DOI: https://doi.org/10.33103/uot.ijccce.22.2.6

Comparative of Viola-Jones and YOLO v3 for Face Detection in Real time

Tameem Hameed Obaida[1], Nidaa Flaih Hassan[2], Abeer Salim Jamil[3]
[1]*Computer Systems Techniques, Al-Furat Al-Awsat Technical University, Najaf Technical Institute, AL Najaf, Iraq*
[2]*Department of Computer Science, University of Technology, Baghdad, Iraq*
[3] *Department of Computer Technology Engineering, Al-Mansour University College, Baghdad, Iraq*
[1]*tameem.daham@atu.edu.iq*, [2]*110020@uotechnology.edu.iq*, [3]*abeer.salim@muc.edu.iq*

Abstract— *This Face detection is considering one of the important topics for recognizing human, it is the first step before the face recognition process, it is considered one of the biggest challenges in the field of vision computer. In recent years Many algorithms for detection have appeared, which depend on extracting the features of the human face, and works continue to develop them to this day. This paper aims to make a comparison between two of the most commonly face detection methods, Viola Jones (V_J) and YOLO v3. This comparison is made to determine which of the two algorithms is being most useful when used to detect faces in digital video. These algorithms are used in many applications, including image classification, medical analysis of image, and objects detection in real time (especially in surveillance cameras). Both algorithms are applied to detect faces in the real time video. The experimental results of a sample consists of 20 video frames show that V_J algorithm consumes less time in comparison with YOLO v3 algorithm, but its results are less accurate, unlike the YOLO v3 algorithm, which is slower in detect face with high accurate rate.*

Index Terms— *Viola jones, Yolo v3, V_J algorithm, Accuracy rate.*

I. INTRODUCTION

Face recognition and tracking are one of the most interesting fields of human computer interaction. There are relatively small distinguishing facial characteristics and it is the most fascinating job to observe them. Video detection and tracking facial artifacts are difficult activities. The issues associated with the undistracted presence of faces and the setting in which they occurred in the video. Since the sizes, positions, occluded and disordered face and face poses in video are not limited, difficulties can arise. If the face camera and surroundings are in motion, it may be difficult to recognize the face and monitor the face from the video series. The detection and tracking depend primarily on unique visual and motion attributes. Thus, these algorithms may meet difficulties in detecting and tracking the object. In addition, in order to process the large amount of data, an efficient algorithm is required to save computational costs[1,2].

In many applications such as driving assistance, video monitoring or facial recognition, the issue of face detection is a basic problem. Clearly, because of the differences in appearance such as : location and orientation of the face in the image, appearance of (glasses, hat, beard) that partially masked face [3]. In this paper, comparison between Viola-Jones Haar Cascade Classifier algorithm and the Convolutional Neural Network

Received 18/September/2021; Accepted 24/November/2021
 ISSN (Print) 1811-9212 ISSN (Online) 2617-3352

DOI: https://doi.org/10.33103/uot.ijccce.22.2.6

(CNN) is presented, in addition to explaining the advantages and disadvantages of each method in the field of face detection.

This paper is organized as follows: Section 2, presented recent work for detecting face using Viola Jones (V_J) and YOLO v3, the basic steps of two algorithms are explained in Section 3. Section 4 discusses the results of applying these two algorithms in real time video, and finally Section 5 presents the conclusion of this paper.

II. RELATED WORK

There are many algorithms used to detect faces in the video, this paper is focused on V_J and Yolo v3 algorithms, which are the most popular for their high ability to detect faces. The following researches have been chosen as the most recent effective in their performance:

Jaber et.al,2017, analyzed Viola Jones algorithm and proposed a method for improving its accuracy and speediness based on coding of eyes, using an Open CV, its goal is to eliminate errors in eye-based detection, then training algorithm is implemented which include negative and positive images to produce distinguished features, these features are stored in an XML file for detecting faces and eyes. The results obtained as follows, Viola Jones algorithm got 60.12% accuracy, while the proposed method got 90.89% accuracy. But both methods detected only the front faces, thus faces must be at a close distance from the camera in order to extract the accrue features [4].

In 2017 , Dang and Sharma compared number of detection algorithms in terms of accuracy by the DetEval program in order to obtain accurate results, because the algorithm deals well with the values of the boxes around the face. V_J getting the highest accuracy(0.27321),SMQT Features(0.26792),Neural Network (0.339450) SVM(0.01392850) sequentially. Their work is done on the front faces only, and no consideration is given to head movement, expressions, or occlusion[5].

Pratama et.al in 2020, compared two dark and bright contrast colors to read objects in three-dimsions image, the comparsion is accomplished by calculating the average of pixel number of objects in image. Thus, V_J face detection algorithm had been proven to have many weaknesses, since it is not possible to detect the face in the tilt condition. Their results indicated that the accuracy of this algorithm reaches 90% in the case of front faces and close to cameras only [6]**.**

Rahmad et.al,2020, compared V_J with HOG, using multiple conditions including head position, expressions, and occlusion. The V_J algorithm achieved an accuracy of 75.33 percent and HOG achieved an accuracy of 80.22 percent as average for this conditions [7].

Mao et.al,2019, proposed a lightweight method to object detection called Mini_YOLOv3, based on the Darknet_53. They used technique for reducing the dimension first, then increasing in order to (restore) it . The proposed algorithm was tested on data in the COCO_MS database and proved that it has the same accuracy as the YOLO v3, but it takes half the time to detect the objects [8].

III. FACE DETECTION ALGORITHMS

Face detection and tracking algorithms in digital video are very important, since they precede face recognition, correctly detecting a face makes the recognition process easy. Many algorithms have been used for face detection, V_J and Yolov3 algorithms are

Received 18/September/2021; Accepted 24/November/2021
 ISSN (Print) 1811-9212 ISSN (Online) 2617-3352

DOI: https://doi.org/10.33103/uot.ijccce.22.2.6

considered in this work because they are widely used by programmers and developers, and the following is a description of each algorithm[5] :

A. Viola-Jones face detection (V_J)

In 2001, Paul Viola and Michael Jones proposed the V_J Object Detection system, which was the first framework to offer competitive detection rates for objects. It can be used in real time to detect objects, but it is mostly applied to face detection applications. This framework's detection rate is quietly high (true positive rate) and very low false-positive rate, making the algorithm strong and rapidly processing the images as well. The primary objective is face detection, not recognition, which is to distinguish faces from non-faces as the first recognition step[5, 9]. In V_J algorithm, the four main steps taken are:

1. Haar Feature Selection:

As all faces of human have similar attributes, the region of the eyes is darker than the region bridge of the nose, etc. Using the Haar feature, also known as digital image features, based on Haar base functions, these attributes are compared[5,10].

In the V-J algorithm, Haar classifiers are used to detect face's characteristics. In computer vision, Haar features are used to recognize the density of pixels in a Zone in a traceable way. These features in image represented as rectangle regions, and the classifiers consist of two or three rectangles to be continuously be scanned for window features[7]. *Fig. 1* illustrates the features of Haar:

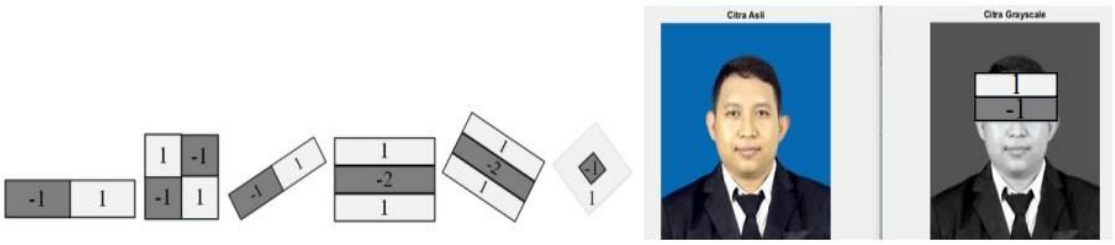

FIG. 1. HAAR FEATURES [7].

2. Creating an Integral image:

Computing the rectangles adjacent to the rectangle present at (x,y) into a single image representation. *Fig. 2* illustrates how to compute an integral image from pixel values.

The value of each point in an integral image is the sum of all the pixels up and to the left, including the target pixel, and the sum of the pixels in the orange box is then determined. By formula usage D - B - C + A, the integral image value is 113 – 42 – 50 + 20 = 41[5].

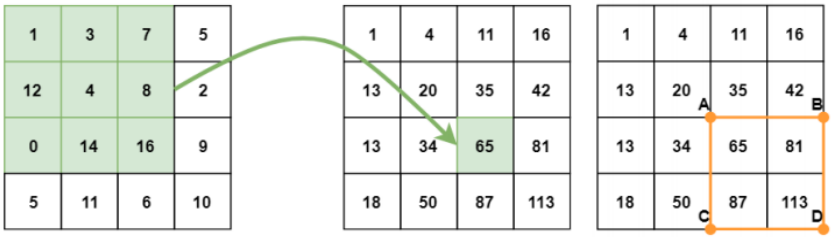

FIG. 2. INTEGRAL IMAGE [7].

Received 18/September/2021; Accepted 24/November/2021

ISSN (Print) 1811-9212 ISSN (Online) 2617-3352

DOI: https://doi.org/10.33103/uot.ijccce.22.2.6

3. AdaBoost Training:

AdaBoost is a learning algorithm which consists of a collection of weak binary classifiers to generate a powerful new one; it is used to be trained. This algorithm is able to extract important small visual features from a large amount of features[3].

4. Cascading Classifiers:

The V_J method of combining classifiers for organizing a collection of features in the form a multilevel classification. This discards the background windows quickly, in order to do further computations on face-like regions. *Fig. 3* shows Cascade Classifiers[3].

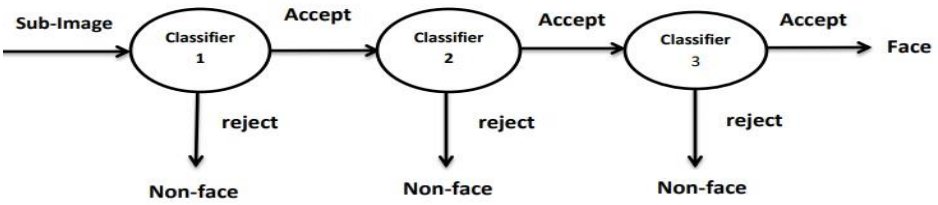

FIG. 3. CASCADE CLASSIFIERS [7].

5. face detection Flowchart

Haar's rectangular features are applied to the face in order to calculate the difference between black and white pixels by scanning the gray scale image. If the face is found, then a rectangle is drawn on the face. Many classifiers pre-trained for the face and eyes are already available in Open CV and xml files were saved in the Open CV. *Fig. 4* illustrates a face detection diagram using V_J algorithm [4].

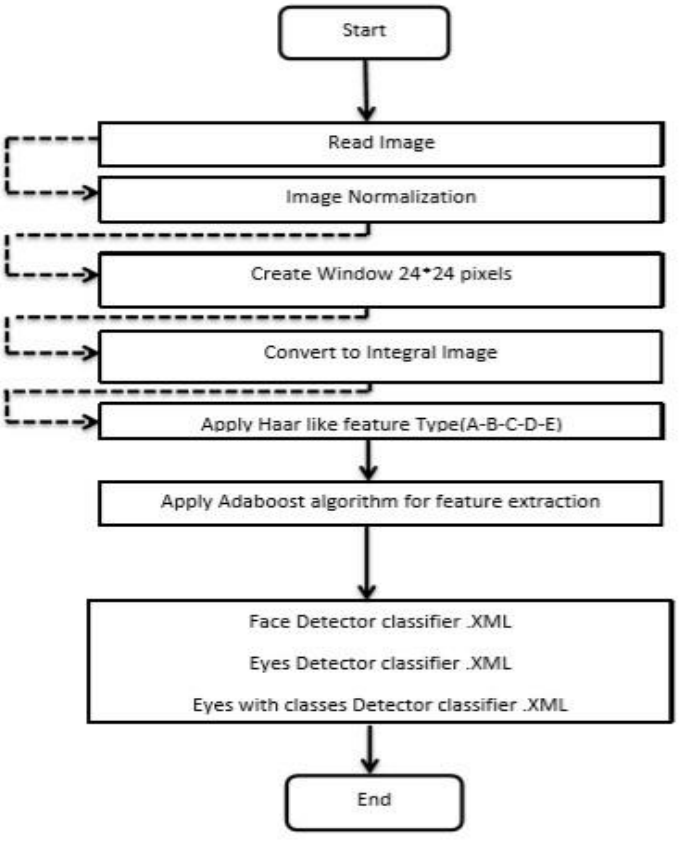

FIG. 4. CLASSIFIER OF OPENCV [4].

Received 18/September/2021; Accepted 24/November/2021

ISSN (Print) 1811-9212 ISSN (Online) 2617-3352

DOI: https://doi.org/10.33103/uot.ijccce.22.2.6

B. YOLOv3 object detection Algorithm

Deep learning methods have been used in detection of object in recent years, where it uses features of low level to construct high level features that are more abstract and to show the data hierarchically in order to enhance object detection. As compared with conventional algorithms for detection, the approach based on deep learning-based object detection has superior performance for multi-classification tasks in terms of robustness, accuracy and speed [11, 12].

With the constant growth of the convolutional neural network (CNN), a lot of progress has been made in face detection due to the use of modern CNN-based object detectors[13], including RCNN,SSD,YOLO ,Taking advantage of the strong deep learning approach to extract features of image. Nevertheless, the CNN needs to build its own network structure, and optimizes the network weight parameters by training. *Fig. 5* shows detection of facial based on CNN[8].

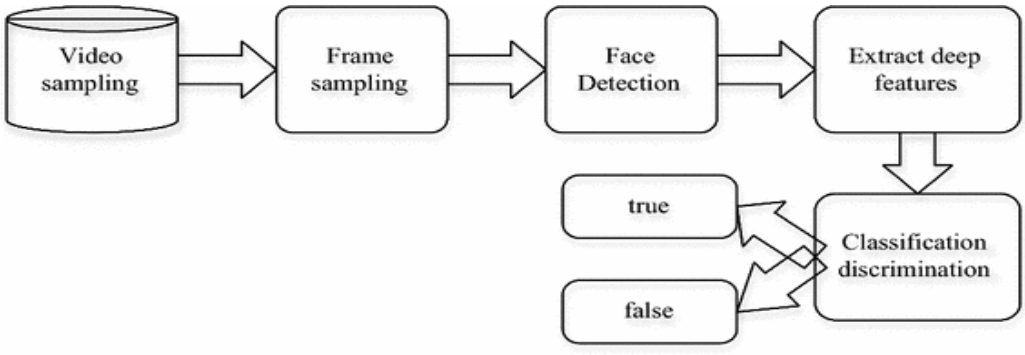

FIG. 5. DETECTION OF FACIAL BASED ON CNN [14].

Use CNN to extract the face image feature; the distribution of probability for each image of the face is counted. As the classification outcome of the video, maximum probability is used of all the sample faces in the video[14, 15].

A broad variety of united pipeline framework-based methods have been suggested by researchers in recent time, one of which is the you_ only _look _once v2 (Yolo v2) approach. In order to raise the recall, this method uses the batch normalization to enhance the confluence and avoid over-fitting, and anchor boxes to predict bounding boxes. Recently, Huang and Pedoeem proposed a shoaly real-time detection way based on the Yolo v2 method for non-GPU computers; in order to speed up the detection quickness, their method decreases the input image size by half [11, 16].

YOLOv3 is a version of YOLO and YOLOv2 that has been enhanced [17]. With a single feed forward CNN, it explicitly forecasts class probabilities and bounding box offsets from full images. It completely removes the generation of region proposals and sampling features, the method of YOLOv3 divides the input image into small grid cells of S×S. The grid cell is responsible for detecting the object if the center of an object falls into a grid cell. Each cell predicts the B bounding box location data and compute the object degrees corresponding to these bounding boxes[11].

It is possible to get each object score as follows:

$$C_i^j = P_{i,j}\,(\text{Object}) * IOU_{\text{pred}}^{\text{truth}} \quad (1)$$

Whereby C_i^j is the abjectness score, $P_{i,j}\,(\text{Object})$ object function, (IOU) it is the intersection over union. As one portion of the loss function, the YOLOv3 technique uses

Received 18/September/2021; Accepted 24/November/2021
 ISSN (Print) 1811-9212 ISSN (Online) 2617-3352

DOI: https://doi.org/10.33103/uot.ijccce.22.2.6

binary cross entropy of expected object degrees and true object degrees. It is possible to express it as follows[11]:

$$E_1 = \sum_{i=0}^{S2} \sum_{j=0}^{B} W_{ij}^{obj} [\hat{C}_i^j \log(C_i^j) - (1 - \hat{C}_i^j) \log(1 - C_i^j)] \quad (2)$$

Where S2 is the number of grid cells, C_i^j and $\hat{} C_j^j$ are degree the predicted, and B is the number of bounding squares, four projections: :tx,ty, tw,th of bounding box, (CX, CY) is a group of grid cells in the image's upper left corner. From the upper left corner of the image, the central location of the final bounding boxes projected is set by (bx, by). It is calculated as follows[11]:

$$\begin{aligned} b_x &= \sigma(t_x) + c_x \\ b_y &= \sigma(t_y) + c_y \end{aligned} \quad (3)$$

σ() is a Sigmoid Function. The height and width of the bounding box projected are determined as follows:

$$\begin{aligned} b_w &= p_w e^{tw} \\ b_h &= p_h e^{th} \end{aligned} \quad (4)$$

As one aspect of the loss function, the YOLOv3 approach uses the square error of coordinates prediction[8, 11]. It is possible express it as follows:

$$\begin{aligned} E_2 = {} & \sum_{i=0}^{S2} \sum_{j=0}^{B} W_{ij}^{obj} [(\sigma(t_x)_i^j - \sigma(\hat{t}_x)_i^j)^2 + (\sigma(t_y)_i^j - \sigma(\hat{t}_y)_i^j)^2] \\ & + \sum_{i=0}^{S2} \sum_{j=0}^{B} W_{ij}^{obj} [((t_w)_i^j - (\hat{t}_w)_i^j)^2 + ((t_h)_i^j - (\hat{t}_h)_i^j)^2] \end{aligned} \quad (5)$$

C. Yolo v3 Architecture

YOLO v3 incorporates all residual blocks, up sampling and skip connections. YOLO v3 is a completely convolutional network, and by applying a 1 ×1 kernel on a feature map, its eventual output is generated. 1 × 1 is the form of the detection kernel (B × (5 + C)). So, the size of the kernel is 1 ×1 ×255. Also, YOLO v3 works at 30 frames per second in order to increase the accuracy compared to previous versions.[18].

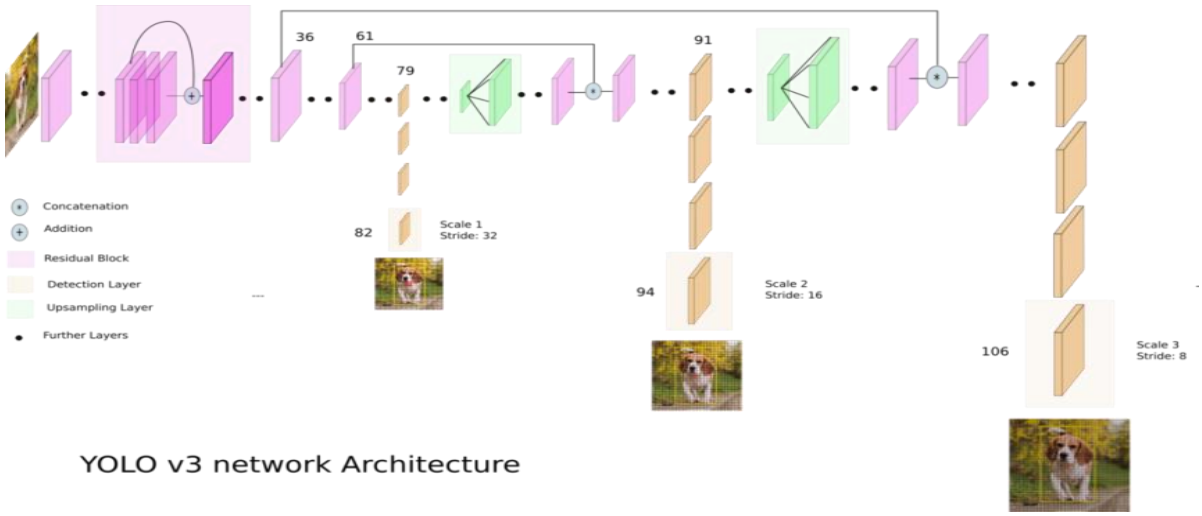

FIG. 6. YOLO V3 ARCHITECTURE [18].

Received 18/September/2021; Accepted 24/November/2021
 ISSN (Print) 1811-9212 ISSN (Online) 2617-3352

DOI: https://doi.org/10.33103/uot.ijccce.22.2.6

IV. RESULT AND DISCUSSION

In this paper, a comparison was made between V_J and YOLO v3 algorithms for face detection in terms of speed and accuracy. Accuracy is considered one of the very important criteria as it depends on the rotation of the head at different angles, expressions and tilt. After performing the real-time video test, it was found that the ability of V_J algorithm to detect face is limited up to the angle of 30 by 98%, occlusion 27% and expressions 86%. The Python programming language was used to make the comparison.

Fig. 7 shows the results obtained by moving head in different directions. Among the 20 snapshots that were taken for the video, there are eleven shots in which the face was not detected; V_J algorithm achieved good results in detecting the front face, in addition to the movement of head in some different directions and angles up to the 30 angle. This means that it is weak at detecting faces at angles of 45, 60 ... etc. In addition to its inability to detect when the head is tilted. That is, the detection accuracy rate is 45% relative to the head movement (rotation). As for *Fig. 8*, shows the results obtained by a group of people with different facial expressions, such as fear, surprise and sadness....etc.

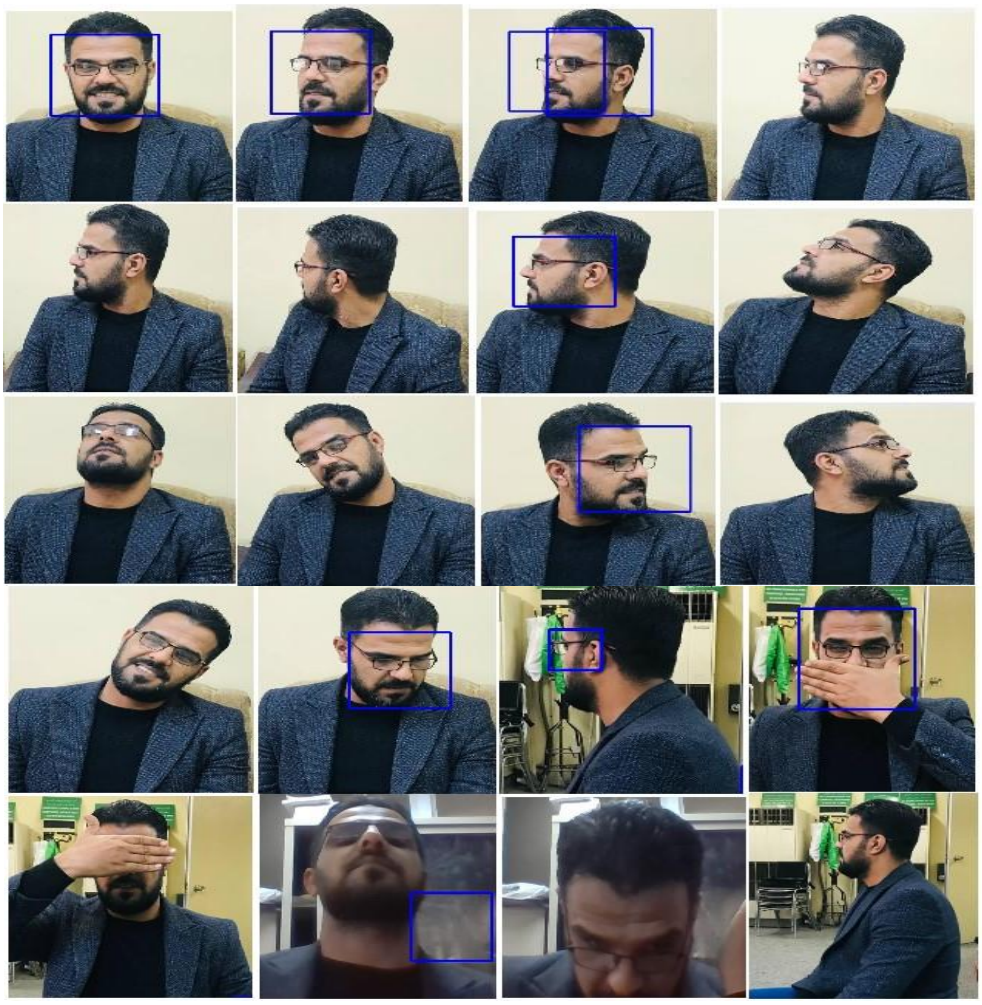

FIG. 7. SHOWS THE DIFFERENT ANGLES OF THE FACE USING V_J ALGORITHM.

Received 18/September/2021; Accepted 24/November/2021

ISSN (Print) 1811-9212 ISSN (Online) 2617-3352

DOI: https://doi.org/10.33103/uot.ijccce.22.2.6

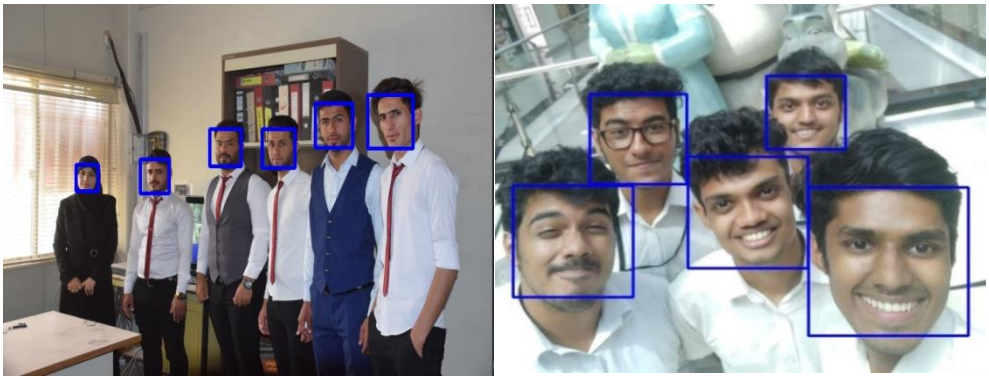

FIG. 8. SHOWS THE FACE DETECTION FOR A GROUP OF PERSONS AND DIFFERENT EXPRESSIONS.

When testing the video in real time using the Yolo v3 algorithm, the accuracy rate was about 90% in detecting the face and in all angles, occlusion 89% and expressions by 98% .It also achieved good results if the head was tilted left or right. *Fig. 9* shows the results obtained by moving the head in different directions. Among the 20 snapshots that were taken for the video, there were two shots in which the face was not exposed, in which the movement of the head is up and down strongly. As for *Fig. 10*, shows the results obtained by a group of people with different facial expressions and *Fig. 11* represents Graph of test results.

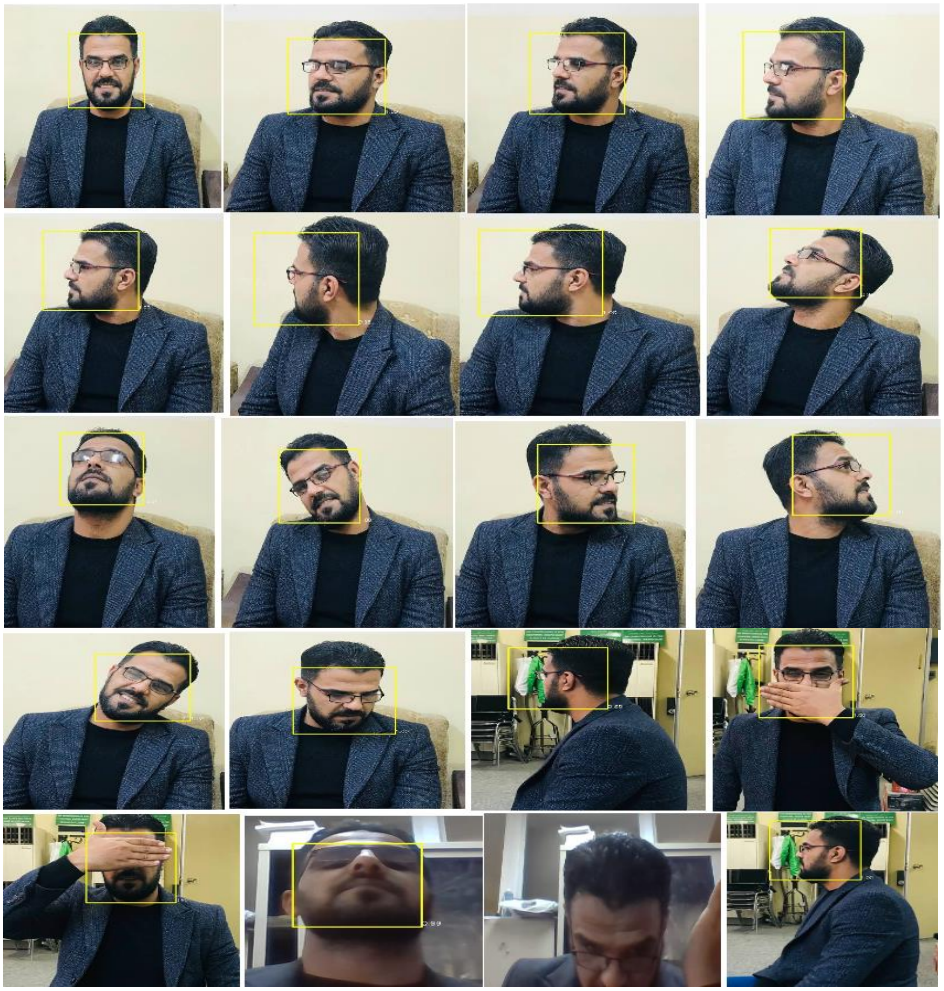

FIG. 9. ILLUSTRATION OF THE DIFFERENT ANGLES OF THE FACE USING THE YOLO V3 ALGORITHM.

Received 18/September/2021; Accepted 24/November/2021

ISSN (Print) 1811-9212 ISSN (Online) 2617-3352

DOI: https://doi.org/10.33103/uot.ijccce.22.2.6

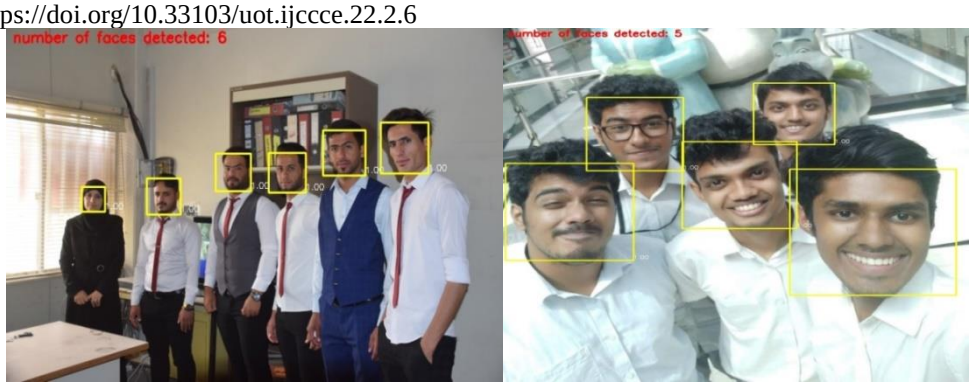

FIG. 10. SHOWS THE FACE DETECTION FOR A GROUP OF PERSONS AND DIFFERENT EXPRESSIONS.

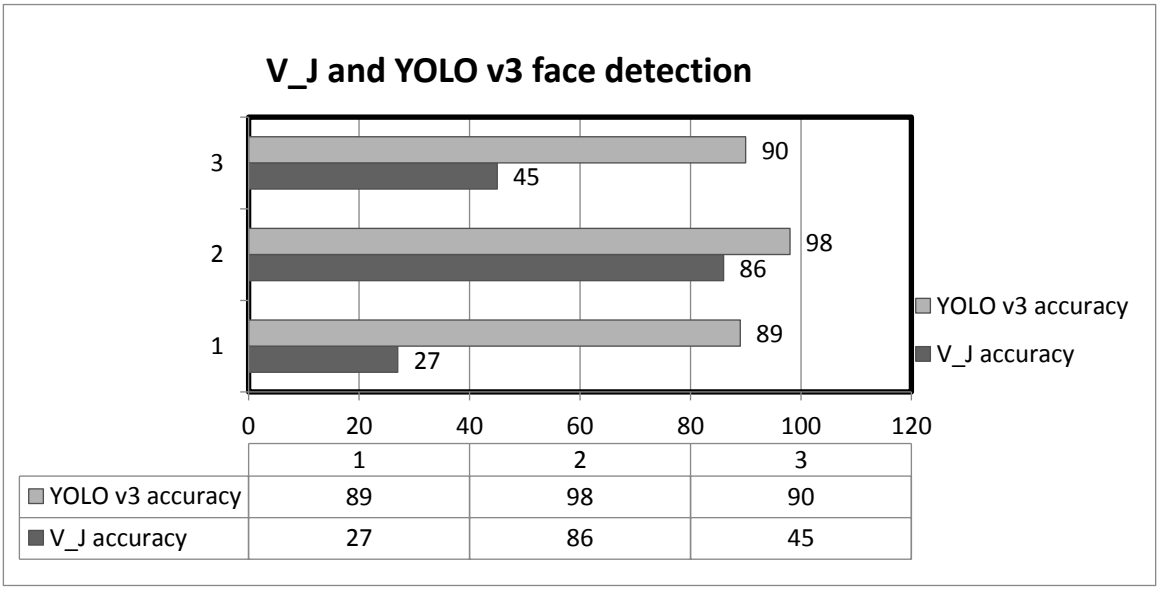

FIG. 11. GRAPH OF TEST RESULTS.

V. CONCLUSIONS

Face detection and locating it in the video is one of the topics that has gained a great importance. V_J and Yolov3 algorithms are two successful algorithms for detecting faces; they have been used in many different applications. In this paper comparison between these algorithms is done, the comparison is focused on cases of head movement in many angles. Based on the results obtained in this paper, V_J algorithm obtained an accuracy rate of 45%, while the Yolov3 deep learning algorithm obtained an accuracy rate of 90%, these rates refer to the V_J algorithm that can achieve good results in detecting the front and close faces of the camera, as well as in expressions, but its results is low in the case of occlusion, while for the YOLO v3, it achieved good results in the case of head movement at different angles, as well as in expression and occlusion. Additionally from practical view, it has been found that the V_J is faster, but its accuracy rate is less, unlike YOLO v3, it is slower but has more accuracy rate. For the future work, and through this study, the results of testing the two algorithms showed that each algorithm has advantages and disadvantages in terms of speed and accuracy. Therefore, it is possible to use Viola algorithm in the case of speed and also it is possible to use algorithm Yolov3 in the case of accuracy, or improve the algorithms, or combine them in order to obtain a fast and accurate algorithm to be suitable for work.

Received 18/September/2021; Accepted 24/November/2021
 ISSN (Print) 1811-9212 ISSN (Online) 2617-3352

DOI: https://doi.org/10.33103/uot.ijccce.22.2.6

REFERENCES

[1] A. Dey, "A contour based procedure for face detection and tracking from video," in 2016 3rd International Conference on Recent Advances in Information Technology (RAIT), 2016: IEEE, pp. 483-488.

[2] A. S. Jamil, A. M. S. Rahma, and N. H. M. Ali, "Using Shape Representation to Design Panorama Video System," AL-MANSOUR JOURNAL, no. 25, 2016.

[3] A. P. Mena, M. Bachiller Mayoral, and E. Díaz-Lópe, "Comparative study of the features used by algorithms based on viola and jones face detection algorithm," in international work-conference on the interplay between natural and artificial computation, 2015: Springer, pp. 175-183.

[4] A. M. A. Hossen, R. A. A. Ogla, and M. M. Ali, "Face detection by using OpenCV's Viola-Jones algorithm based on coding eyes," Iraqi Journal of Science, pp. 735-745, 2017.

[5] K. Dang and S. Sharma, "Review and comparison of face detection algorithms," in 2017 7th International Conference on Cloud Computing, Data Science & Engineering-Confluence, 2017: IEEE, pp. 629-633.

[6] A. P. SP, D. Dannika, and A. P. Wahyu, "FACERECOGNITION SYSTEM USING VIOLA JONES ALGORITHM & CASCADE CLASSIFICATION," PalArch's Journal of Archaeology of Egypt/Egyptology, vol. 17, no. 4, pp. 2732-2740, 2020.

[7] C. Rahmad, R. A. Asmara, D. Putra, I. Dharma, H. Darmono, and I. Muhiqqin, "Comparison of Viola-Jones Haar Cascade classifier and histogram of oriented gradients (HOG) for face detection," in IOP conference series: materials science and engineering, 2020, vol. 732, no. 1: IOP Publishing, p. 012038.

[8] Q.-C. Mao, H.-M. Sun, Y.-B. Liu, and R.-S. Jia, "Mini-YOLOv3: real-time object detector for embedded applications," Ieee Access, vol. 7, pp. 133529-133538, 2019.

[9] H. Ayad, N. F. Hassan, and S. Mallallah, "Image Categorization Based Color Detector," Engineering and Technology Journal, vol. 34, no. 5 Part (B) Scientific, 2016.

[10] M. Emaduldeen and R. M. Hassan, "Image Seam Carving Based on Content Aware Resizing by Gradient Method," AL-MANSOUR JOURNAL, no. 25, 2016.

[11] L. Zhao and S. Li, "Object detection algorithm based on improved YOLOv3," Electronics, vol. 9, no. 3, p. 537, 2020.

[12] Z.-Q. Zhao, P. Zheng, S.-t. Xu, and X. Wu, "Object detection with deep learning: A review," IEEE transactions on neural networks and learning systems, vol. 30, no. 11, pp. 3212-3232, 2019.

[13] H. I. Abdulrazzaq and N. F. Hassan, "Modified Siamese Convolutional Neural Network for Fusion Multimodal Biometrics at Feature Level," in 2019 2nd Scientific Conference of Computer Sciences (SCCS), 2019: IEEE, pp. 12-17.

[14] W. Liu, "Video face detection based on deep learning," Wireless Personal Communications, vol. 102, no. 4, pp. 2853-2868, 2018.

[15] N. F. Hassan and H. I. Abdulrazzaq, "Pose invariant palm vein identification system using convolutional neural network," Baghdad Science Journal, vol. 15, no. 4, 2018.

[16] F. Han, H. Zhu, and J. Yao, "Multi-Targets Real Time Detection from Underwater Vehicle Vision via Deep Learning CNN Method," in The 29th International Ocean and Polar Engineering Conference, 2019: OnePetro.

[17] L. Z. Chun, L. Dian, J. Y. Zhi, W. Jing, and C. Zhang, "Yolov3: Face detection in complex environments," International Journal of Computational Intelligence Systems, vol. 13, no. 1, pp. 1153-1160, 2020.

[18] J. Redmon and A. Farhadi, "Yolov3: An incremental improvement," arXiv preprint arXiv:1804.02767, 2018.

Received 18/September/2021; Accepted 24/November/2021
 ISSN (Print) 1811-9212 *ISSN (Online) 2617-3352*