Machine Learning Based Digital Recognition of Identical Twins to Support Global Crime Investigation

Asogwa Tochukwu Chijindu[1], Ituma Chinagolum[2]
[1]*Computer Science, Enugu State University of Science and Technology, Enugu, Nigeria*
[2]*Computer Science, Ebonyi State University Abakaliki, Nigeria*

Abstract: **This paper presents the application of machine learning for digital recognition of identical twins. The main aim here is to present a novel episteme that will differentiate two similar suspected faces of different identities using bag of features. The process includes face detection, extracting bag of features among other image processing techniques, training, classification and result prediction. Some existing bibliographies will be reviewed, analyzed and a new system will be proposed. This work will highly improve criminal forensic investigation, eliminate impersonation, and stop mistake identity arrest of suspect to mention a few among other importance and applications. The implementation and result will be demonstrated employing mathlab development tool.**

Keywords: **face detection, bag of features, training, classification**

I. INTRODUCTION

The events of 11 September 2001 have made it painfully clear even to anuninformed observer that we have entered a period of history where statesand even superpowers can be challenged in unorthodox ways [1]. Repeatedly in our daily times, national news and social media, we hear about the crimes of credit card fraudulent acts, armed robbery, impersonation, computer breaking's by hackers, security breaches in a company or government building to mention a few. In most of these cases, the criminals took advantage of the setback in the conventional access control systems which do not grant access by "who we are", but rather by "what we have", such as ID cards, keys, passwords, emails, username, PIN numbers, or mother's maiden name, however none of these means really define us. In other words if someone steals, duplicates, or acquires this identity, means he or she will be able to access our data or our personal property any time they want, and as a result we find ourselves in a world where there is an increased rate of insecurity continuously affecting both the developed, developing and the under developed habitat of humanity.The security of our habitant cannot be waved aside by a flip of hands as government at all levels is struggling day and night to curb these challenges that are now technologically driven.According to [2] any socio-economic development of a person, organization or a nation depends not only on its ability to provide a sustained security of information, lives and properties which is vital to its continued existence but also on its ability to strategically generate competitive intelligence through surveillance, as the need to maintain and secure information or physical property is becoming both increasingly important and challenging.

Recently, Biometric technology became available to allow recognition and verification of "true" individual identity through biometric systems to help support criminal investigation. This system refers to technologies that analyze and measure human body characteristics features for security applications [3]. These features include fingerprints, palm, DNA, voice patterns, irises, hand measurements and facial patterns, which are called biometric features.

Biometric access control are automated methods of verifying or recognizing the identity of a living person on the basis of some physiological characteristics, such as fingerprints or facial features, or some aspects of the person's behavior, like his/her handwriting style or keystroke patterns [4].Since biometric systems identify a person by biological characteristics, they are difficult to forge. Among the various biometric identification methods face recognition has been described as the “Holy Grail” of biometric identification systems, due to a number of significant advantages over other methods of identification (as well as the difficulties encountered in the quest to obtain a practical working system). However, with the current state of the art, these advantages do not include operating performance in terms of recognition accuracy. When compared with other identification technologies, face recognition cannot compete with the low error rates achieved using iris or fingerprint systems. It is one of the few biometric methods that possess the merits of both high accuracy and low intrusiveness [5]. It has the accuracy of a physiological approach without being intrusive. For this reason, since the early 70's [5] face recognition has drawn the attention of researchers in fields from security, psychology, and image processing, to computer vision. The major problem yet to be solved despite these series of research works has been to differentiate identical twin with face [3].

There have been many attempts to solve this problem. The early approaches are aimed for gray level images only, view-based detectors are popular in this category, including Rowley's neural networks classifier [6], Sung and Poggio's correlation templates matching scheme based on image invariants [7] and Eigen-face decomposition [8]. Model based detection is another category of face detectors [9] but they all ended up solving the problem of detection and not recognition. [10] came close with their research on face recognition, but their work cannot differentiate twins.

This born out the author's motivation to embark on this research work, providing a new applicable dimension to this trend to differentiate and recognize two similar faces or different persons, employing machine learning techniques to train bag of extracted features from the faces (real and query images) after it has been processed employing image processing techniques for accurate recognition using mathlab as the implementation tool.

Statement of the Problem

a) Since 2009, the Nigerian security services have regularly reported that they have killed Boko Haram leader and warlord AbubakarShekau, only for him to reappear in videos posted on the internet less than a year later as Boko Haram's new leader [11]. In order word anassumed innocent man has been killed mistaken for Shekau. (mistaken identity)
b) According to [12] and [13], at the end of 2012, six women were raped in Marseille France. Evidence, including DNA, led police to not one, but two suspects (Elwin and Yohan) which are identical twins. When asked to identify the attacker, victims recognized the twins but couldn't say which one had assaulted them.
c) Imposters taking advantage of face transplant techniques to access classified information.
d) Recent increase in signature forgery and duplication of access key cards by impersonators.

Aim and Objectives of the Study

The aim of the researcher is to develop a novel system that apply machine learning to recognize identical twins to support global crime investigation with the following set out Objectives:

i. To develop a system that recognizes the real image identity of two identical suspects.
ii. To employ machine learning and bag of features for the recognition and verification process
iii. To perfect existing system of face recognition and verification system.
iv. To develop a system that support and improves forensic investigation

II. LITERATURE REVIEW

In 2013, [14] researched on differentiating identical twins using conditional face recognition algorithms, however the performance of their work are broken down by expression of face, gender, lightening and age. [15] Worked on a study of face recognition of identical twins by human. However, their work recorded 78.825% accuracy.[16] Researched on geometric approach for face recognition and revealed that theproblem of face recognition is as a result of the difference in human pose, face expression, hairstyle, image style and lighting conditions. [17] Studied face recognition system using Adaboost algorithm. Their work converts the input images from a camera to binary pattern and face location candidates using the AdaBoost Algorithm. However, despite the success of the selected literatures reviewed above, none can precisely differentiate identical twins.

III. METHODOLOGY

In this section, the various methods to be adopted for the development of the proposed system will be discussed below while figure 2: explain the process flow sequentially.

Image capture: this process involves capturing the facial image using a HD image acquisition device (quality of image is important). It is implemented using the image acquisition tool in mathlab.

Face detection: this employs computer vision tool to detect human face and facial features using the viola and jones algorithm specially developed for face tracking and detection.

Image Binarization: this process converts the facial image into a bi-level in preparation for preprocessing.

Histogram equalization and transformation: this is a preliminary processing techniques employed to enhance the image and eliminate background noise.

Edge Detection

This method simplifies and minimizes the image data to be processed. "Edge detection begins with the examination of the local discontinuity at each pixel element in an image amplitude, orientation, and location of a particular subarea in the image that is of interest. Also it detects the boundary between two homogeneous regions and finds out the outlines of an image background. "This process of partitioning a digital image into multiple regions or set of pixels is called image segmentation".

Image Normalization

Once the face and its features are segmented, the next stage is to normalize these face parts, to enable generation of general facial analysis through feature extraction. The normalization process help achieve constant illumination in the facial image.

Bag of Feature extraction

This is the final step of the image processing step. This involves a statistical method of dimensionality reduction that represents the discriminative part of the image in a compact feature vector, this is implemented with mathlab using the histogram of oriented gradient technique to extract the training and label features. The feature extracted are called bag of statistical features from both the query (y) and training dataset (x).

Recognition

Based on the feature extraction techniques of statistical method that extracts large amount of statistical values from the face, a machine learning algorithm is required for the classification and training of the extracted features hence a supervised machine is employed under the holistic approach to classify and train the extracted features for recognition.

Training and Classification

For the training and classification of the statistical extracted features, we employ a supervised machine learning technique called the k-nearest neighbor classifiers. It is a simple algorithm and a non parametric method that classifies data based on similarity measures such as distance metrics which are defined by the standard Euclidean distance and Euclidean distance [3].Given an *m1*-by-*n* data matrix X, which is treated as *m1* (1-by-*n*) row vectors x_1, $x_{2.....}$, ..., x_{m1}, and *m2*-by-*n* data matrix Y, which is treated as *m2* (1-by-*n*) row vectors y_1, y_2, ...,y_{m2}, the various distances between the vector x_s and y_t are defined as follows [18]:

Euclideandistance: d_{st}^2=(xs−yt)(xs−yt)……..equation (i)

Standardized Euclidean distance:

d_{st}^2= (xs−yt)V−1(xs−yt)....equation (ii)

Where V is the m-by-*n* diagonal matrix whose *j* the diagonal element is $S(j)^2$, where S is the vector containing the inverse weights.

Prediction of Label (Result)

For the prediction of the classified label: the mathematics function below is employed as in [18].

$q = \underset{q=1,,,,,k}{\arg\ min} \sum_{k=1}^{k} T\left(\frac{k}{x}\right) C\ (\frac{q}{k})$, ...Equation (iii)

Where: q is the predicted classification.

k is the number of classes.

$T\left(\frac{k}{x}\right)$is the posterior probability of class *k* for observation *x*.

$C\ (\frac{q}{k})$ is the cost of classifying an observation as *y* when its true class is *k*

Implementing the function given a set *x* of *n* points and a distance in equation (i) and (ii) respectively, k nearest neighbor (K-NN) search finds the (k) closest points in *x* to a query point or set of points y (see figure 1).

Given a number of columns of x to be less than 15(for simplicity sake) the equidistance is calculated and the label is predicted as shown in equation (i). The clustering diagram below further explains the classification model and label prediction. However, according to [18] if the number of x is more than 15, the ExhaustiveSearcher model is employed, this model considers the either the hammering distance, angles, correlation, spearman, jaccard, seuclidean, mehalanobis or custom distance function for classification.

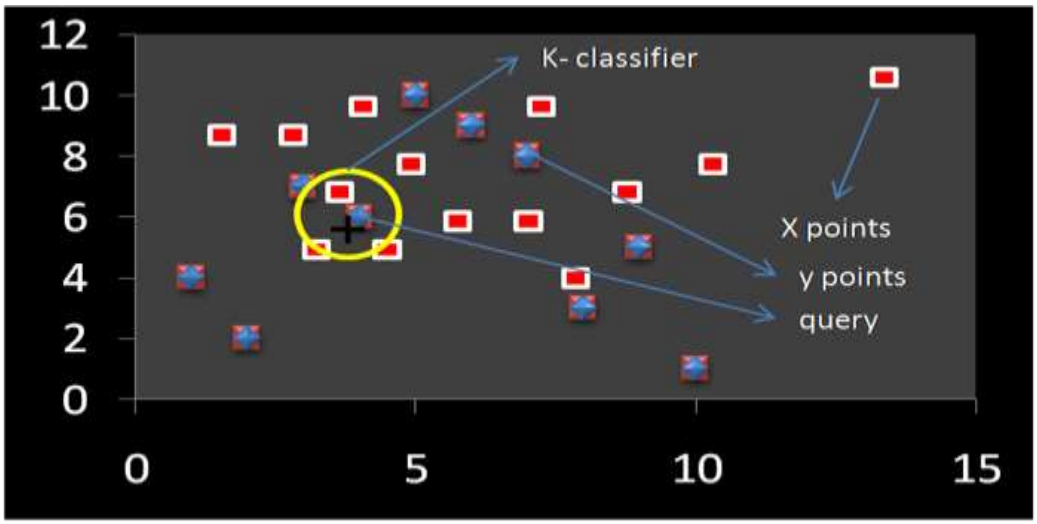

Figure 1: KNN classification model

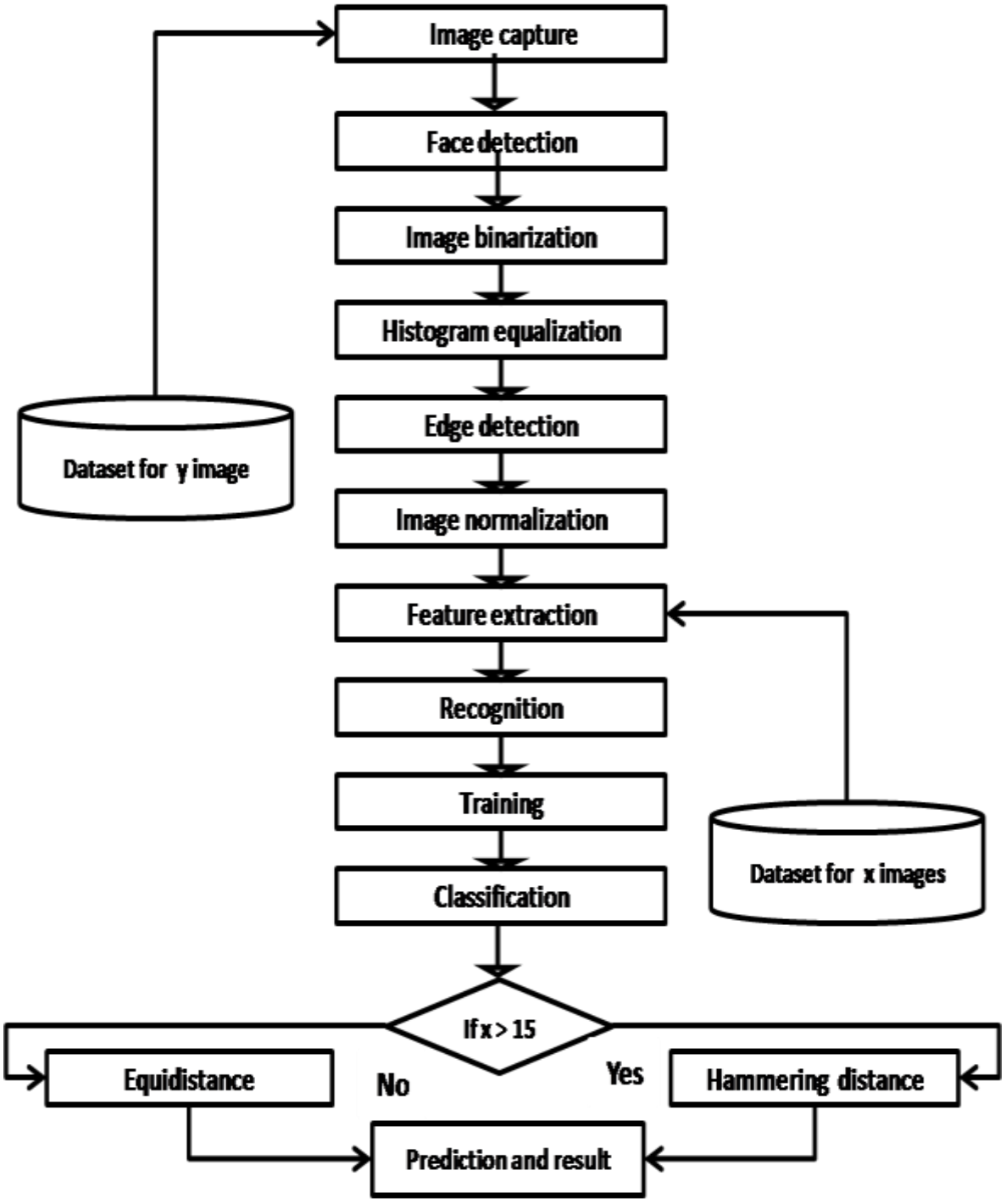

Figure 2: process flow chart

IV. IMPLEMENTATION RESULT

This research work is experimented and developed using mathlab and the results are presented below. Figure 3 shows the evidence of two identical twins which will be added to training dataset(circled in yellow) shown in figure 5. One of figure 4: images will be use for search as the query image (see result in Figure 6). Figure 6: also provides the preliminary filtration model using the histogram transformation analysis while figure 7: demonstrate other image processing results. The feature extraction and normalization results are presented in figure 8: respectively. Finally the recognition result is displayed in figure 9.

Figure 3:y images of training dataset

figure 4: x query images of identical twins

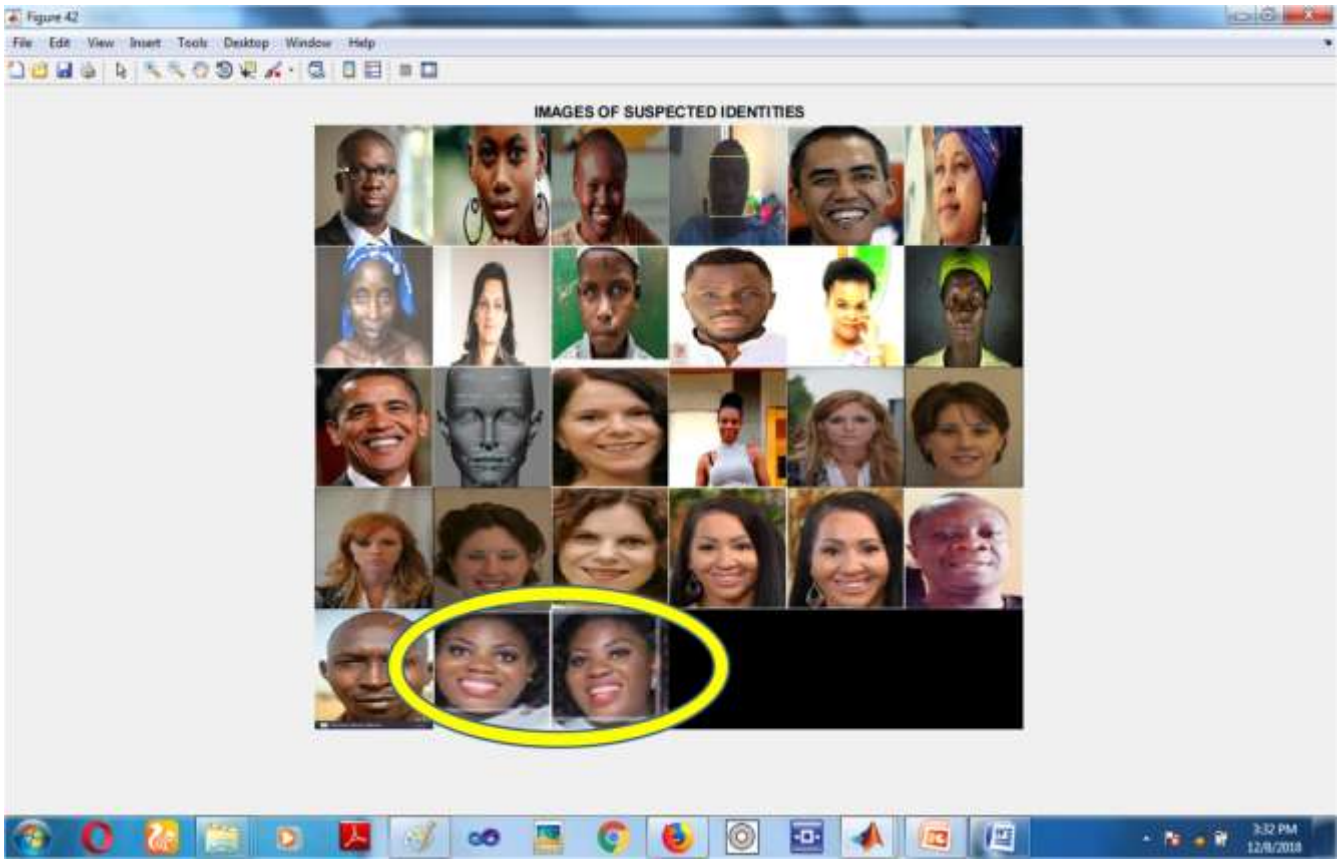

Figure 5: dataset of montage images (suspected identities)

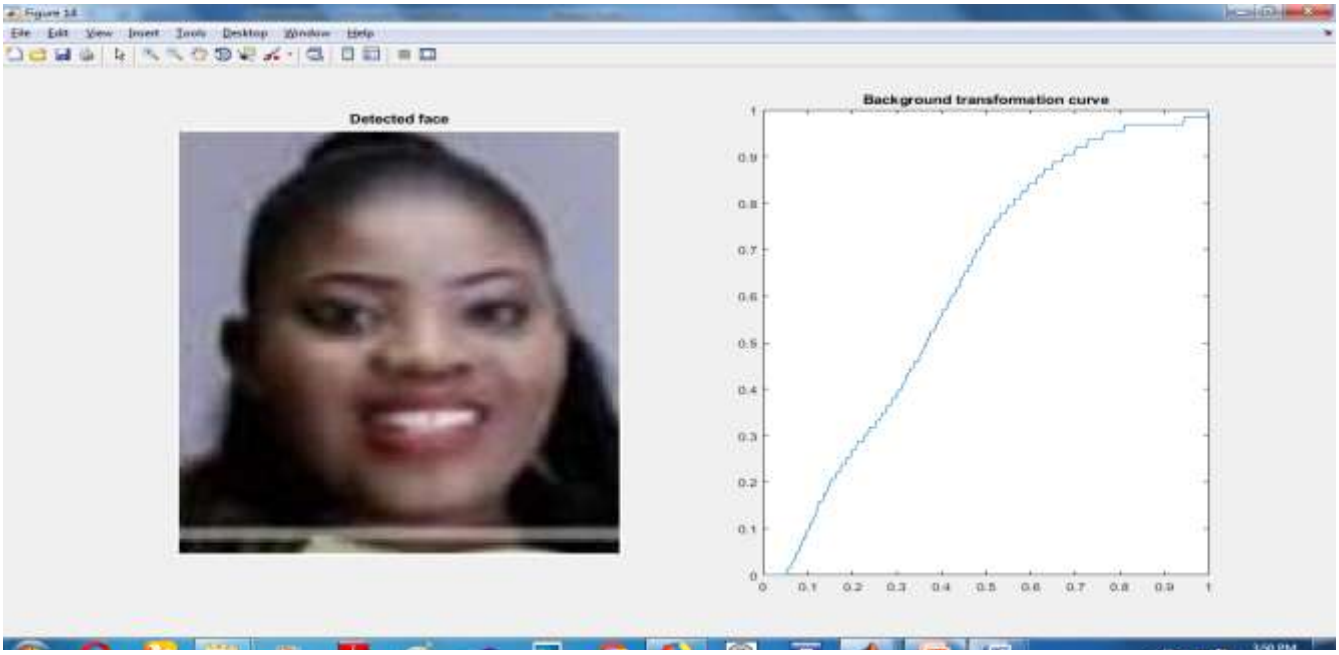

Figure 6: detected query image with histogram transformation graph

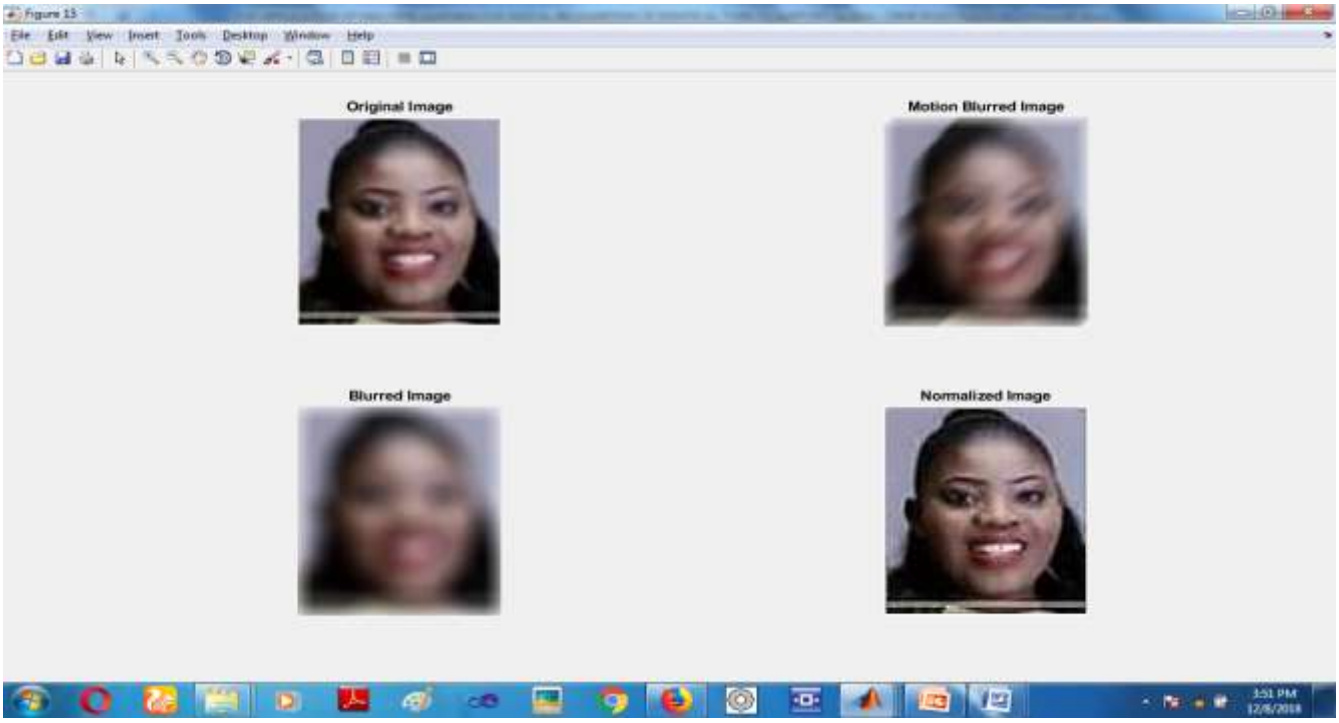

Figure 7: image filtration and normalization result

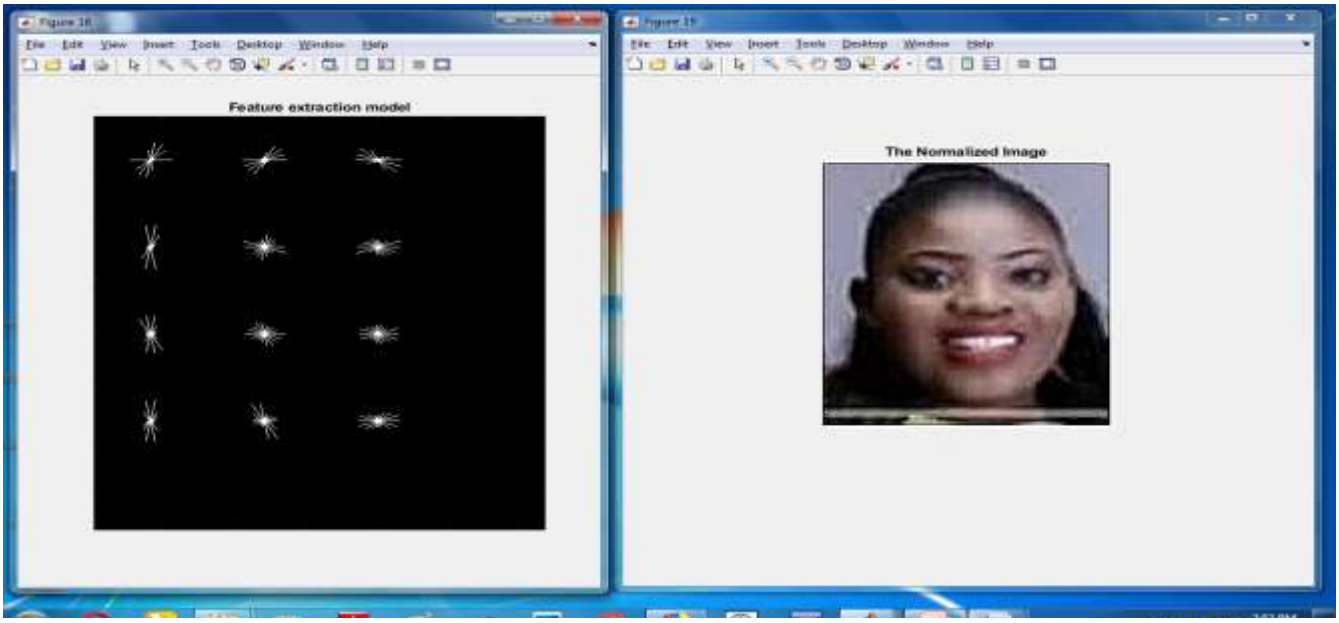

Figure 8: model of the feature extraction process

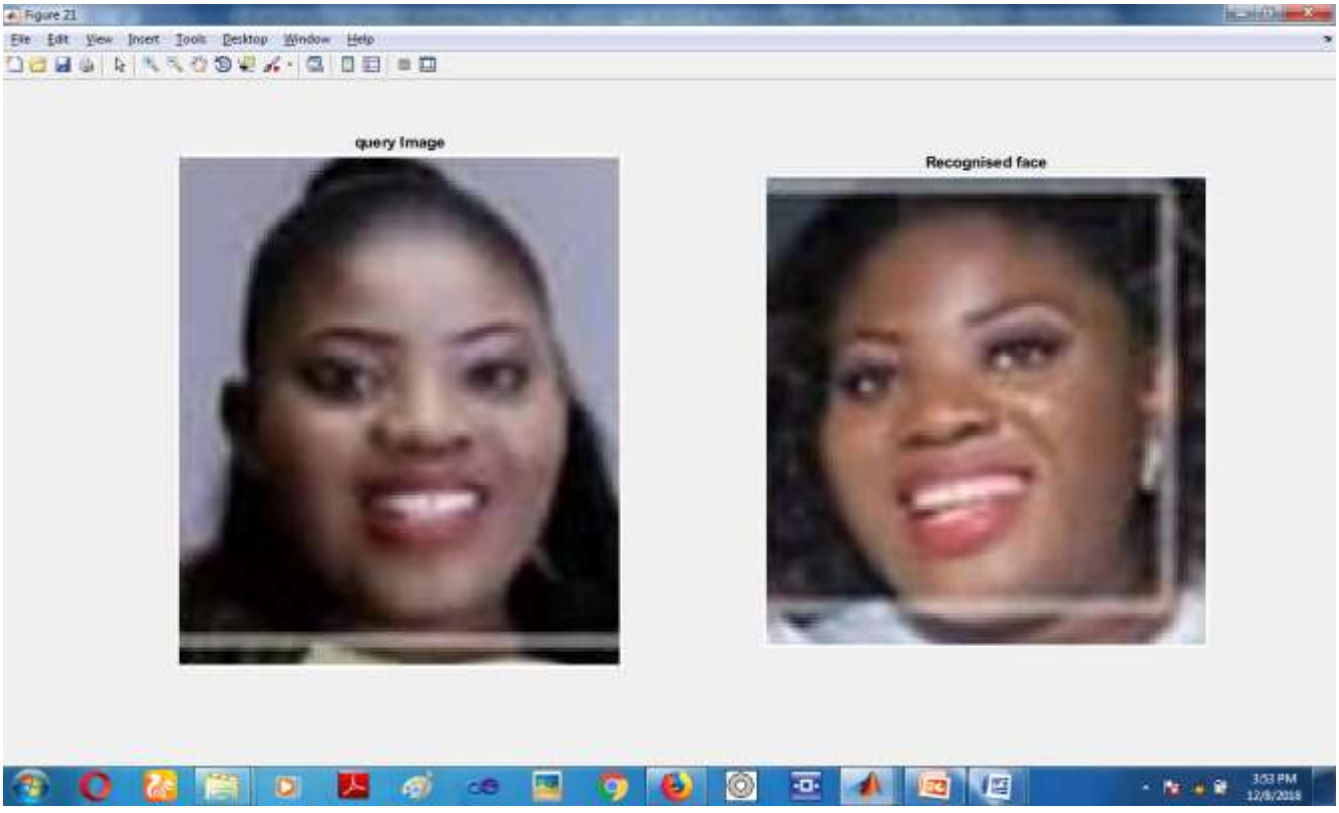

Figure 9: Recognition result

V. CONCLUSION

This work has successfully provided a new pathway to support digital forensic investigation, employing artificial intelligence (machine learning) to improve existing face recognition systems. Secondly, K-NN classifier is employed to train and classify the extracted features from both the query (y) and class images (x) using the Euclidean model. The result is a classof (x, y) K nearest neighbor variables which the algorithm is labeled as the recognition result. We believe that this work can be further improved using neural network instead of support vector machine which was employed for this research due to the limited number of images dataset we used as training set, also the time for the recognition process can also be improved in future work. Other images tested with this work provided a recognition result above 92%.

REFERENCES

[1]. Erik Mannik, Terrorism; its past, present and future prospects
[2]. Desouza. Facial component analysis and image processing for face detection. 46(2):235–245, May 2011.
[3]. Chioma O. Implementation Of Daugman's Algorithm And Adaptive Noise Filtering Technique For Digital Recognition Of Identical Twin Using Mathlab, 2018
[4]. V. Vijayan, K. Bowyer, P. Flynn, D. S Huang, L. Chen, M. Hansen, O. Ocegueda, S. Shah, and I. Kakadiaris. Twins 3D Face Recognition Challenge. In International Joint Conference on Biometrics, 2011.
[5]. K. Kelly, design and implementation of face detection system , Volume 1, Issue 2, March-April, 1990, ISSN 091-273.
[6]. H. Rowley, S. Baluja, and T. Kanade. Neural network-based face detection. In Proc. IEEE Conf. on Computer Visioin and Pattern Recognition, pages 203-207, San Francisco, CA, 1996
[7]. Kah-Kay Sung and TomasoPoggio. Example-based learning for view-based human face detection. A.I. Memo 1521, CBCL Paper 112, MIT, December 1994
[8]. B. Moghaddam and A. Pentland. Probabilistic visual learning for object representation. In S.K. Nayar and T. Poggio, editors, Early Visual Learning, pages 99--130. Oxford Univ. Press,1996.
[9]. O. Jesorsky, K. J. Kirchberg, R.W. Frischholz. Robust Face Detection Using the Hausdorff Distance. In Proc. Third International Conference on Audio- and Video-based Biometric Person Authentication, Halmstad, Sweden, 2001
[10]. Vimal, VirenderKadyan, Implementation and Performance Analysis of Face Recognition Using MATLAB. IJCST Vol. 6, Issue 2, April - June 2015 ISSN : 0976-8491 (Online) | ISSN : 2229-4333.
[11]. John Campbell, From Africa in Transition. Boko haram's AbubakarShekau; Dead or Alive. July 26, 2017.
[12]. Alison Gee. Twin DNA test: Why identical criminals may no longer be safe. 15 January 2014
[13]. https://www.bbc.com/news/magazine-25371014
[14]. Swati Y. Dhote, A. D. Gotmare, M. S. Nimbarte. Differentiating Identical Twins by Using Conditional Face Recognition Algorithms. (IJSR) ISSN (Online): 2319-7064 Index Copernicus Value (2013)
[15]. Soma Biswas, Kevin W. Bowyer and Patrick J. Flynn. A Study of Face Recognition of Identical Twins byHumans
[16]. Samal D. and Starovoitov V. "Approaches and methods to face recognition. A survey," Instute of Engineering Cybernetics, 54 pages, 2014.
[17]. Senthilsingh and Manikandan. Face recognition system using Adaboost algorithm. IEEE Transactions on Neural Networks, Volume: 16 , Issue: 2 , pp. 436 446, 2014
[18]. Mathworks documentation 2017a.