavelet transform is a process whereby the features of the image are localized in scales while removing noise. The basic idea behind wavelet filtering is to transform and spatially represent image features in the magnitude of wavelet coefficient. Wavelet coefficients are small value of noise which are transformed without affecting the quality of the main image and then reconstructed. This process was used to remove the impulse noise in the image captured while the modeling was presented at the system design section.

3.2.9 Normalization

This was used to vary the changes of pixel intensity values in the image captured. The aim is to convert the image into a range of pixels values which are familiar to ensure that striped noise is removed.

3.2.10 Training

Training a dataset in artificial intelligence is a technique used to learn a machine learning or deep learning algorithm of particular data patterns, and the reference model learning used for time series classification decisions. In this study, the technique was used to train the dataset images collected so as to learn the pattern of each image as a reference convolutional face model and used for time series face recognition. This training process was done using convolutional neural network. The convolutional neural network is a deep learning algorithm which specializes in solving image-based pattern recognition problems and it uses the images to develop a facial recognition system.

3.2.11 Face Recognition

This is the output of the result achieved when the convolutional neural network was used to train and classify a time series image data. The classification was done using the reference convolutional face model and then compares with an input image feature vectors to make a decision based on probability index of the best match.

3.3 Process Diagram

The process diagram was used to present the logical interaction of the various methods discussed to achieve the proposed system as shown in figure 3.

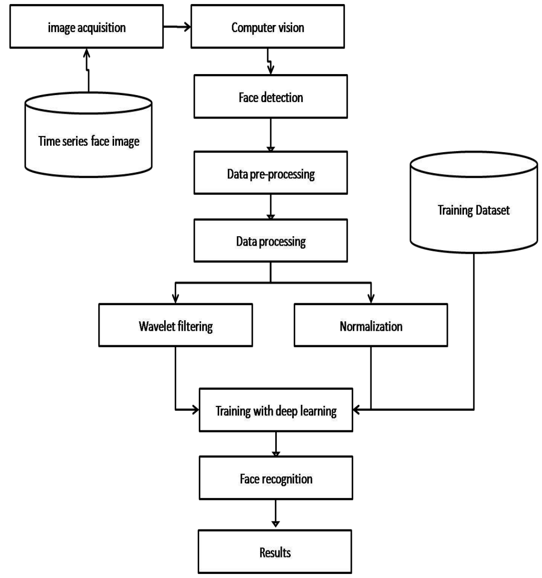

Figure 3: The Process Diagram

The process diagram presents the face detection and recognition system workflow. The diagram shows how the training dataset containing images of suspected persons was trained to generate a reference model, and then use it to recognize time series face images. This was done using image acquisition process which intelligently collects the image of the suspect with the help of computer vision technology. The computer vision was used to make the camera only search for face-like images, based on Haar features and Adaboast algorithm to detect face and capture it. The data captures which is always characterized with noises; such as background noise, strip noise and impulse noise were processed using histogram equalization technique to remove the background noise; wavelet filter to remove the impulse noise and then normalization process to mitigate the strip noise. The processed data was then feed to the deep learning algorithm for training using convolutional neural network which train the input data and compare with the reference model to recognize face.

3.4 System Design

The system design was presented in three main sections which are: the face detection, data processing and face recognition. The design was done using Unified Modeling Language and mathematical models where necessary to present clear explanation and relations between all necessary actors and parameters which contributed to the realization of the new system.

3.4.1 The Face Detection Modeling Diagram

The face detection model shows how the camera was able to read, track and detect only face in a wide varieties of other images within the camera scene. This was achieved using the logical data flow diagram which shows the marriage between the camera device and the configured Viola Jones algorithm in figure 2 to make it intelligent. The configuration was done also considering the standard and requirement of the NPF image set which as specified in table 1.

The face detection model shows how the camera reads image data from the scene and use computer vision to search for Haar features which represents face using facial features like eye, nose and mouth. These features are labeled as judicial points (m) and the process continues until the points satisfy the requirement of face detection is greater or equal to 8 points. The reason for this value of judicial point was based on the study of RaceFaces (2020) which revealed that 8 facial points satisfy the requirement for face detection.

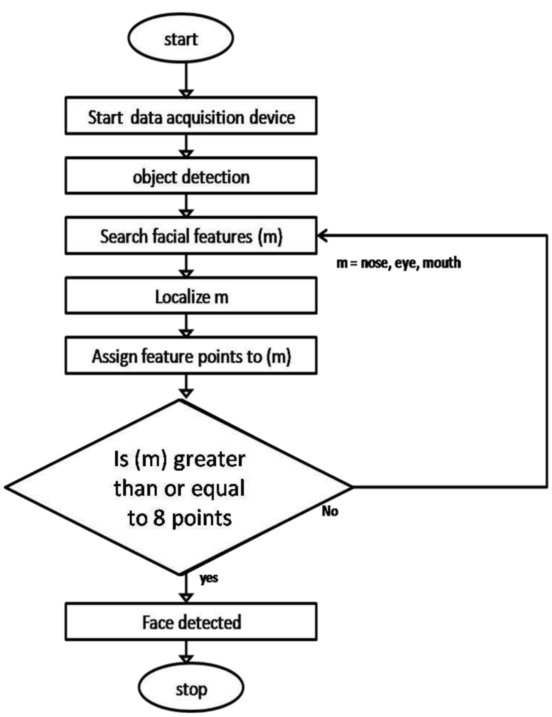

Figure 4: The Face Detection Data Flow Modeling Diagram

The Data Flow Modeling diagram in figure 4 presents the data flow of the facial detection system. The diagram shows the interaction between the camera and the computer vision algorithm to identify and capture a suspected person face for investigation via facial recognition.

3.4.2 Data Processing Model

The data processing model was developed using the Discrete Wave Transform (DWT) which uses three filters simultaneously to process image data and remove noise. The filters are the Haar wavelet, orthogonal filter and bi-orthogonal filters based on Wavelet Decomposition Theorem, which segments the images into various sub band of Low Low (LL), Low High (LH), High Low (HL) and High High (HH). These regions represent the various part of the image, with the major part within the low low pass bands, while the remaining parts like the edges are shared among the other bands.

The DWT which presents a two-dimensional representation of the face data detected and captured is modeled using equation 1.

$$F(u,\ v) = \frac{1}{\sqrt{M}} \sum_s \sum_t W_q(y_o, s_, t) q_{yo,s,t}(u,v) + \frac{1}{\sqrt{M}} \sum_{x=H,V,D} \sum_{y=yx}^{\infty} \sum_s \sum_t W_q^s(y_, s_, t) \psi^s{}_{y,s.t}(u,v) \quad (1)$$

Where the wavelet coefficients are presented as:

$$W_q(y_o, s_, t) = \frac{1}{\sqrt{M}} \sum_{u=0}^{s-r} \sum_{v=o}^{T-1} f(u,v) q_{yx,s,t^{(u,v)}} \quad (2)$$

$$W_q^s(y_o, s_, t) = \frac{1}{\sqrt{M}} \sum_{u=0}^{s-r} \sum_{v=o}^{T-1} f(u,v) q^s{}_{y,s.t}(u,v) \quad (3)$$

x = (H, V, D)

Where M is the power of 2, $\frac{1}{\sqrt{M}}$ is the normalization factor, u and v are the dilation and translation parameters in terms of discrete values, $W_q^s(y_o, s_, t)$ and $W_Q(y_o, s_, t)$ are approximation coefficients, F(u, v) is first scale input, y is scale parameters, s is shift function, T is time reverse scale, t is time domain, f is frequency domain, W or (ψ) *is wavelet function*.

In two dimensional 2D and 3D the scaling function q(u,v) and three 2D wavelets of $\psi^H(u,v), \psi^V(u,v) and\ \psi^D(u,v)$ are used to deduct the wavelet representation of the image as the models below:

q(u,v) (4)

$\psi^{V}(u,v)$ (5)

$\psi^{H}(u,v)$ (6)

$\psi^{D}(u,v)$ (7)

The models in equation 4 is the scale factor used to decompose the wavelet coefficients of the images into various scales as shown in DWT decomposition modeling diagram in figure 5.

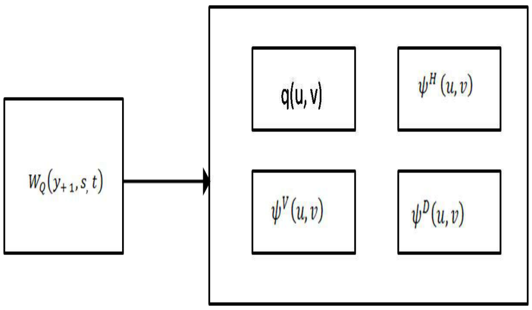

Figure 5: The Wavelet Decomposition Diagram

The figure 5 shows how the DWT was able to scale the face input whose wavelet coefficients are represented with equation (2) and (3) to form the decompositions block modeled with equation 4 which represents the LL sub section of the band containing the main part of the face data and the other equations 5 to 7 which presented the edges and other features of the image as shown in figure 6.

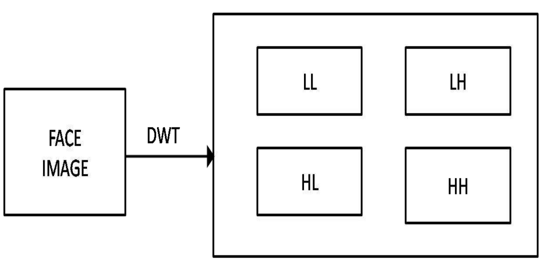

Figure 6: The Scaling of the Face Image with DWT

3.4.3 The Proposed Deep Learning Model

The deep learning model was developed using Convolutional Neural Network (CNN). This is deep learning algorithm which has proven in other areas of application to be the best in solving image-based pattern recognition problem. This was modeled in this research to train the wavelet image processed and to recognize face accurately. The CNN is made up of input layer responsible for the dimensioning of the image into a compatible size and then fed to the convolutional layer for scanning and extraction of the best feature parts, however these high quality of image pixels scanned, pooled and arranged in the last convolutional layer contains noise reflection originating from the high feature pixels which requires normalization before feeding to the fully connected layer. This was done using rendering technique to normalize the pixels behaviors and then fed to the fully connected layer for training using artificial neural network and then predict the output with Softmax function. The block diagram of the CNN is presented in figure 7.

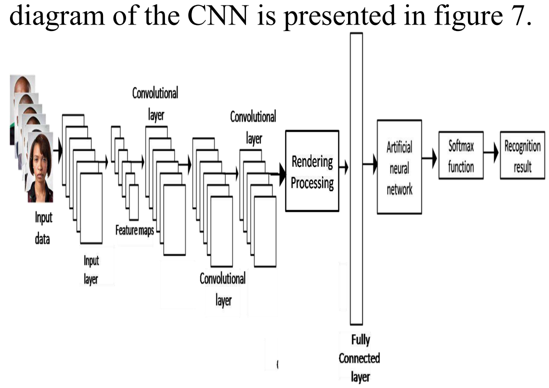

Figure 7: The CNN Block Diagram

3.4.4 Input Layer

This is the first layer of the CNN which collects data from wavelet transform and dimensioned using (h x W x C) where h is the image height, W is the weight and C is the color channel which is 3 for colored image. Hence the dimension of the input is (190 x 260 x 3). This value represents the size of the data that is fed to the convolutional layer. For the dimensioned image to be fed to the convolutional layers, filtering and pooling process were used.

3.4.5 Filtering

Filtering is a process of scanning the various pixels of the dimension image to extract interesting key features of the image in a pixel format. The filter size was specified as:

$$F = (f_w \times f_h \times d) \quad (8)$$

Where f_w is the filter weight, f_h is the filter height and d is color dimension. This filter was used to scan the various sections of the image and extract pixel information from each scan to form a matrix. During the filter scanning process, areas where filter do not cover are padded before scanning to ensure compatibility between the mother image and the filter. The model of the feature maps scanned by the filters is presented in equation 9.

$$F_o = \left[\frac{F_i + 2_p - k}{s}\right] + 1 \quad (9)$$

Where Fo is the output features, Fi is the input features, p is the convolutional padding, s is the strides size, k is the convolutional kernels size. During these feature extraction process, rectified linear unit (ReLU) was used to introduce nonlinearity in the features to transform all negative values into null. The next stage was to pool the features into matrix presentation of a convolutional layer. The pooling method is presented the next section.

3.4.6 Pooling

When feature maps were scanned from the main image using the filter, the scanned pixels were extracted using pooling method. The feature maps are the interesting image pixels collected during scanning and were pooled using the maximum pooling technique which only selects the highest pixel value of array of extracted pixel matrix to form a convolution layer as shown in figure 8.

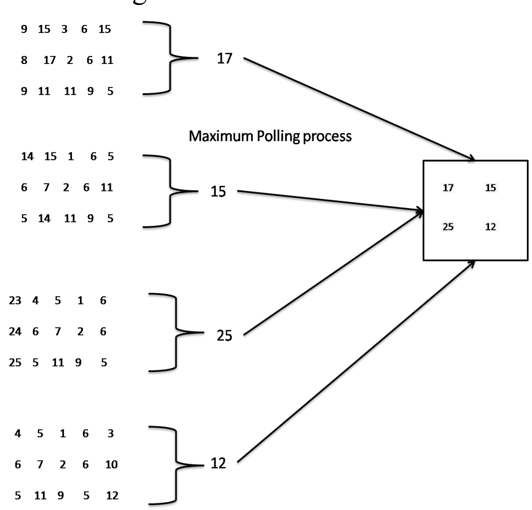

Figure 8: The Maximum Pooling Process

The figure 8 presents some of the output 5 by 3 matrixes of pixels scanned by the 5 by 3 filter determined by equation 3.8. From the output as shown in the matrixes above, the highest values were only selected and used to form the next convolutional layer. The reason is that the pixel with the highest value contains the most important part of that image section while the total output was presented as a convolutional layer as shown in figure 9.

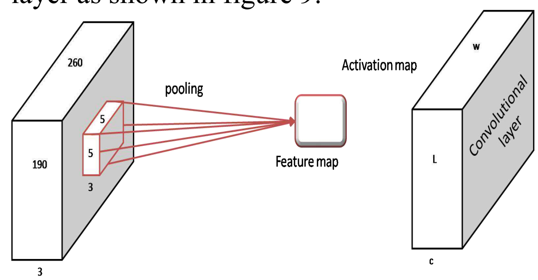

Figure 9: The Convolutional Process

The figure 9 presents the data flow diagram of the convolutional layer setup. This was achieved using the 5 by 3 filter to scan the 190 by 260 face image. The scanned output was presented as feature map and pooled into a matrix of pixels to form a convolutional layer. The total pixel used to form the convolutional layer is presented as equation 10.

$$C_O = ((w * h) + 1)*nf) \quad (10)$$

Where the total pixels in the convolutional layer is given as (C_o), w is filter weight (w), filter height (h), number of filters (nf) and filter bias term (1). The model in equation 10 presents the total pixels values in the first convolutional layer only; however, to generate pixels for other convolutional layer, the number of previous pixels was added (np) and presented as:

$$C_O = ((w * h * np) + 1)*nf) \quad (11)$$

This model was used to sum up all the feature maps extracted per convolutional process in an array of matrix. The activation size is determined using the relationship between the number of image pixels and the depth as shown in equation 12.

$$A_s = (w * h * d) \quad (12)$$

In this model of equation 12, the parameters of w, h and d are defined based on each layer specifications.

3.4.7 Rendering Model

Rendering is an image process approach which is applied in addressing many limitations of other image processing technique. According to Kajiya (1986), the process describes and manages the effect of radiance reflected from an integral convolution of image using geometric approximation. In equation 11 which described an array of convolutional matrix representing the best features of the face image, the pooling technique and the convolutional process of CNN ensured that the best features of the image were extracted and represented using in matrix arrays, however this process after completion in the last convolutional layer induced certain degree of light intensity due to the very high quality of the image in terms of pixels, and this is not recommended to feed to the convolutional layer directly for training in order that a reliable recognition result will be achieved.

Secondly due to the series of convolutional process performed, and the pooling technique which is selective to only extract the best part of the image, this do not necessary mean that other parts are not useful either, hence the rendering process was also used to address this issue based on the law of conservation of energy as discussed in (Owen, 1999). The rendering model is presented using the relationship between the radiance reflected from each image pixel and reflection distribution factor from all angles of feature points as presented in equation 13.

d $$L_r(x,y) = f_r(x,w,y) L_f(x,w) \cos\theta dw \quad (13)$$

Where x is the total reflectance from each image pixel, w is the convolutional angle, and θ is the angle of the feature point, dw is the differential angle of the pixel, L_f is the surface, y is the direction of radiance, L_r is the surface radiance.

The rendering model in equation 13 was applied after the last convolutional layer of the CNN considering the emitted radiance from the image pixel and presented as an integral rendered convolutional image as equation 14.

$$L_s(x,y) = L_e(x,y) + \int f_r(x,w,y) L_f(x,w) \cos\theta dw \quad (14)$$

Where radiance on the convolutional image at the last convolutional layer is L_s, radiance emitted from the image feature is presented as L_e, The equation 14 can be remodeled as equation 15.

$$L_s(x,y) = L_e(x,y) + \int g(x,k) f_r(x,w,y) L_f(x,w) \frac{\cos\theta \cos\theta^i dA}{||k-x||^2} \quad (15)$$

Where $||k-x||^2$ is the distance from k to x; $\cos\theta^i = (n.w)$; $L_s(x,y) = L_e(x,y)$ and

$$g(x,k) = \begin{bmatrix} 1 & if\ x\ is\ visible\ to\ k \\ 0 & otherwise \end{bmatrix}$$

3.4.8 Fully Connected Layer

This section of the CNN flattens the rendered convolutional image in equation 15 and then fed to a neural network for training to learn the image features for time series face recognition. The training was enabled using back propagation algorithm in figure 10, and was monitored for accuracy to ensure that there is no overshoot performance using the loss function model in equation 16.

$$L = \sum_{i=l}^{K} (P_i - D_i)^2 \quad (16)$$

Where L is the loss function, k is the number of observations, P is prediction and D is the training target. The model presents the loss between P and D, where K is the number of class in the face dataset and i is the output loss which is scalar. The Deep Learning settings are shown in Table 2.

Table 2 Deep Learning Settings

Maximum number of Epoch to train	15
Epoch between display	1
Maximum time to train in sec	Infinity
Maximum validation failure	5
Scale factor for length	190
Scale factor for weight	260
Initial step size	0.01
Minimum performance gradient	1e-6
Cost horizon	7
Control horizon	2
Number of bias functions	1
Number of Channel	3

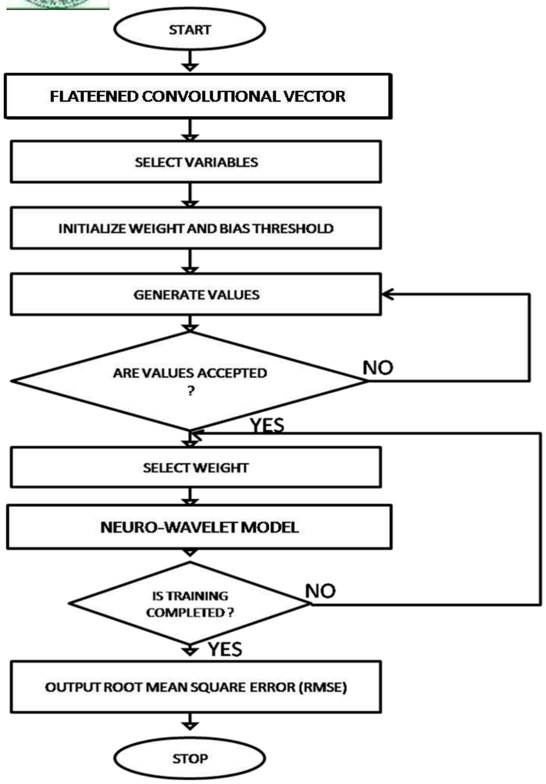

Figure 10: The Back Propagation Algorithm

3.4.9 Output Layer for Recognition of Result

This is the final layer of the network which produces the desired output of the training process. This layer is designed using a Softmax activation function which transforms the learned feature vectors into probability distributions consisting of various probabilities proportional to the various exponential of the face input data. The model of the Softmax function is presented in equation 17.

$$\sigma(\vec{z})_i = \frac{e^{z_i}}{\sum_{j=1}^{k} e^{z_j}} \quad (17)$$

Where σ is the softmax function; e^{z_i} is the standard exponential function for input vectors from the ANN output, k is the number of classes of the multiset classifiers; e^{z_j} is the standard exponential function for output vector, $\vec{z}$ is the input vector.

An improved CNN modeling diagram is shown in figure 11.

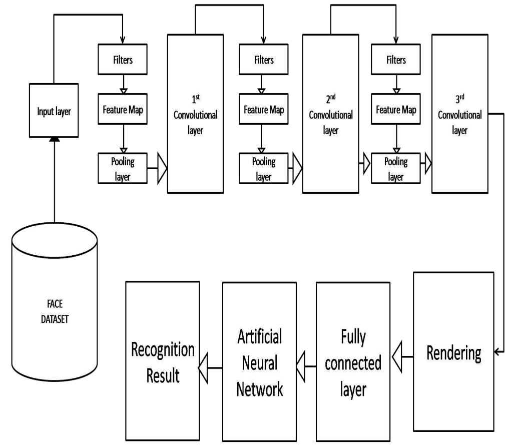

Figure 11: An Improved CNN Modeling Diagram

The figure 11 presents the model of the novel convolutional neural network developed with the potential for précised facial recognition result. From the diagram, the face dataset is fed to the CNN via the input layer for dimensioning and then scanning using filter and pooling layer. The outputs are arranged in matrix array to form a convolutional layer and then the scanning process was repeated until the third convolutional layer. At this stage the rendering process was introduced for normalization before feeding to the fully connected layer which uses artificial neural network to train the data for time series facial recognition purposes. The system flow chart is presented in figure 12.

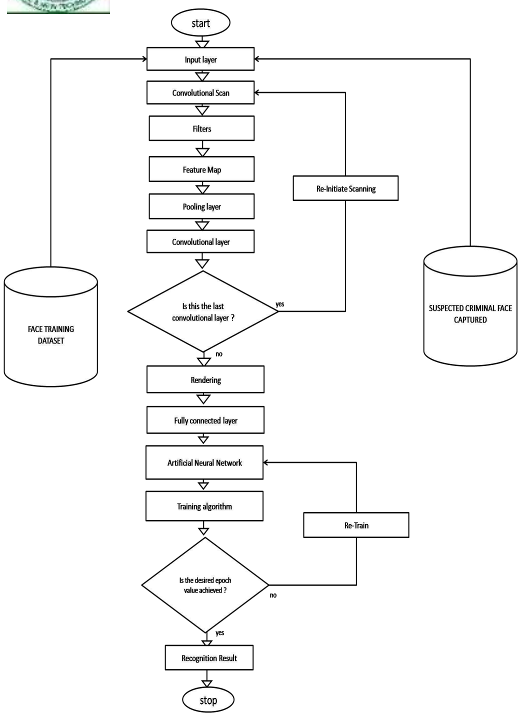

Figure 12: System Flowchart

The system flowchart presents the data flow activity of the new system developed, showing how the training and test data was processed by the system to make accurate recognition process. From the flowchart, the convolutional neural network is first trained and used to recognize faces of suspected and captured person. The convolutional neural network model was developed using rendering approach which normalizes the convoluted series of high value pixels and then used artificial neural network to train the data for recognition result.

3.5 The System Implementation

The face recognition system developed was implemented using the mathematical and modeling diagrams developed, neural network toolbox, deep learning toolbox, image processing toolbox, wavelet toolbox, signal processing toolbox, image acquisition toolbox and then deployed with Matlab User Interface software as a criminal face detection and verification system as shown in figure 13.

Figure 13: The Facial Recognition System

3.6 Use Case and Domain Analysis

Use Case Unified Modeling Languages are deployed in research to explain in details the main components of the requirement definition. They describe the activity through which the system will satisfy the aforesaid functional requirements. The Use Cases would then be deployed in constructing the process model which explains the operations (user action) in a more formal manner. The process uses diagrams to document an object-based decomposition of systems showing the interaction between these objects and the dynamics of these objects.

The researcher's aim here is to provide a common vocabulary of object-based terms and diagramming techniques that is rich enough to model any system development project from analysis to design. Use case diagrams give a user point of view of the system with different users referred to as the Actors.

In the User Case of face recognition diagram shown in figure 14, the actors include: police, new officers, existing officers, user; while the supporting actors include the suspect, query image.

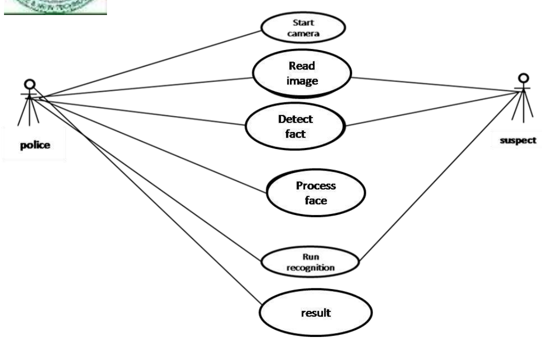

Figure 14: Use Case Diagram of the Face Detection and Recognition System

Primary Actor: Police
Secondary Actor: Suspect
Brief Description of Event: From the modeling diagram above the police agent starts the camera which reads and captures the suspected face. The face detected is processed to remove impurities and then trained for recognition.
Pre-condition: It is assumed that the initial suspected image of crime suspect is already in database.
Post-condition: When recognition process occurs, a match is expected to be achieved as verification result.
Main flow of Events:

1. The use case diagram starts with the surveillance video streaming
2. The camera reads the face image using computer vision
3. Face detection
4. Data processing
5. Recognition

The Use Case for new users is shown in figure 15.

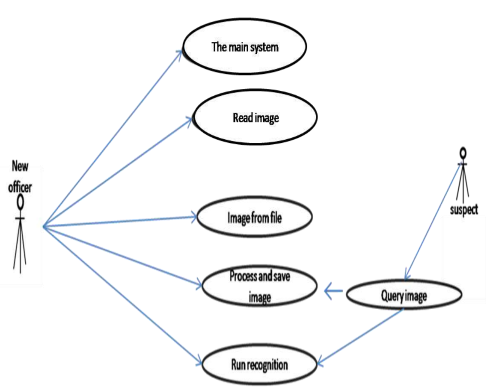

Figure 15: Use Case for New Users

Brief description: The investigating officer accesses the main system, and then uses the camera to upload query images and then save for recognition.
Primary Actor: New officer
Secondary Actor: Suspect
Pre-condition: We assume that the initial suspected image of crime suspect was captured and stored already in database.
Post-condition: When recognition process occurs, a match is expected to be achieved as verification result.
Main flow of Events:

1. The new officer gained access to the software
2. The software reads the image from the stored location
3. Image is processed
4. The recognition process is initiated to compare both images.

The Use Case for an existing user is shown in figure 16.

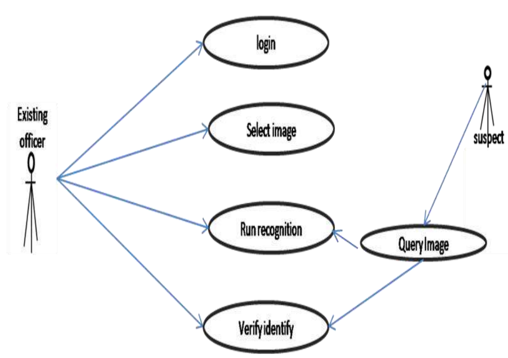

Figure 16: Use Case for Existing User

Brief description: The existing officers login and select image from file for investigation and match with the suspected query image for recognition.
Primary actor: Existing officer
Secondary actor: Suspect
Pre-condition: We assume that the initial suspected image of crime suspect was captured and stored already in the database.
Post-condition: When recognition process occurs, a match is expected to be achieved as verification result.
Main flow of events:

1. The existing officer gains access into the main system.
2. The already captured and saved image is selected from file for investigation.
3. The suspect's image is loaded for recognition.
4. The query image is processed.
5. The recognition process is initiated to compare both images.

The Use Case Diagram for image acquisition is shown in figure 17
.

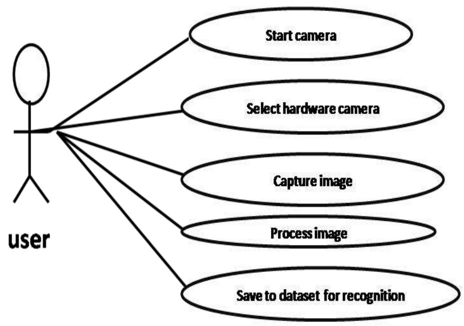

Figure 17: Use Case Diagram for Image Acquisition

Brief Description: The use case diagram in figure 17 is used to model the user processes for image acquisition with the new system. First, the image acquisition tool (camera) is started and camera captures the image. The image is processed and saved for recognition purpose.

Primary Actor: User

Pre-condition: The camera has already been installed and connected to the system.

Main flow of Events:

1. The camera is started.
2. The particular one for surveillance is selected.
3. The suspect's picture is captured (query image).
4. The query image is processed.
5. The processed image is saved to dataset or file.

Alternative Flows

1. The face can be detected without the eyes open.
2. The image is processed.
3. It is then ready for recognition.

Key scenario

1. Face recognition without iris analysis.

Post-Condition

1. The image is processed and compared with a query image for recognition.

IV RESULTS AND DISCUSSION

4.1 Result of the CNN Model

The CNN was evaluated using the accuracy and loss function as modeled in equation 17. The reason for these two performance evaluation metrics was to measure the reliability of the model developed before integrating for crime investigation via facial recognition system. Since false alarm was among the major challenge which affected the reliability of existing system, the evaluation measured is a means to examine the performance of the model to address the existing problem of false alarm. In the training process, the data was divided by the deep learning toolbox into training and test set. The training data was used to learn the CNN model of the training dataset features, while the test set was used to check the probability for correct face classification. This process was performed iteratively at different epoch, and when the result achieved has higher loss function, not approximately zero, a back propagation approach was used to return and adjust the neurons in the fully connected layer, and the process continues until the CNN learn the face training dataset, then the result validated as shown in the figure 18.

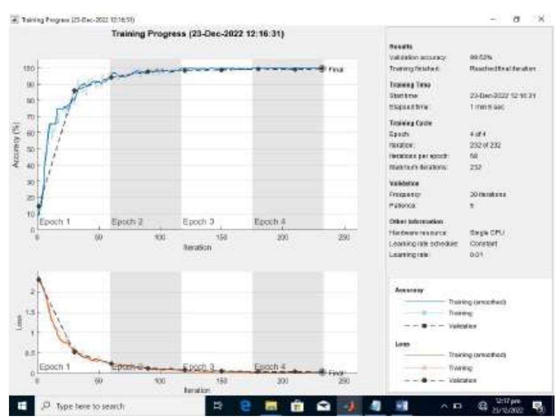

Figure 18: The Results of the CNN Training

The figure 18 presented the training performance of the CNN model developed using accuracy and loss parameters. From the result the testing process occurred within 6 epochs which has 50 iteration step intervals. This means that 50 iterations took place between each epoch. From the result, the accuracy of the CNN classification is 99.52%, while the loss function is 0.48%. The implication of the result is that the CNN correctly classified the training face data 99.52 times out of 100 testing times. The 0.48 also presented the percentage wrong classification, however because the percentage of wrong classification is approximately zero, it shows that the error can be tolerated and that the CNN model for face verification is reliable.

4.2 Result of the CNN model Integrated as Facial Recognition System

The idea of this section is to check the practical implication of the CNN model for face recognition and digital crime management and then ensure that the same good result achieved from the validation process was replicated with the new solution developed. To test the face recognition system developed with the CNN model, a query face image was captured and used to test the system as shown in the figure 18.

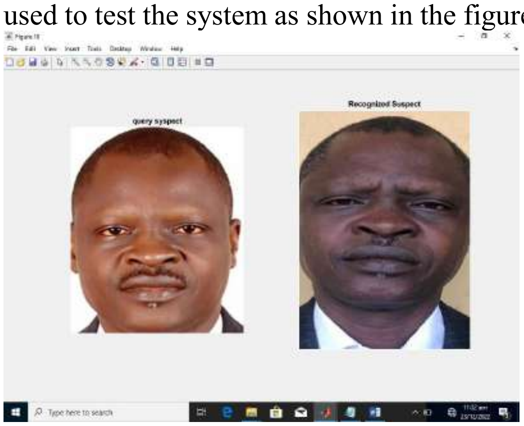

Figure 18: Result of Suspected Query Person

The figure 18 presented the query face which was used to test the face recognition system. The data was capture using the face detection model in the figure 3. The model used the process of computer vision to track the face in the image data, via Viola and Jones approach which used Haar and Adaboost features to search for face like features in image frame and then captured as shown in the figure 18.

Having captured the face image, we recall that the aim of the study is for crime management using facial recognition. Crimes are always committed in the open space (outdoor environment) which are attributed to many factors that affects quality of data captured such as the impact of sunlight, dust, atmospheric condition, etc, which all directly induce noise on the image captured.

Zeenathunisa et al. (2011) added that various noises in form of impulse noise, background noise and striped noise all add noise to the face data. To process the data and mitigate the noise, histogram equalization was used to transform the image as in figure 19.

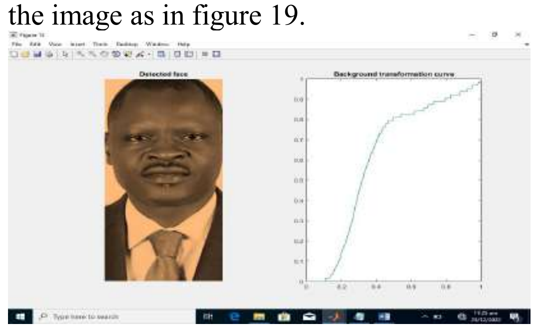

Figure 19: The Data Pre-processing Result with Histogram Equalization

The figure 20 shows another result of face recognition output. It used another face of an individual to evaluate the performance output of the software.

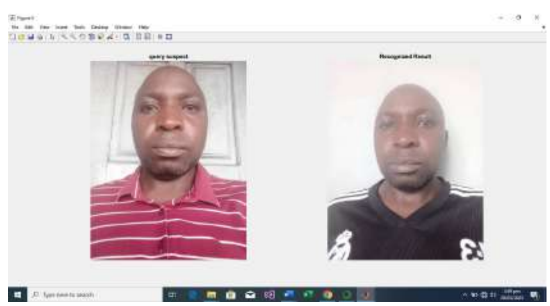

Figure 20: Another Result of Face Recognition Output

From the result of the figure 4.9 and 4.10 respectively, the performance of the face recognition software was evaluated considering other query images and the result showed that the software was able to correctly recognize the query image from the dataset, using the CNN-based face recognition model.

V. SUMMARY AND CONCLUSION

5.1 Summary

Facial recognition system is not only an easy way to capture criminals during crime investigation, but also the easiest means to identify and verify a person's identity without any reasonable doubt. However, to achieve this standard requires precision in system design, which was not fully recognized in the conventional system as reviewed; hence it makes them lack reliability. This research was able to solve this problem using an improved deep learning algorithm which employed advance and necessary steps with techniques to develop a system which can verify someone's identify via facial recognition with high accuracy.

5.2 Conclusion

"Unknown gunmen" has remained a familiar issue covering the Nigerian news on daily basis. Most recently, criminals hide under this terminology to commit all forms of atrocities which have cost lots of lives and labeled bad name to the country. To solve this problem, this research has presented a reliable facial recognition system, developed with the best algorithm to solve image classification problems. This was implemented with high level programming language and deployed for used in the Nigerian Police Station after it has been tested and validated. The result showed reliability, easy to use, affordability and has the capacity to recognize accurately every face of a criminal suspect. The performance was also compared with existing state of the art algorithms and it was observed that the new system has a percentage improvement of 0.52% which is good among other advantages it has over the existing systems.

REFERENCES

Ana Betrand (2012) "Face localization and facial feature extraction based on shape and color segmentation". 16th IEEE international conference on pattern recognition; Volume 2, Issue 2, ISSN 2091-2730

Andrew Enughwure (2015); "A Combination Approach to Face Recognition, "Bishop University. Volume 1, Issue 4 46(2):235–245.

Arun Alvappillai and Peter Neal Barrina (2018)" Face Recognition using Machine Learning" IEEE Computer Society Conference on Computer Vision and Pattern Recognition.

Calister Obi (2015) "Challenges of Insecurity and Terrorism in Nigeria: Implication for National Development" OIDA International Journal of Sustainable Development, Vol. 08, No. 02, pp. 11-18,8 Pages Posted: 29.

Girshick, J. Donahue, T. Darrell, and J. Malik (2014) "Rich feature hierarchies for accurate object detection and semantic segmentation". In CVPR.

Haider Mehraj and Ajaz Hussain Mir (2020) " A Survey of Biometric Recognition Using Deep Learning" EAI Endorsed Transactions on Energy Web 03 2021 – 05; Volume 8; Issue 33.

Karina Enriquez (2018)" Faster Face Detection using Convolutional Neural Networks & the Viola-Jones Algorithm" California State University Stanislaus, 1 University Circle, Turlock, CA 95382

Levine, S., Finn, C., Darrell, T. & Abbeel, P. (2016). "End-to-end training of deep visuomotor policies," J. Mach. Learning Res., vol. 17, no. 39, pp. 1–40.

Lin S.D, Chen K. (2019) "Illumination InvariantThermal Face Recognition using Convolutional Neural Network. IEEE International Conference on Consumer Electronics -Asia (ICCE-Asia);

Liu X, Guo Z, You J, Kumar B. (2020) "Dependency-Aware Attention Control for Image Set-Based Face Recognition". IEEE Transactions on Information Forensics and Security.

Mohd B., Sandeep S., Nelson K (2012), " Survey on Noise Removal in Digital Images" IOSRJCE; ISSN 2278-0661; vol 6; Issue 4; PP 45-51

Prajyot H. Mohite, Pratibha P. Shingare (2020)" Deep Learning Approach For Facial Expression Recognition" Journal of critical reviews;ISSN- 2394-5125 Vol 7, Issue 14.

Russell, S. and Norvig, P. (2010). Artificial Intelligence: A Modern Approach, 3rd Edition. New Jersey, Prentice Hall

Vimal, Virender Kadyan (2015) "Implementation and Performance Analysis of Face Recognition Using MATLAB. IJCST Vol. 6, Iss ue 2, ISSN : 0976-8491 (Online) | ISSN : 2229-4333.

Wei T. Jing X. Y., H. S. Wong, D. Zhang, (April 2006.) Face recognition based on 2D Fisherface approach, Pattern Recognition, 39, 4, pp. 707–710,

Yogita Mahajan, Shanta Sondur (2018)"Aging face recognition using deep learning" International Journal of Engineering and Applied Sciences (IJEAS) ISSN: 2394-3661, Volume-5, Issue-8.

Zeenathunisa S, Jaya A., Rabbani M., (2011) "A biometric approach towards recognizing face in various dark illumination", IEEE, PP1-7.

