International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 07 Issue: 02 | Feb 2020 www.irjet.net p-ISSN: 2395-0072

Real-time Analysis of Video Surveillance using Machine Learning and Object Recognition

Smita Pawar[1], Anurag Bambardekar[2], Rahul Dhebri[3]

[1]Professor, Dept. of Electronics and Telecommunication, Xavier Institute of Engineering, Mumbai, Maharashtra, India
[2,3]Student, Department of Electronics and Telecommunication, Xavier Institute of Engineering, Mumbai, Maharashtra, India

---***---

Abstract - *Widely accepted standards for video surveillance are very poor and inadequate in critical situations and often they fail to recognize or even identify suspicious activities. Our goal is to explore the feasibility of our proposed methodology of implementing a novel video surveillance system. This paper aims at studying the various existing algorithms in computer vision and implement algorithms and techniques best suited for our system. The objective is to develop a better system which utilizing machine learning, computer vision and image processing algorithms to detect and analyze objects of interest in a scenario. We employ algorithms for detection and recognition of faces as well as suspicious objects and persons on the input feed provided by CCTV cameras, various criminal activities can be detected, and authorities will be assisted to take the desired action early as possible. Implementing effective security measures in critical environments is of the utmost importance and is very difficult as it involves a lot of IT infrastructure as well as human inputs.*

***Key Words*: Real-time, Video surveillance, Machine learning, Computer Vision, Image Processing, CCTV cameras, Face recognition, Object recognition.**

1.INTRODUCTION

Our objective is to build an effective novel framework which can be used across different domains. The proposed framework will have different aspects. The first aspect is to build an effective face detection mechanism. It should be secure & efficient. We aim to detect humans in CCTV footage from a single camera and eventually from multi-camera systems in an indoor environment or a restricted environment. We also want to analyze the detected faces and then estimate and find parameters such as facial features, age and gender as well as recognize the earlier detected faces. Tracking faces using head pose estimation is also desired to be achieved. Another aspect is to detect the events from the video. This is the most challenging part of the framework. It will include the algorithms to detect movements. As a business owner, one of the top priorities is protecting your property against theft and break-ins as well as dishonest employees. Remote surveillance to monitor your system live and react quickly to any activity on your site is possible through the surveillance system. Secure the perimeter of a property with video surveillance cameras to thwart trespassers and create a safer environment. Information obtained from CCTV can be used to classify different kinds of objects (e.g., pedestrians, groups of people, motorcycles, cars, vans, lorries, buses, etc.) moving in the observed scene, to understand their behaviours and to detect anomalous events. Crucial information like classification of the suspicious event, specific information about the class of detected objects in the scenario, etc.) can be transmitted to a remote operator for augmenting its monitoring capabilities and, if necessary, to take appropriate decisions.

2. Survey of Existing System

2.1 Automated Video Surveillance

Mrs. Prajakta Jadhav et al. wrote computer programs using the best suitable language/tool which with the help of behavioural analysis can understand routine things. It will learn with respect to time and will start reporting things which are abnormal. These abnormal things will further be reported to different entities like police or doctor or an individual for analysis. This project is combination of electronics and computer science. Proposes to detect abnormal events from recordings rather than in true "real time"Focuses on building an effective storing mechanism which should be secure & memory efficient.[1]

2.2 Real Time Facial Expression Recognition for Nonverbal Communication

This paper published by Md. Sazzad Hossain and Mohammad Abu Yousuf represents a system which can understand and react appropriately to human facial expression for nonverbal communications. The considerable events of this system are detection of human emotions, eye blinking, head nodding and shaking. The key step in the system is to appropriately recognize a human face with acceptable labels. This system uses currently developed OpenCV Haar Feature-based Cascade Classifier for face detection because it can detect faces to any angle. The false detection rate is increased due to variation in skin colour or lighting condition changes.

For Head nodding and shaking, their system can deal with small motions. So, it fails when there is large motion.

Since Haar Cascades are used, it is difficult to identify parameters such as emotions, blinking of eye and head nods from side profile of the face. [2]

2.3 Real Time Monitoring of CCTV Camera Images Using Object Detectors and Scene Classification for Retail and Surveillance Applications

Anand Joshi, in his paper, focussed on monitoring surveillance video and detect threat perception and theft scenarios. He chose datasets containing images of handguns, knives, human hand and everyday objects observed in the retail environment. Using these collections of images, he prepared three classes of image datasets. a) guns b) knives c) hand and d) Everyday Objectsobserved in retail environments [3] and created over 1000 labels accordingly in the database. Images from the following data sources were used for this purpose. Knives Images Database, which contains 9340 negative examples and 3559 positive examples, Internet Movie Firearms Database, which contains 8557 images, Hand Dataset which contains about 14700 hand images from various sources. EgoHands Dataset containing 120000 images. ImageNet dataset. which contains more than 1.2 million images in over 1000 categories. He trained and evaluated the data set on different models to see which one gives the best result.[3]

2.4 Real Time System for Facial Analysis

Janne Tommola, Pedram Ghazi, Bishwo Adhikari, and Heikki Huttunen in this work, describe the functionality of their demo system integrating a number of common real-time machine learning systems together. The demo system consists of a screen, webcam and a computer, and it estimates the age, gender and facial expression of all faces seen by the webcam. Apart from serving as an illustrative example of a modern human-level machine learning for the general public, the system also highlights several aspects that are common in real-time machine learning systems. [4]First, the subtasks needed to achieve the three recognition results represent a wide variety of machine learning problems: (1) object detection is used to find the faces, (2) age estimation represents a regression problem with a real-valued target output (3) gender prediction is a binary classification problem, and (4) facial expression prediction is a multi-class classification problem.[4] Moreover, all these tasks should operate in unison, such that each task will receive enough resources from a limited pool. The face detection uses the SSD detector with MobileNet.

2.5 Detection of Real Time Objects Using TensorFlow and OpenCV

This paper by Ajay Talele, Aseem Patil and Bhushan Barse introduces a new computer vision-based obstacle detection method for mobile technology and its applications. Each individual image pixel is classified as belonging either to an obstacle based on its appearance. The method uses a single lens webcam camera that performs in real-time, and also provides a binary obstacle image at high resolution. In the adaptive mode, the system keeps learning the appearance of the obstacle during operation. The system has been tested successfully in a variety of environments, indoors as well as outdoors, making it suitable for all kinds of hurdles.

System.This paper presents a new method for obstacle detection with a single webcam camera. It also presents a new method of vision-based surveillance robot with obstacles avoidance capabilities for general purposes in indoor and outdoor environments.[5]

YOLO imposes strong spatial constraints on the bounding box predictions since each of the grid cells only predicts two boxes and can have only one class.

This spatial constraint then limits the number of nearby objects that our model can predict.

The model struggles with the small objects that appear in groups

3. Proposed System

We present a new method to robustly and efficiently analyze CCTV footage in real-time. We propose a fully automatic and computationally efficient framework for the analysis of Real-Time Video Surveillance.

OpenCV is an open-source computer vision library that contains image processing functions and over 2,500 algorithms used for things like facial recognition.OpenCV can accelerate CUDA and OpenCL GPUs. OpenCV supports deep learning platforms like TensorFlow. OpenCV is built using a layering process.

We use OpenCV to perform Human Face analysis and extract facial features, track the faces, detect age, gender and other parameters which are essential to profile a person. We also perform Movement analysis in a closed environment to monitor the subjects and eventually detect for anomalies.

TensorFlow is a platform that is based on dataflow graphs and is useful in training with deep neural networks. We utilize Google's TensorFlow API to create a digital framework that will identify handguns and knives in real-time video. By utilizing the different models, our system is trained to identify handguns and knives in various orientations, shapes, and sizes, then the intelligent gun/knife identification system will automatically interpret if the subject is carrying any suspicious object. Our experiments show the efficiency of the implemented intelligent gun/knife identification system.

Currently, code models and libraries such as TensorFlow, OpenCV, dlib etc. for object detection identification have been examined. First trials were on a machine which does not have GPU support for these frameworks. Subsequently, we started to work on a machine having GPU support.

We have worked on a pre-trained model named Mobilenetv1 with TensorFlow. The model is tested with various test images.

Major objectives: Face Detection, Face Landmarks Extraction, Face Recognition, Age & Gender Estimation, Human Pose Estimation, Weapon Detection, Detect MotionTrajectory Tracking and Alerting Concerned Authorities.

4. Algorithms:

4.1 Object Detection: Tensorflow was used, which is Google's open-source machine learning library for carrying out the task of object detection and recognition and TensorRT engine was used to build the model.

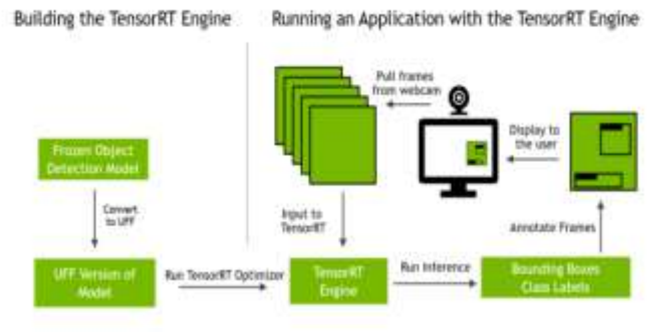

Figure -1: Building and runningTensorRT engine

4.2 Face Detection: For achieving the goal of face detection major face detection techniques were used and compared. Haar Cascade was used in the first stage for recognition but it suffered when a side profile of human face was presented. So later we moved on to use 'facerecognition' library.Face Classification by using CNN takes input as an image, then it processes it by extracting feature classses and classify them into the different categories. The hidden layer of CNN consists of Convolutional layer, Activation Function(ReLu, Sigmoid & any other), pooling layers, fully connected layers and normalization layers.

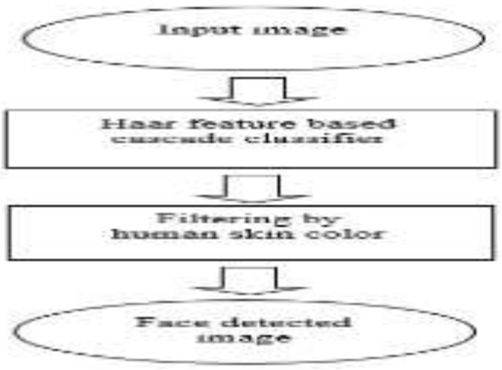

Figure -2:Haar-Cascade Face Detection[5]

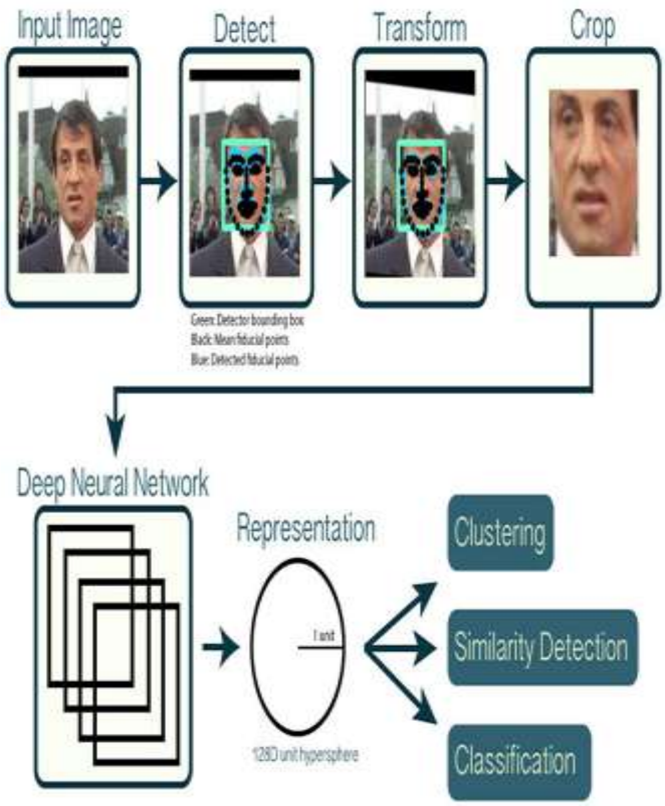

Figure -3: Face detection using DNN

5. Results:

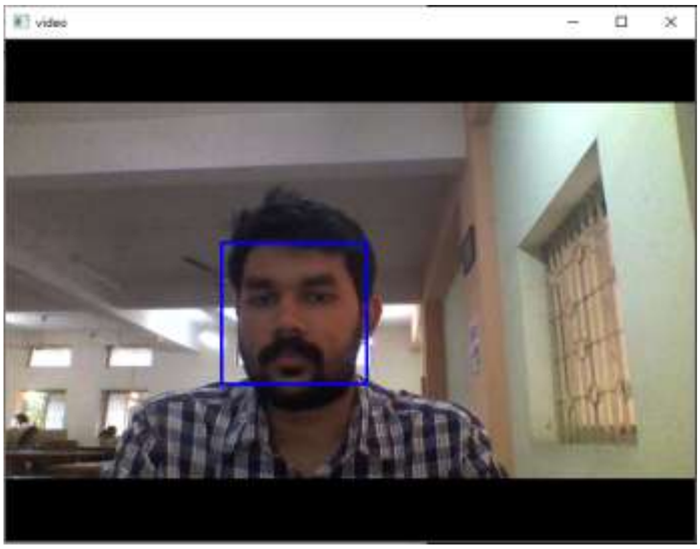

Figure -4: Face Detection Using Haar Cascades

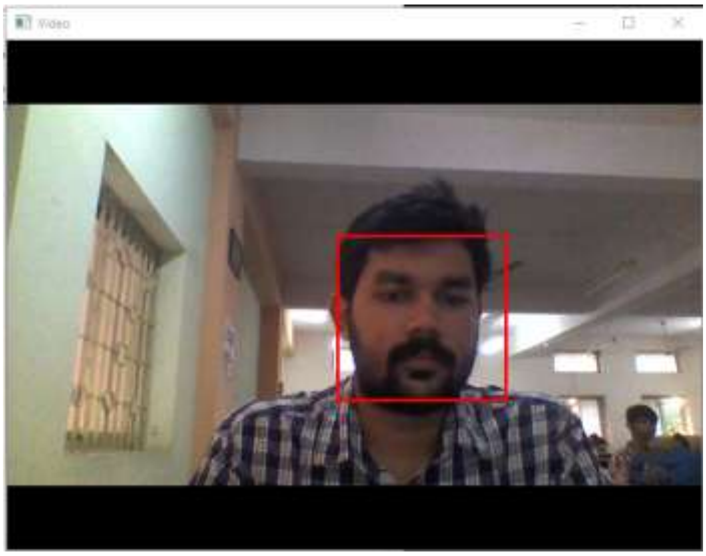

Figure -5: Face Detection using 'facerecognition' library

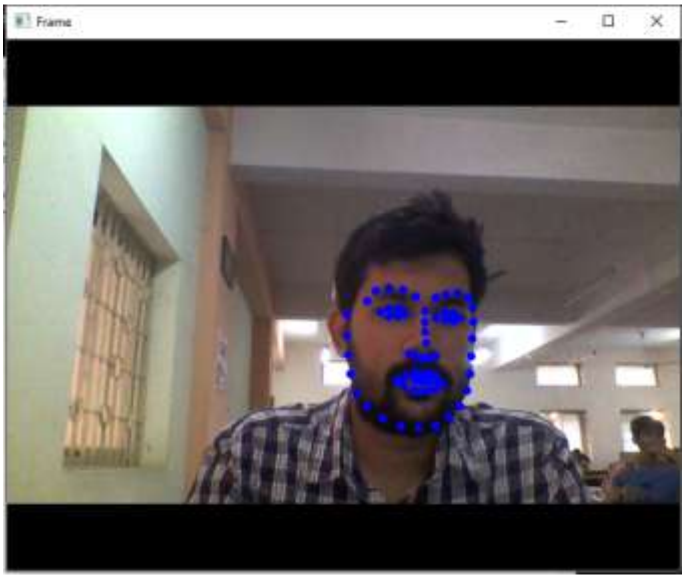

Figure -6: Landmark Extraction

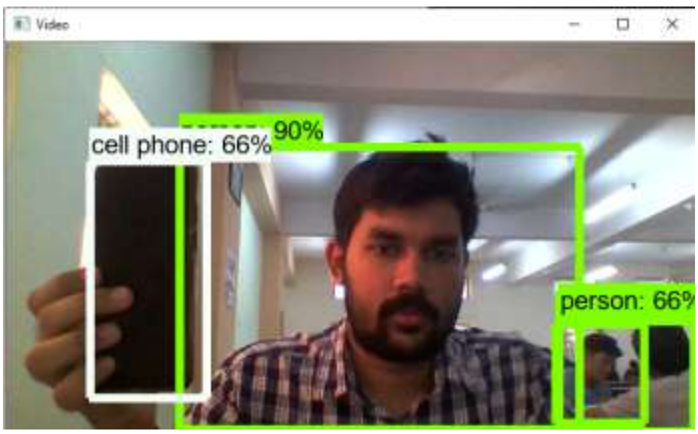

Figure -7: Object Detection

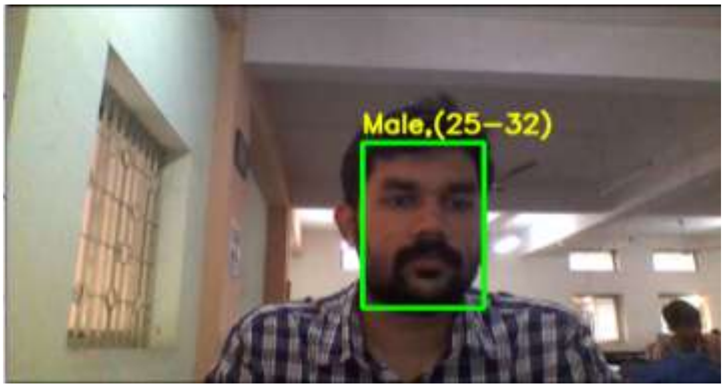

Figure -8: Age/Gender Estimation

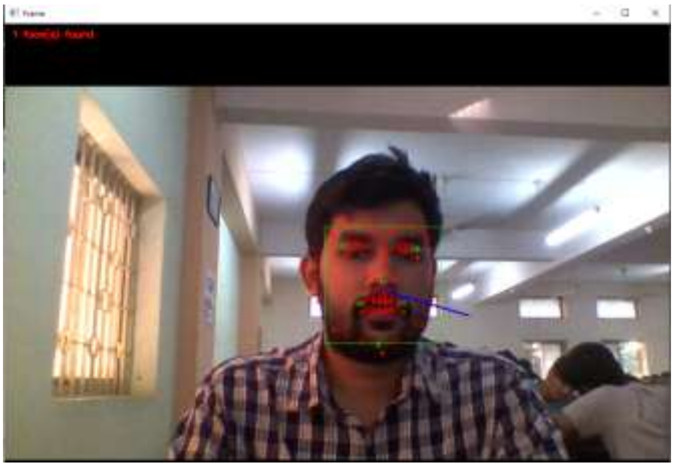

Figure -9: Head Pose Estimation using 68 point model

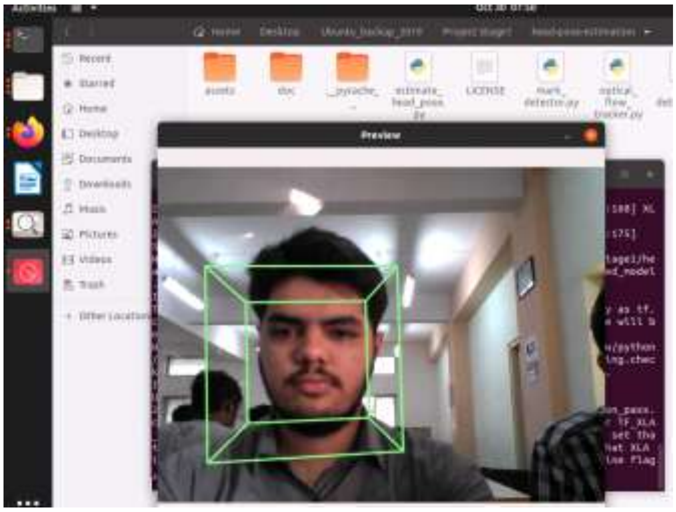

figure -10: Head Pose Estimation

Figure -11: Human Pose Estimation

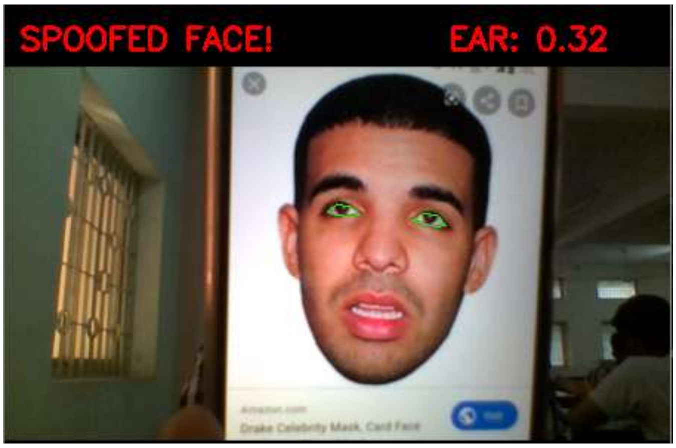

Figure -12: Spoof Face Detection

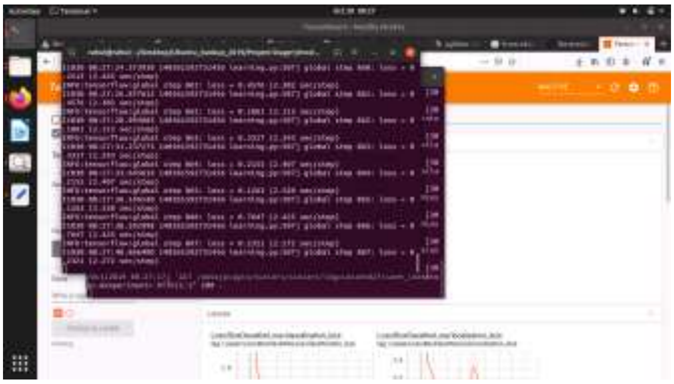

Figure -14: Model results on Tensorboard

Figure -15: Training.

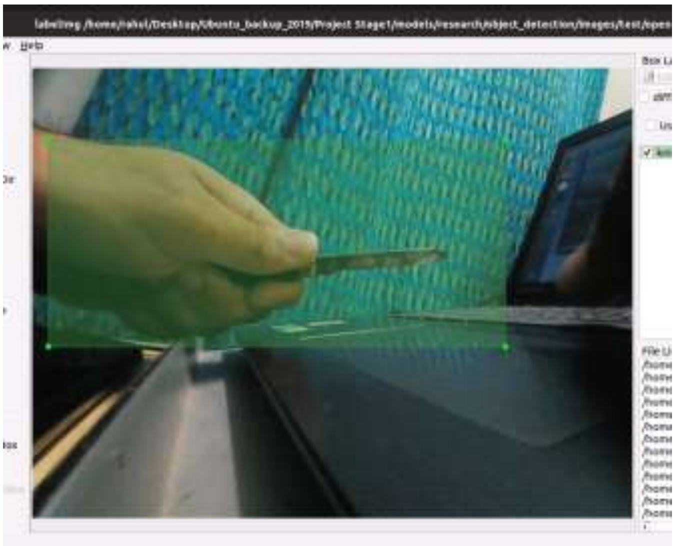

Figure -16: Knife detection

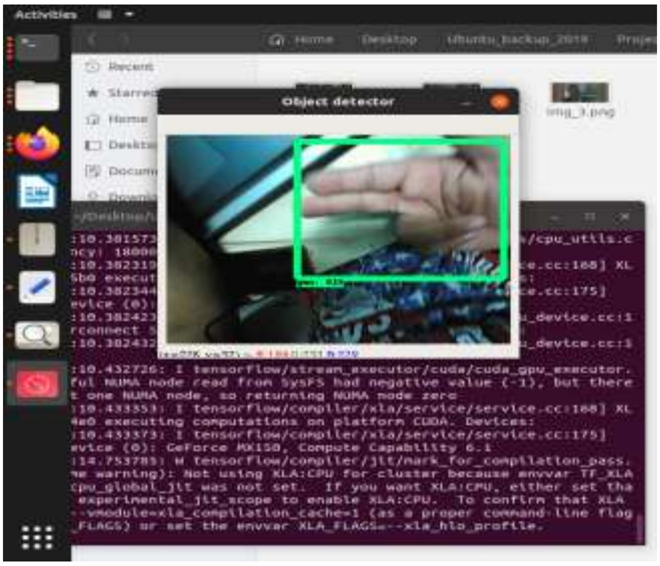

Figure -17: Gun Gesture detection

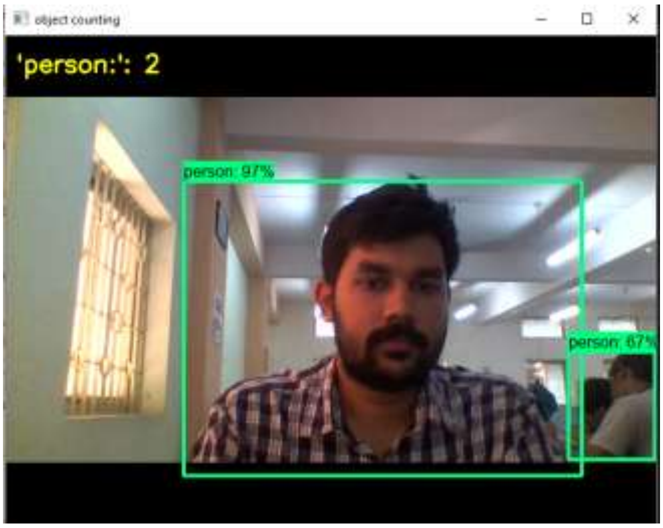

Figure -18: Pedestrian counting using Tensorflow

Figure -19: Motion tracking using Background subtraction

Figure -19: Occlusion of face by a Helmet

6.Conclusion: In this paper, we have presented a prototype system for real time analysis of surveillance video. This research has considerable implications for the effective operation of CCTV surveillance. The information we extracted was sufficient to enable not only the generation of accurate, human-readable commentary on surveillance video such as facial anlysis, motion tracking and anomaly detection in video frames. Image and video-processing techniques have been implemented that could be used within a semi-automatic process to help operators maintain global situational awareness of the entire scene when focussing on potentially interesting activity.

ACKNOWLEDGEMENT

We would like to thank our project guide Prof. Smita Pawar who has been a source of inspiration. We are also grateful the authorities, faculties of Xavier Institute of Engineering who have helped us to be better acquainted with recent trends in the technology.

REFERENCES

[1] Jadhav, Mrs Prajakta, Mrs Shweta Suryawanshi, and Mr Devendra Jadhav. "Automated Video Surveillance." (2017).

[2] Hossain, Md Sazzad, and Mohammad Abu Yousuf. "Real time facial expression recognition for nonverbal communication." Int. Arab J. Inf. Technol. 15.2 (2018): 278-288.

[3] Joshi, Anand. "Real Time Monitoring of CCTV Camera Images Using Object Detectors and Scene Classification for Retail and Surveillance Applications." (2017).

[4] Tommola, Janne, et al. "Real time system for facial analysis." arXiv preprint arXiv:1809.05474 (2018).

[5] Talele, Ajay, Aseem Patil, and Bhushan Barse. "Detection of Real Time Objects Using TensorFlow and OpenCV." Asian Journal For Convergence In Technology (AJCT) (2019).