

Performance Comparison of Detection, Recognition and Tracking Rates of the different Algorithms

Meghana Kavuri¹, Kolla Bhanu Prakash²

Department of Computer Science and Engineering, Koneru Lakshmaiah Education Foundation, Guntur, India

Abstract—This article discusses the approach of human detection and tracking in a homogeneous domain using surveillance cameras. This is a vast area in which significant research has been taking place from more than a decade and the paper is about detection of a human and its face in a given video and stores Local Binary Pattern Histogram (LBPH) features of the detected faces. Once a human is detected in the video, that person will be given a label and him/her is tracked in different video taken by multiple cameras by the application of machine learning and image processing with the help of OpenCV. Many algorithms were used for detecting, recognizing and tracking till date, thus in this paper, main thing is the comparison of the proposed algorithm with some among the state-of-the-art algorithms. And also shows how the proposed algorithm is better than the other chosen algorithms.

Keywords—Detection; recognition; tracking; local binary pattern histogram; Kalman filter; particle filter

I. INTRODUCTION

The observation or monitoring of activity, behavior and other information by a system which includes several Closed Circuit Television (CCTV) cameras for observation and a set of algorithms to track a person is called Surveillance system. Technology has evolved a lot in the last few decades, previously there were no security cameras neither in Banks, Railway Station nor in other places. There were only security guards which protect these areas. Once the security cameras came into existence, it became easy to find people passing within the range of CCTV camera by simply searching through the videos recorded. Inventions increase people's expectations, although security camera reduces the human effort one has to search for an individual through the entire video which takes a considerable amount of time. So people thought what if the searching task can be accomplished by machine, it would save both human effort and time. A combination of machine learning algorithms with image processing is used to make a machine learn to reorganize a person and track that person in the given footage. The project is all about a system which has been designed to track a human in given videos.

Detecting a face from a video and tracking a person in that video is a very challenging task as there will be several changes including the resolution when a surveillance camera is used, and also the processing and computation of the frames taken from a surveillance camera in a real-time environment will be a critical and very challenging task because of many reasons. Some of the reasons where the difficulties arise are because of poor image quality, set up areas, limited networks, and compression of digital video, incorrect configuration, and

weather changes. Apart from these, there are some more serious problems that need to be taken into consideration are:

1) *Features*: Features of the person (spectacles, eyes, facial features, different hairstyles, body changes, facial hair) matter a lot while he/she is being tracked. Thus even after a period of time, the person who is already recognized earlier should be detected and recognized. Even if the person loses weight; if they have changed their hairstyle; if they are with/without beard and mustache; on the application of make-up and with/without spectacles, the person needs to be detected, recognized and tracked.

2) *Timings of the day*: Even if it is Morning/Afternoon/Evening/Night, the person needs to be detected and tracked.

3) *Occlusion*: If too many persons or even if a crowd is passing by, then also it should be able to identify and track the person.

4) *Weather*: Whatever may be the weather condition, detection and tracking should be efficient.

5) *Facial coverage*: Nowadays men and women cover their faces due to high temperatures, dust, and pollution.

6) *Age*: As a person ages too, facial features change due to the reduction of muscle and formation of wrinkles as the skin loosens.

7) *Twins/Triplets*: It is a bit tough for humans only to differentiate between twins/triplets.

These are some of the serious problems that are needed to be taken into consideration while implementing the algorithms [6]. Suspicious behaviours in shopping malls is also another serious problem, which also has to be considered [5].

Focusing on Face detection, there are many algorithms which had evolved from time to time. As the difficulties, objectives, and requirements increase, more efficient algorithms are being generated for accuracy and effectiveness as outcomes. Some of the detection algorithms by Colour, Motion, Model-based face tracking, Edge-Oriented Matching [1], Hausdorff Distance [2], Weak Classifier Cascades [3], Histogram of Oriented Gradients (HOG) and Deep Learning [4], etc. Many approaches are also there for the detection of a face like the frame-based detection, which is designed for still images and Typical Approaches, Real-Time & Multi-View Approaches [7], [8].

In this paper, the main point is comparison of different tracking algorithms from the state-of-the-art with the proposed algorithm which uses Local Binary Pattern Histograms

(LBPH) [9] and Tensorflow for detecting, recognizing and tracking the recognized person in other videos. LBPH approach is compared with Principal Component Analysis (PCA), Linear Discriminant Analysis (LDA) for animal recognition previously [10], which showed efficient results for a small test dataset. Tensorflow can be used in heterogeneous and large scale environments; it is mainly an advantage when it comes to the performance of an algorithm [11].

II. APPROACH

Firstly, the human detection phase takes place, then face detection and then face recognition. As soon as the detection is performed in the video, the rectangle will be drawn and tracking will be done through the video. It doesn't mean that the faces detected in the video is similar to humans detected because, only the trained faces will be recognized and if there are untrained faces, they will just be detected as humans, but the face cannot be detected. After training some people on the machine, it now tries to recognize the individuals. After all these steps are successfully completed, then it moves to the next phase i.e., tracking of the recognized individual in the video.

After taking the video as input, it then sets a Histogram of Oriented Gradients (HOG) [12] descriptor for the human detection using the help of Support Vector Machine (SVM) and OpenCV libraries. Non-maximum suppression is used to reduce false positive cases.

Then, in the second phase, Haar features are used to detect human faces. The detected faces will be shown separately in another window. After the face detection phase, a person should be mainly recognized before his face gets recognized. Local Binary Pattern Histograms (LBPH) and Haar cascades are used for training the machine. Detected faces are considered as datasets and are used to train the machine and the LBPH values for the respective faces are stored in a YAML file.

III. FLOW CHART

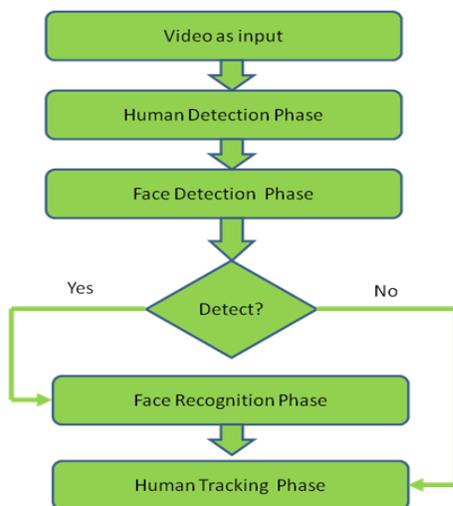


Fig. 1. Describes the Various Algorithms used in different Phases.

IV. DATA FLOW DIAGRAM

Faces can be recognized by importing the values stored in the YAML file into the memory and LBPH face recognizer program tries to recognize the faces using the LBPH features. After the recognition of a face, a label will be given to the person. As soon as a person gets recognized in camera 1, he/she is labeled as subject1, and will be tracked with the same label even if he/she passes through camera 2. The step by step procedure involved in all the phases is displayed through the Fig. 2 and the flow of data is shown from one phase to another.

Main aim is to compare the proposed algorithm with the Kalman filter and Particle filter and compare the results. Kalman Filter is moderately consistent in tracking. And it was only tested for tracking a single person [14]. Kalman filter predicts the present state variables and their uncertainties, and after the outcome of the next state, previously predicted values will be updated with some weighted average. But in real-world applications, Particle filter [15] can also be used for non-linear systems and non-Gaussian systems as Kalman Filter is not that efficient to handle the non-linear systems. Now comparison of these algorithms with the proposed algorithm is done.

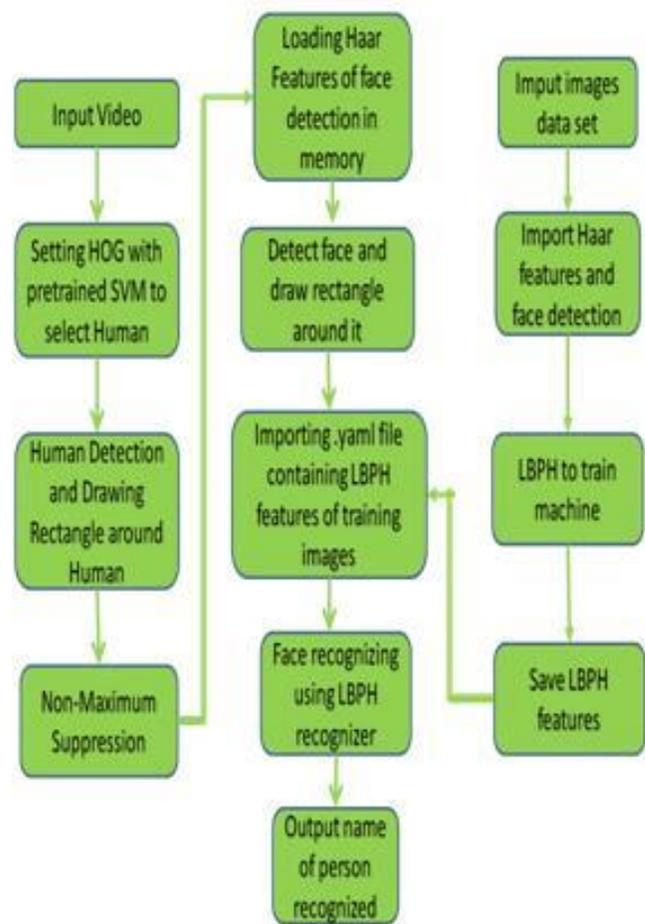


Fig. 2. Data Flow Diagram Showing the Flow of the Algorithm.

V. EXPERIMENTAL RESULTS

The proposed algorithm has been tested within some video sequences captured by a camera. Its purpose is to detect humans, recognize and track them throughout the sequence. Below are the tables which give the results of the performance of the proposed algorithm when compared to the other chosen algorithms.

TABLE. I. THE COMPARISON OF DETECTION RATE

Sequences	Algorithms		
	Kalman Filter	Particle Filter	Proposed Algorithm
Sequence1	89.8%	90.3%	91.4%
Sequence2	76.5%	77.1%	78.7%
Sequence3	66.2%	82.4%	84.3%

TABLE. II. THE COMPARISON OF RECOGNITION RATE

Sequences	Algorithms		
	Kalman Filter	Particle Filter	Proposed Algorithm
Sequence1	82.4%	82.5%	87.6%
Sequence2	72.6%	70.6%	74.7%
Sequence3	-	71.4%	73.8%

TABLE. III. THE COMPARISON OF TRACKING RATE

Sequences	Algorithms		
	Kalman Filter	Particle Filter	Proposed Algorithm
Sequence1	77.3%	79.6%	82.9%
Sequence2	65.6%	69.7%	72.1%
Sequence3	-	62.5%	71.2%

From Table I, II and III, comparison of the performance of the three algorithms can be clearly seen. Kalman filter algorithm had given less performance when compared to other two, because Kalman filter in rapid situations, it showed very slow reaction speed and was unable to recognize the person and also couldn't track them. But if Extended Kalman Filter, Unscented Kalman Filter [13] were used, then could get some positive results and Kalman Filter has the inability to manage dynamic background pixels. On the other hand, Particle Filter showed better results when compared to Kalman Filter, but it is not effective in occlusion environments and also when fast rotations take place computational time increases.

VI. RESULTS

As the methodology, working of the program are discussed and shown in the data flow diagram and the flow chart above, now showcasing the results against some test cases. The output of each and every test case using the screen shots of the output is provided. The program will require more computation as the training dataset size increases and also depends on the system configuration. After trying the program on two systems which were available with us, one of which

had the following specifications: AMD A4 2.2 GHz dual-core processor and 4GB RAM and the other system had the following specifications: Intel i5 2.4 GHz Quad-core processor 8 GB RAM. For an input video of 30 frames per second, the 1st system gave an output speed of just 4-5 fps. On the other hand, the 2nd system gave an output with a speed of 10-12 fps.

Initially, we tested the performance of the algorithms by considering the homogeneous environment. So the recognition of faces can be done only if training is done only for those faces beforehand. So the machine is trained for two people. One of them is labeled as 1 and the other one is labeled as 2. The output for four videos is shown, where each of the test videos represents different scenarios. Different stages of face detection and face recognition are shown in the results provided by the program. Here are the following screen shots of the different stages of the output obtained accordingly to the given input test videos.

The screen shots of the output provided by the program for the first video are shown in Fig. 1. In this video, only one person will be seen. The machine is trained to recognize that person and label him as 1. Initially, the frame is empty as seen in Fig. 3(a). When the person enters the frame, the program correctly detects the presence of a person and draws a rectangle around him. The program has correctly recognized the presence of person 1 as seen in Fig. 3(c) and it continues to track his face and correctly label it as 1. A false positive is triggered due to the sleeve of person 1 and the program has incorrectly detected it as a face and labeled as 1 as it can be seen in Fig. 3(o). Then again, the program has removed the false positive and correctly recognized person 1 in the upcoming figures until the person 1 goes out of the frame. In most of the cases, the human body is not detected because the full human body is not present in those frames.

The screen shots of the output provided by the program for the second person given in Fig. 4. Similar to video 1 in this video too only one person will be seen. The machine is trained to recognize him and label him as 2. Initially, the frame is empty seen in Fig. 4(a). As the person enters the frame, the program correctly detects the presence of a person in the upcoming frames as it can be seen in the figures. The program recognizes the person 2 and labels him for the first time in Fig. 4(m). The person continues to correctly label the face of a person 2 as 2 till the person moves out of the frame.

The screen shots of the output provided by the program for the third video are given in Fig. 3. Unlike previous videos, person 1 and person 2 both appear in this video, but they enter the frame one after the other and not together. Initially, the frame is empty as seen in Fig. 5(a). Then the person who is labeled as 1 in our training samples enters the frame. In Fig. 5(d), the program has recognized the presence of person 1 for the first time and has even successfully detected the face of person 1. Since in the frame, as they enter the body of person 1 is visible, the program has correctly drawn a rectangle around him as seen in Fig. 5(e). The system is successful in detecting the face of person 1 and labeling it as 1 till the time he is present in the frame.

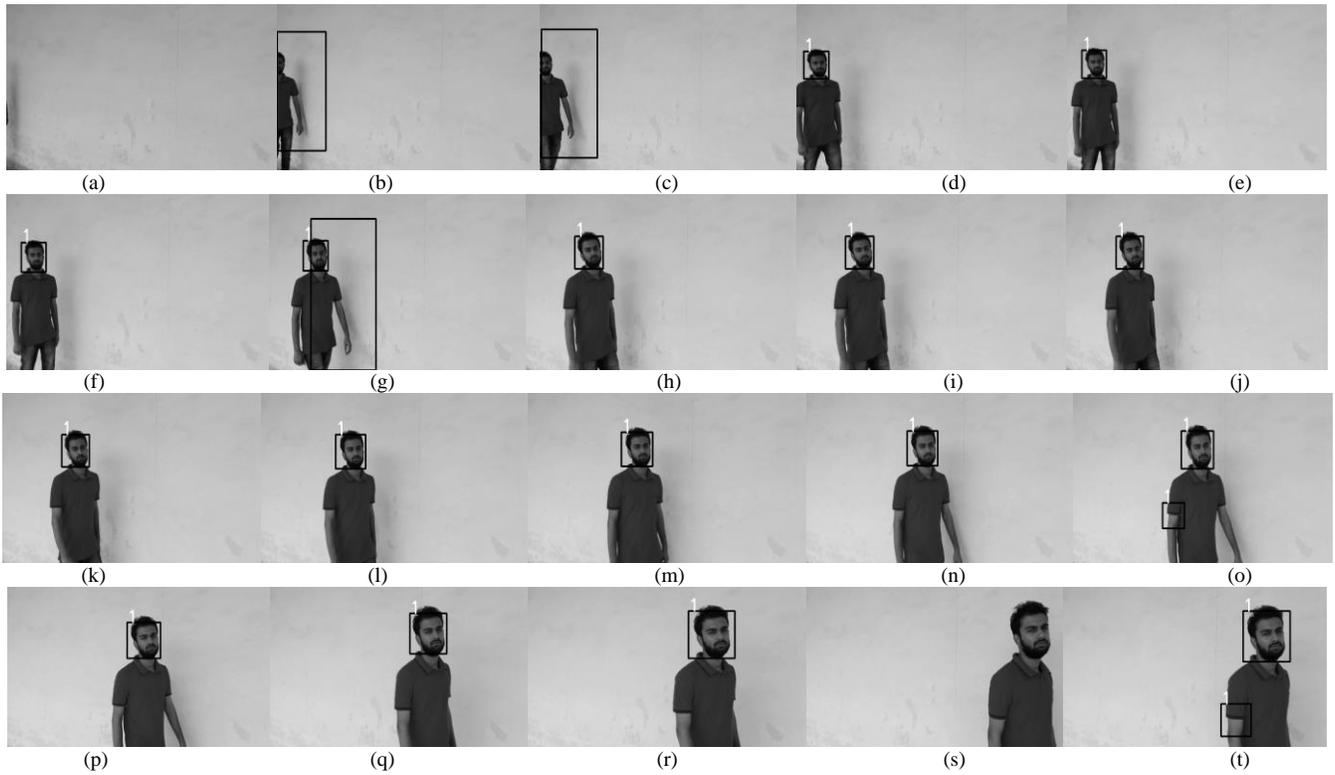


Fig. 3. Output of Video 1: Single Person Video; Frames o, True False Face Detected Cases, no Face Detected in Frame s, in many Images Human is not Detected Due to the Absence of Full Human Body.

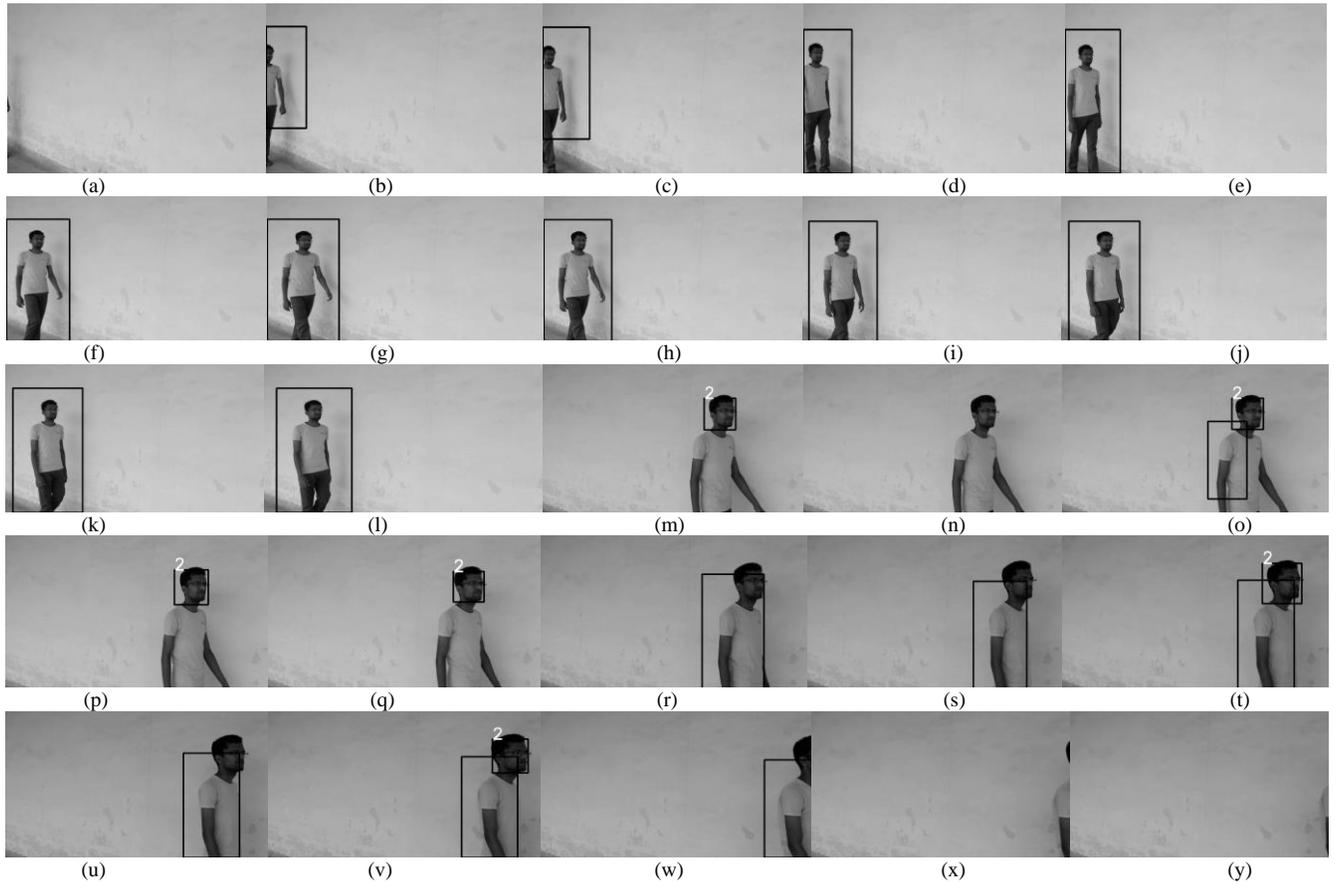


Fig. 4. Output of Video 2: Single Person Video; no Face Detected in many Images due to unclear Face.

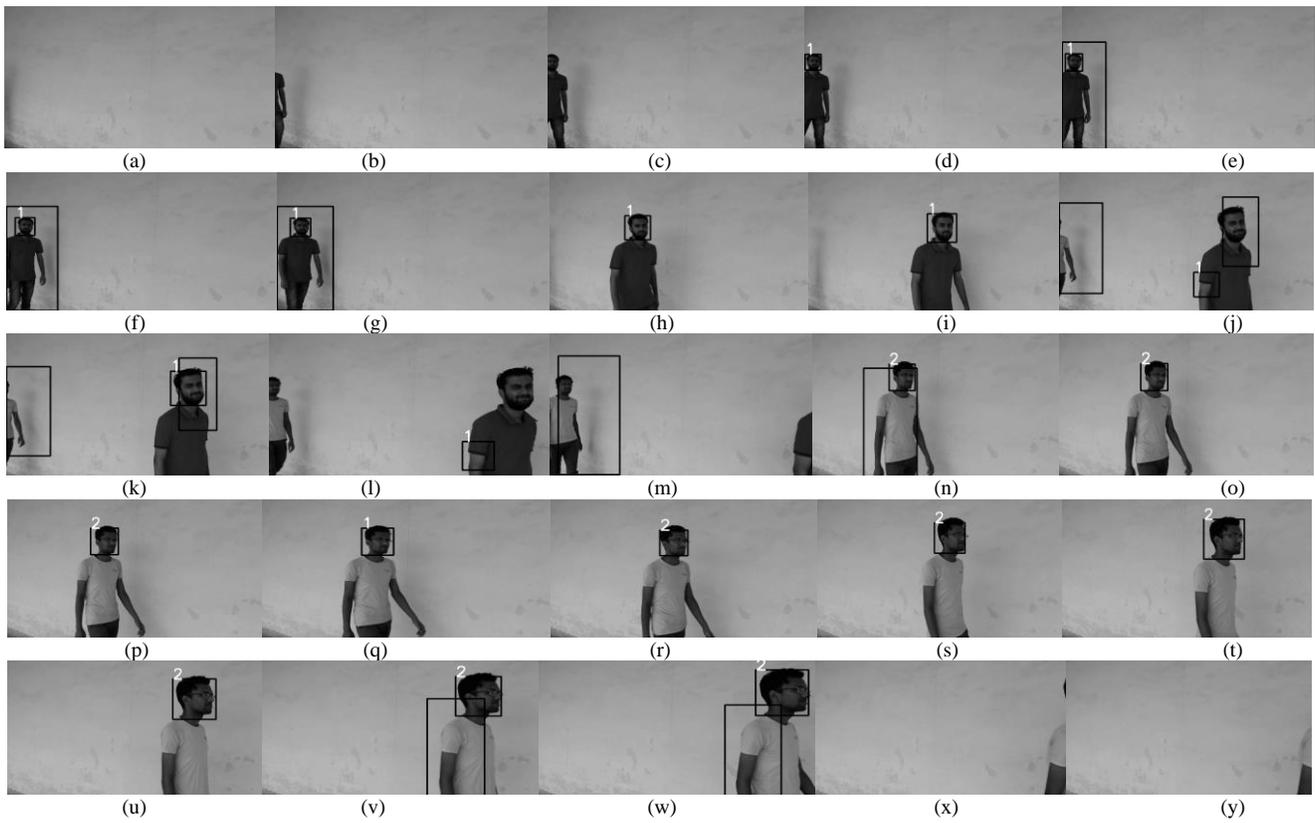


Fig. 5. Output of Video 3: Multiple Person Video, Walking One after other; False Faces are Detected in Frame g, j, l, and Frame q is Wrongly Detected as Subject 1.

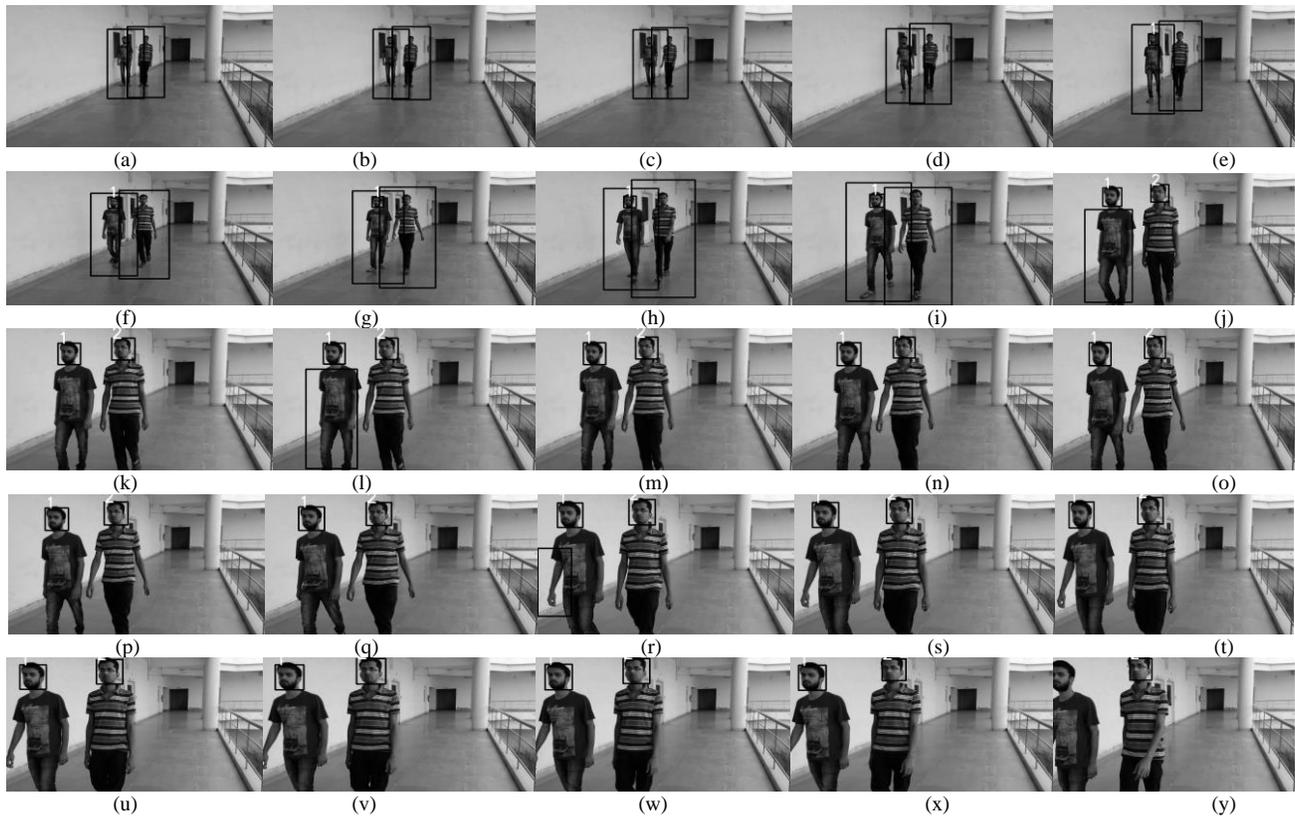


Fig. 6. Output of Video 4: Multiple Person Video, Walking Together; no Face Detected in Frames. a-d Due to the Larger Distance from the Camera.

Although, a false positive was detected by the system in Fig. 5(j) where the system false fully detects the sleeve of person 1 as a face. In the same frame, since the person 2 also enters the frame, a rectangle is drawn around him. Till Fig. 5(l), the presence of a person is correctly detected by the program and is shown by drawing a rectangle around that person. Now in Fig. 5(m), we see that person 1 is completely out of the frame and person 2 is correctly detected. The program was successful in detecting a face in Fig. 5(n) and has correctly labeled it as 2. The program then again fired a false positive as we can see in Fig. 5(q), it has incorrectly labeled person 2 as 1. In Fig. 5(r), the program has recognized and labeled the person 2 correctly as person 2 rectifying the false positive. The program continues to label the person 2 correctly until the person goes out of the frame.

The screen shots of the output provided by the program for video 4 are shown in Fig. 6. In this video, both person 1 and person 2 appear together in the same frame. We can see here since the first frame, the program has correctly detected both the persons. It has drawn a rectangle around both the humans as it can be seen in the figures given above. As they are moving towards the camera, the rectangle is drawn around them, also move with them in one way tracking them. In Fig. 6(e), the program has successfully recognized the person 1 and labeled him as 1. And it continues to track both of them until it finally recognizes the person 2, also as seen in Fig. 6(j). In this and the following figures, it can be seen how the program has successfully recognized both the persons and has correctly labeled them. As it can be seen in Fig. 6(n), the program false fully recognized the person 2 as person 1 and thus labeled person 2 also like 1, but in the following figures, the program has correctly recognized the person 1 as 1 and the person 2 as 2 and it continues to recognize them till the last frame. In this video, we can get better human detection because, in many frames, the full body of a human is present.

VII. CONCLUSION

The proposed algorithm focuses mainly on human detection, recognition, and tracking, given a video as input to the program and can get the labels of the identified persons in the video. Only the limitation is that the user needs to train the machine beforehand, and then it will recognize the persons. A train.py program is used to train the machine for the people to be tracked. It took 30 fps when the input is given and we got 15 fps as the output. Also as the system configuration increases, better results will be shown. At some point of severe occlusion conditions, it doesn't give such accurate results. In future addition of a feature that the program itself

learns how to detect a new person whenever they appear and thus a label generation should be done automatically. So that whenever same person re-appears next time, he/she would be easily identified on its own and thus it implies the reduction of computational speed and also false positives.

REFERENCES

- [1] B. Froba, Ch. Kulbeck, "Real-time face detection using Edge-Orientation matching", in the Conference on Audio- and Video-Based Biometric Person Authentication, 2001.
- [2] O. Jesorky, K. J. Kirchberg, and R. W. Frischholz, "Robust face detection using the Hausdorff distance", in the proceedings of the Third International Conference on Audio- and Video-Based Biometric Person Authentication, pp.90-95, 2001.
- [3] P. Viola, "Robust real-time face detection", in the International Journal of Computer Vision, pp. 137-154, 2004.
- [4] S. Chang, D. Xiaoqing, F. Chi, "Histogram of the oriented gradient for face recognition", Tsinghua science and technology, vol. 16, pp. 216-224, April 2011.
- [5] R. Aroyo, J. J. Yebes Luis, M. B. Selvan, G. D. J. Almazan, "Expert video-surveillance system for real-time detection of suspicious behaviors", in Expert systems with Applications, vol. 42, pp. 7991-8005, November 2015.
- [6] T. Nagaria, Dh. Chourishi, "A comprehensive survey and detailed study on various face recognition methods", in the International Research Journal of Engineering and Technology (IRJET), vol. 5, December 2018.
- [7] S. Rakesh, S. Abhishek, Deepak Agarwal, "Analysis of different face recognition algorithms", in the International Journal of Engineering Research and Technology (IJERT), vol. 3, November 2014.
- [8] K. Sunil Reddy, "Comparison of various face recognition algorithms", in the International Journal of Advanced Research in Science, Engineering and Technology, vol. 4, February 2017.
- [9] S. Ranganatha, Y. P. Gowramma, "Image training and LBPH based algorithm for face tracking in different background video sequence", in the International Journal of Computer Science and Engineering (JCSE), vol. 6, September 2018.
- [10] P Kamencay, T Tmovszky, M Benco, R Hudec, P Sykora, A Satnik, "Accurate wild animal recognition using PCA, LDA, and LBPH", IEEE, 2016.
- [11] M Abadi, P Barham, J Chen, Zh Chen, et al, "Tensorflow: A system for large-scale machine learning", OSDI, Savannah, GA, USA, November 2016.
- [12] N Dalal, Bill Triggs, "Histogram of oriented gradients for human detection", 2005.
- [13] Ch Liu, P Shui, Song Li, "Unscented extended Kalman filter for target tracking", in the Journal of Systems Engineering and Electronics, vol. 22, pp. 188-192, April 2011.
- [14] C Suliman, C Cruceru, Florin M, "Kalman filter based tracking in a video surveillance system", in the proceedings of the 10th International Conference on Development and Application Systems, Suceava, Romania, May 2010.
- [15] F Wang, "Particle filter for visual tracking", G. Shen and X. Huang, CSIE 2011, part I, CCIS 152, pp. 107-112, 2011.