e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
Volume:02/Issue:08/August-2020 Impact Factor- 5.354 www.irjmets.com

A NOVEL APPROACH FOR TRACKING CROWD AND MOVING VEHICLES IN SURVEILLANCE VIDEOS

Ms. Kathyayini Vishwakarma L V*[1] , Mr. Keerthi D S*[2]

*[1]Department of Electronics & Communication, Malnad College of Engineering Hassan, Karnataka, India.

*[2]Assist Professor, Department of Electronics & Communication, Malnad College of Engineering Hassan, Karnataka, India.

ABSTARCT

At present we are approaching a method for detecting the movements of moving person, vehicles such as car and motorbike along with crowd in a surveillance camera system and live stations. By using single shot deep neural network called SSD. It works like creating a number of default boxes for set of images to map their features as exactly compare with the ground truth boxes. At the output time it sends a lot of the matched boxes with different aspects ratios, scales and shapes and adjust similarities with the original images and predict whether it is human or vehicles etc. Additionally it gives different resolution pictures features to handle with it size and shape. SSD eliminates all the computations which done in other systems like subsequent re-sampling stage. Encapsulate all the computation in a single neural network only. The other systems like PASCAL, VOC, MSCOCO, RCNN etc. The SSD is also having good accuracy result rate and faster also. It is done with smaller input image size of 300*300 and also achieves 72% mAp on VOC2007 test to compute the detection process.

KEYWORDS: Person Detection, Abnormal Crowd Detection**,** Identifying vehicles, SSD Network model, Python Machine Learning.

I. INTRODUCTION

In the present situation abnormal behavior detection in crowd area is essential because of modern technology and culture. It is one of the interest area of research observing the moving things in a video's or webcam is tough task and computer vision also risk, so for this we have a potential to discover many algorithms and application. SSD is also one among them. The surveillance system is present in everywhere especially in malls, government institution, private sector, companies, ATM sector, banks, private events, house etc. There are the areas where we use surveillance cameras in order to identify the abnormal movements. To achieve good vision analysis we must have good tool for them so we are using good machine learning language also it is done by systems to elaborate the instruction further. Having a better intelligent network we can clearly observe the events. SSD is helps in that area it an implement convolutional filters for mapping the features in different aspects ratio's and scales so it gives different images to compare with originals. We can clearly identify the essential objects and activities and achieve success in our work. Python is a new machine learning language having fast in processing program and execution is robust with the things. Network installation and many modules with package are there to convert the image frames and read them and analyze etc. Due to modern lifestyle and digitalization effect safety camera is part of everybody's life to keep one eye on any events in our risky works so all these processing applications are helpful and we have to improve more innovations in this fields our project will one of them to have small feature extraction from loaded video's and webcam's. Automatic machine analysis and detecting the abnormal crowd is the main concept and with motion analysis also for vehicles and humans distance are done in this concept and achieved by efficient programming.

II. METHODOLOGY

In this approach mainly uses the single shot detector method and feature extraction techniques to recognize the both normal and abnormal activities in a group. With the help of different algorithms we can perform object identification. SSD is the key tool algorithm in this method and which employees a convolutional neural network which is organized as Base convolutions, Auxiliary convolution and

prediction convolution. SSD300 and SSD512 are famous now a days and here we employees SSD300*300 convolutional network for object feature mapping along with this network model we initialize the camera systems and read the frames from camera and convert them into binary large object with size 300*300 and achieve the feature mapping of humans, vehicles, object etc. It use python language for executing program, first it read the network then initialize as per program conditions it run the pre processing program count the confidences of frames and gives crowd result and detect human and vehicles.

a) System Design

The system design consists of the following blocks namely:-

- Initialize Camera system.
- Read image frames from Camera.
- Convert image frames into BOLB with size 300*300.
- Input the BLOB into SSD network to fetch directions.
- Initialize direction to frames.
- Call distance function to frames of BLOB.
- Displaying result crowd and identify abnormal activities of humans and vehicles etc.

In this section we have upload a video's or live video's from camera system. The video's are considered and analyzed by their frames one by one, so first we have to read those image frames for further identification and other functional process, the important thing of this block is to initiates the Camera system to observe / record the events .In this process the uploaded video frames are considered for read. The frames are read one by one and analyzed. With help of open CV library function and can extract the useful feature from frames by using some common image scanning techniques, Also some editing functions are done here with the help of open CV library processing techniques. Here we get the image frames of exact features for identifications are converted into array by using numpy convert the object frames. The BLOB means binary large objects here the binary data's which are stored the value of image in binary form and add pixels values such as 300*300 to network and the pixels values of image are stored with the help of array functions. This process involves the binary large object of 300*300 files is upload to the SSD [single shot detector] neural network. SSD is a very good tool to image identification that is their position, number of persons or vehicles or crowd recognize. SSD is neural network we initialize in our program o find the result. SSD is a most fast detecting system compared with other categories such as YOLO and RCNN network. SSD need only an input image and ground truth boxes to start object identification in its training. The boxes are arranged in a fashion of convolutions with different aspects ratio's in every locations of image. If maps several times with different scales such as 8*8 and 4*4 etc show in above figure For every defaults boxes, we analyze shape offsets and confidence to each objects categories $[(c_1,c_2.....c_p)]$ During timing just match the defaults boxes with ground truth boxes. In the above figure(3) the cat and dog are matched with two defaults boxes and treated as positives and remaining things considered as negative.

b) Working principle of SSD – Network

SSD architecture is mainly involves feed – forward convolution network which gives the fixed size collections to the object of defaults boxes to produce the final output. The working principles of SSD involves following process:-

- Multi-scale feature maps for detection
- Convolution predictors for detection
- Default boxes and aspects ratios
- Training Matching strategy
- Training objectives
- Choosing scales and aspects ratio's for default boxes.
- Hard negative mining
- Data augmentation

Here convolution filters feature layers is send to the end of the truncated base network and using VGG-16 network as a base. The layers used here are progressively decreased and gives good analyzed prediction at different scales. For every feature layer we use multiple and separate convolution model, for detecting the features. For the purpose of good detection we are using a set of convolution filters. For every feature layers that is an existing feature layer which is connected to base layer can produce a number of detecting outputs by make use of these convolution filters that are displayed on the top of the architecture SSD. The size of the feature layers is m*n with p-channels. So the primary analyzing parameter is 3*3*p that produces a number of detections at different categories with respect to its shape, size offsets in defaults box relatively. Kernal is applied to each m*n pixels locations and generate output value. Then the defaults box positions are responsible for the measure of bounding box output values in every feature mapping locations. The bounding boxes are associated with the top of the network in its every feature mapping. In convolution manner the default boxes are tile or attached to feature map, and every position of default box is considered. The position of every box is presently relative to its corresponding cell is fixed. In every cell mapping feature technique process the offsets shape of default boxes in the cell are detected. For every box at 'K' locations, calculate 'c' class scores and 4 – offsets values in compare with original primary default box shapes. So total (c+4) k filters are involved in mapping process at every locations features. So totally relates to (c+4) k.mn outputs for the mapping of m*n.

The small change in training SSD and the training detector is uses the region of approaches and pooling before a final classifier, so the information in ground truth boxes are essential to assigned in the fixed group of detectors outputs. The main thing involved in training is selecting the group of default boxes and respective scales for their detection and data augmentation schemes with tough negative mining. During training period only we have to allocate the correspondence between these two boxes namely default box and ground truth box. Depending on the location every data taken from default box is vary with respect to the ground truth box, with scales and aspects ratio's also vary. Jaccard overlap method is used for matching these two boxes, it should except that everything is matched perfectly with conditions of jaccard threshold valve [0.5],So the network is capable of producing good confidence of multiplying overlapping default boxes instead of selecting only one at maximum overlap .The main objective of training is taken from multiple object training categories. If $x^p ij=1$ if the default boxes are perfectly matched with the ground truth boxes, else $x^p ij=0$. Where i = number of default boxes and j = number of ground truth boxes. The main aim of every network is to reduce the size of its convolution neural network at each layer. It helps in reducing operation and memory consumption time and space with cost effectively and improves high level of translation and scaling effect. So choosing different sizes and calculate every size and finally merged their results. So by comparing lower and higher layer like 8*8 and 4*4, the lower layer contain more information in it because of segmented with base as well they are finally attached to the top and output layer.

III. MODELING AND ANALYSIS

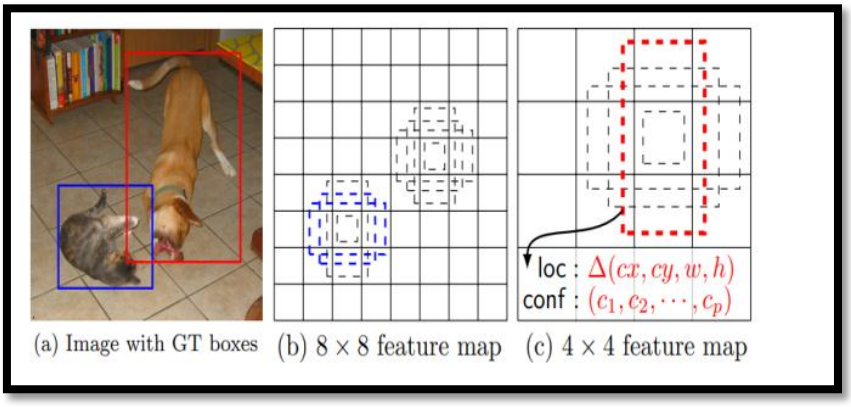


Figure-1: Mapping at different scales.

e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
Volume:02/Issue:08/August-2020 Impact Factor- 5.354 www.irjmets.com

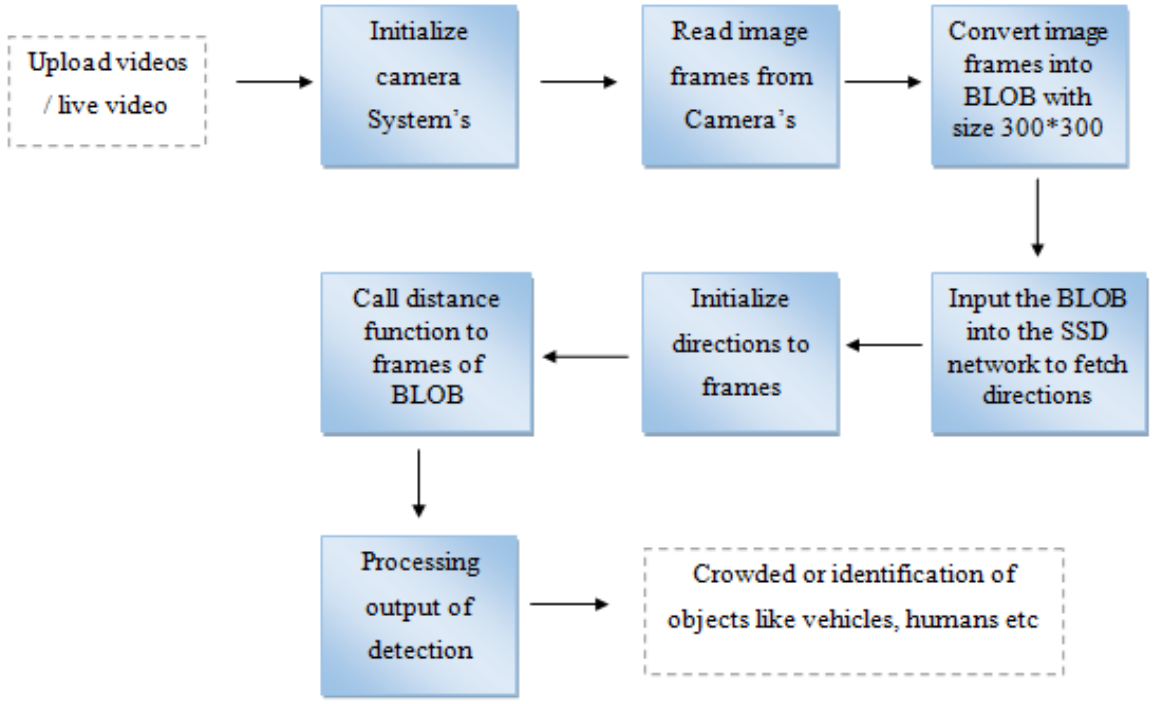


Figure-2: Architecture of Crowd Detection using SSD.

IV. RESULT AND DISCUSSION

Implementation of project

Crowd detection and identify the vehicles is done by extracting the crowd frames from videos or a webcam and then detection is done. For this detection we are using python language. The implementation of this project is done by following steps which are explained in detail below:

Step1 – Download the software and required frame work and their libraries. The latest version of pycharm 3 is installed to 64 bit operating system.

Step2 – The crowd detection is done here so for that we are using SSD. Network model so installed its framework and their libraries.

Step3 – For the purpose of identify the humans and vehicles the supporting systems are installed.

Step4 – Install numpy and its libraries for convert the frames into binary large object.

Step5 – Importing the modulus of open CV and Tensor flow for identify the vehicles and display mark on it at output window.

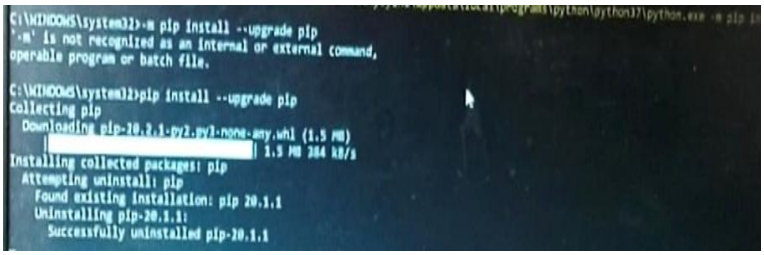


Figure-3: Nmpy libraries installed.

e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
Volume:02/Issue:08/August-2020 Impact Factor- 5.354 www.irjmets.com

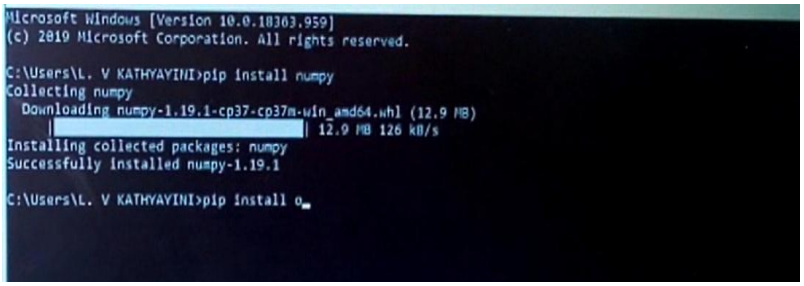


Figure-4: OpenCv libraries installed.

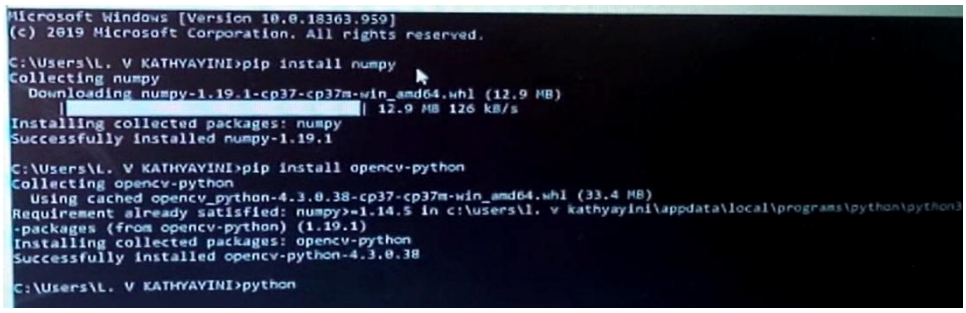


Figure-5: Tensorflow library installed.

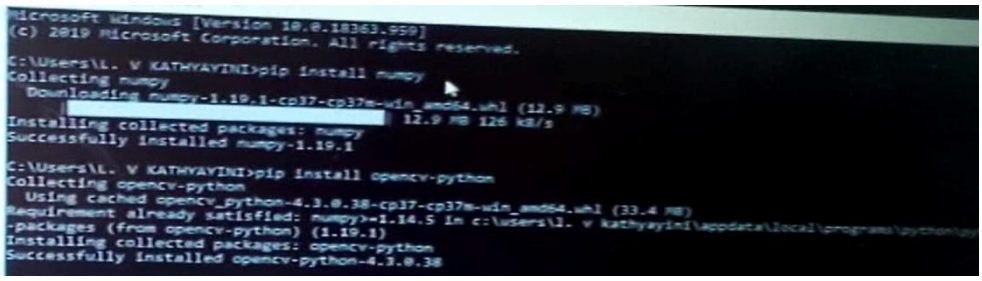


Figure-6: Install pip module.

Importing the required modules

The modules needed to recognize face and iris are cv2, numpy, tensor flow, scipy. After downloading and installation of Python 3.6.7 f or a 64 bit system start importing libraries one by one. Download cv2 by giving pip install Opencv-python and then open the downloaded python IDLE and type as:

Import cv2

Print cv2._version_

If there is no error seen then cv2 is successfully installed.

Download numpy and import it by giving this command in command prompt or cmd.

Import numpy

If no errors are found then numpy is installed. If there is any error then we can create the environment by downloading the visual studio.

Download and import tensor flow and scipy so that we can run the code.

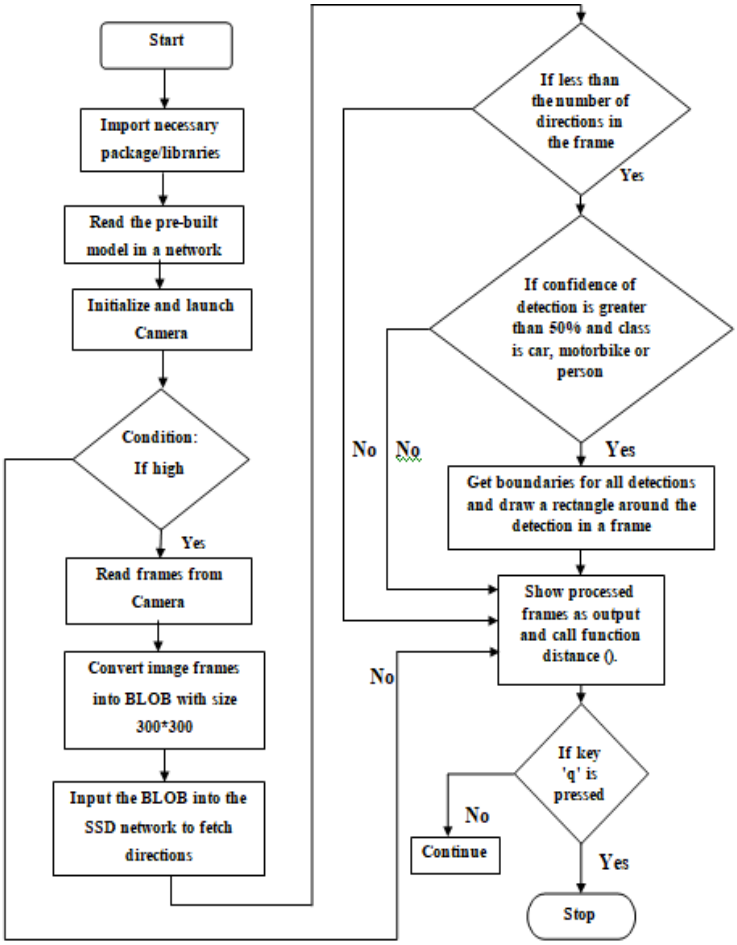
Import tensorflow

Import scipy

Cv2 is the Opencv module and consisting of the function for face detection and recognition. Opencv consists of trainer and a detector. If you want to train your own classifier for any object like car, planes etc then we can utilize Opencv to create one. Numpy arrays are used to store the image. pip install –upgrade and install SSD model. The codes given in appendixes are run using python IDLE with supporting tools and libraries to obtain the desired results.

e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
Volume:02/Issue:08/August-2020 Impact Factor- 5.354 www.irjmets.com

Flow chart for Crowd Detection

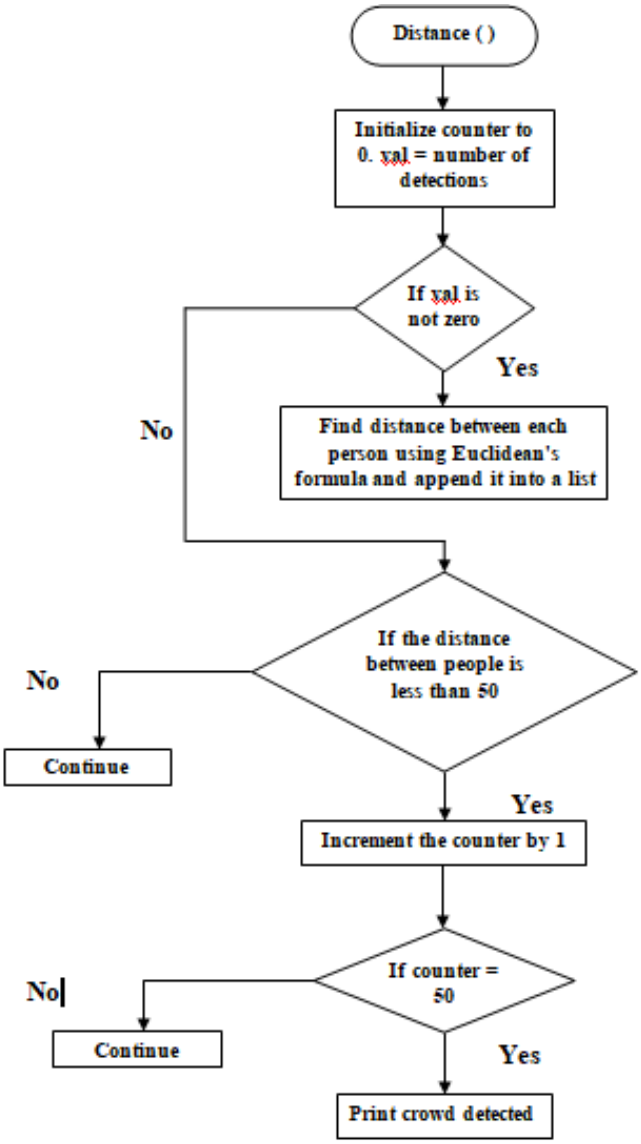


Flow Chart explaining for Detecting:-

The Flow Chart will give the idea of how the code will execute and how the detection is done. Firstly we start the python pycharm 3 software here we using python language so, then Import the Necessary Packages and Libraries to detecting , converting Frames into Binary large object, and counting, reading frames from video's or webcam we need these packages and libraries. Next we start reading pre- built model in a network called ' SSD' model, Before the Execution of Main code we have to initialize this model in a network then Initialize the camera. Applying condition loop to code if conditions are true then it start reading the Frames from camera or web camera. Next convert those Image Frames into BLOB with size 300*300. Then input of this BLOB is fed into the SSD Network Model to fetch the directions. If the number of directions are less then Frames then the two conditions is gets true then Execute further. Then it will goes to next change and check the confidence of direction is greater than 50% and class is considered as car,. Motorbike or person of it is true with condition then goes into Next step that is get boundaries for all detections and draw the Rectangle around the detection in a frame. Next processed the Frames as output and call the function called 'distance ( )'.

e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
Volume:02/Issue:08/August-2020 Impact Factor- 5.354 www.irjmets.com

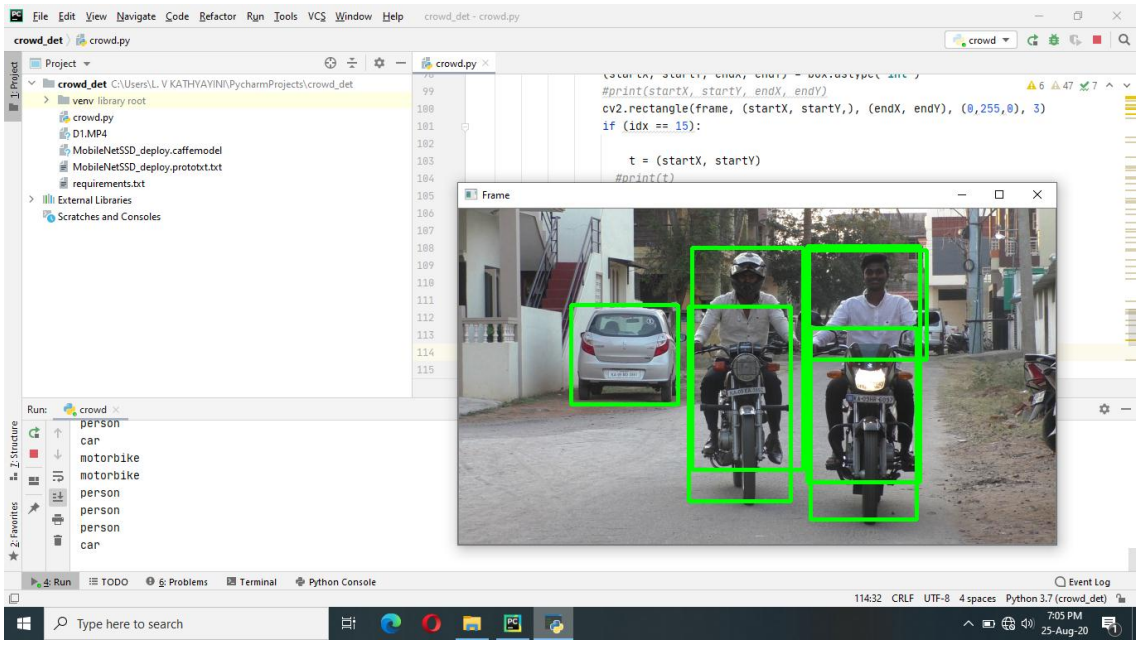
Flow chart for distance function



Flow-chart explaining for Distance Function:-

Once we detect the human, car, motorbike etc. We call this function. Next initialize counter to 'o' and the ' val' is equal to Number of Directions. If only the 'val' is not zero then condition gets true and find the distance between each person using the Euclidian's Formula and append it into a list . If the condition is not true then go to next function. That is if the distance between people is less than 50 Increment the counter by 1. It will count up to 50 after that print crowd is detected. (If counter is not 50 then 'No' and it conditions. If key 'q' is pressed then it Stop the code else continue the program. This is the simple steps which are in programming while executing).

e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
Volume:02/Issue:08/August-2020 Impact Factor- 5.354 www.irjmets.com



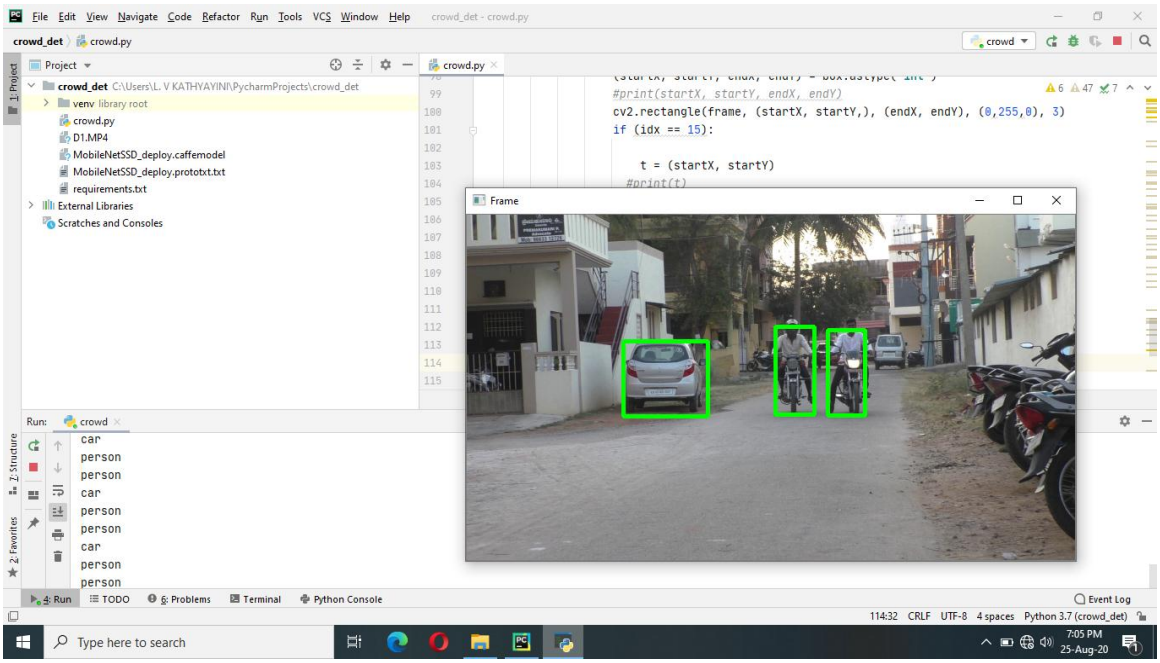


Figure-7: detecting the movements of persons and vehicles.

V. CONCLUSION

In this project Abnormal crowd detection is done using the SSD Neural Network Model. The SSD is the main key tool which is implemented in our code to detecting the objects, Humans, car, Motorbike etc .It is very useful Model with good outputs to its inputs Frames, which makes the detecting process smooth and efficient. The python is s Machine learning language which is fast in processing programs, like speed of Execution, size of memory usage etc. It includes several libraries and packages to convert the Frames easily and read the Frames also very quickly. The surveillance camera systems are more needed in the present System's because of security we have to monitor it continuously. So it is very useful tool to extract the features from cameras as well as been can and risky also, because we need an intelligent and accurate supporting tool for the further identification, recognizing, detecting and analyzing the videos. This project

is a very good approach and very helpful in all the fields also and especially in Government sectors like Bank's, office's, ATM,s, center's, Hospitals and Specially for criminals or Crime Activities are happened areas. To identify the thefts and Criminals we need this Surveillance System. It is good subject in further Researching as well as improving the surrounding fields also, because security is the primary thing which need from everywhere and for everyone it is useful in all the fields. To get high accuracy results and perfect picture need effective systems. Identification also use the good resolution cameras, and also improve a good techniques in Detecting field also because we have good supporting tool means it's half work is clear. Establish and Invent Different Algorithms and Architecture's are implemented in future in this area and get good accurate well defined results.

VI. REFERENCE

[1] K. Buys, C. Cagniart, A. Baksheev, T.-D. Laet, J. D. Schutter and C.Pantofaru, "An 0adaptable system for RGB-D based human body detectionand pose estimation," Journal of visual communication and imagerepresentation, vol. 25, pp. 39-52, Jan 2014.

[2] A. Jalal and S. Kamal, "Improved Behavior Monitoring and ClassificationUsing Cues Parameters Extraction from Camera Array Images," IJMI, 2-18.

[3] A. Jalal, Y.-H. Kim, Y.-J. Kim, S. Kamal and D. Kim, "Robust humanactivity recognition from depth video using spatiotemporal multi-fusedfeature," Pattern recognition, vol. 61, pp. 295-308, 2017.

[4] B. Enyedi, L. Konyha and K. Fazekas, "Threshold procedures and image segmentation," in proc. of the IEEE International symposium ELMAR, pp. 119-124, 2005.

[5] F. Farooq, A. Jalal, L. Zheng, "Facial expression recognition using hybrid features and self-organizing maps", in proc. of ICME, pp. 409-414, July2017.

[6] X. Yang and Y. Tian, "Eigenjoints-based action recognition using naïve bayes-neartest-neighbor," in proc. of the CVPR, pp.14-19, June 2012.

[7] A. Jalal, and S. Kamal, "Real-time life logging via a depth silhouette based human activity recognition system for smart home services," in Proceedings of AVSS, Korea, pp. 74-80, Aug 2014.

[8] A. Jalal, M. Maria and A. Hasan, "Multi-features descriptors for humana ctivity tracking and recognition in Indoor-outdoor environments," in proc. on IBCAST, 2019.

[9] A. Jalal and S. Kim, "The mechanism of edge detection using the block matching criteria for the motion estimation," in Proceedings of HCI Conference, Korea, pp. 484-489, Jan 2005.

[10] A. Sony, K. Ajith, K. Thomas, T. Thomas, and P. L. Deepa, "Video summarization by clustering using euclidean distance," in proc. of the SCCNT, 2011.

[11] S. Kamal and A. Jalal, "A hybrid feature extraction approach for human detection, tracking and activity recognition using depth sensors," Arabian Journal for science and engineering, 2016.

[12] T. H. Chen, T.-Y. Chen and Z.-X. Chen, "A Intelligent People-Flow Counting Method for Passing Through a Gate", in proc. of the CIS, pp.573-578, 2006.

[13] J.-W. Kim, B.-D. Choi and S.-J. Ko, "Real-time vision-based people counting system for the security door", in proc. of the Inter. conf. on CSCC, pp. 1-4, 2002.

[14] V. Rabaud and S. Belongie, "Counting crowded moving objects," in proc. of the conf. on CVPR, pp. 705 -711, 2006.

[15] G. Antonini and J.P Thiran, "Counting pedestrains in video sequences using trajectory clustering," IEEE Trans. on circuits and systems for video technology, vol. 16(8), 2006.

[16] T.-H. Chang and S. Gand, "Tracking multiple people with a multi-camera system," in proc. of the IEEE workshop on multi-object.