International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 08 Issue: 04 | Apr 2021 www.irjet.net p-ISSN: 2395-0072

Detecting and Tracking Objects in a Live Video Stream

[1]***Vedant Praful Zilpilwar***, *Department of Computer Engineering, Pimpri Chinchwad College of Engineering(PCCOE), Pune, India*
[2]***Amey Ajay Athavia***, *Department of Computer Engineering, Pimpri Chinchwad College of Engineering(PCCOE), Pune, India*
[3]***Akash Prithviraj Deshmukh***, *Department of Computer Engineering, Pimpri Chinchwad College of Engineering(PCCOE), Pune, India*
[4]***Saurabh Balkrishna Padman***, *Department of Computer Engineering, Pimpri Chinchwad College of Engineering(PCCOE), Pune, India*
[5]***Prof. Shailaja Pede***, *Department of Computer Engineering, Pimpri Chinchwad College of Engineering(PCCOE), Pune, India*

Abstract - *Machine learning, the future technology is destined to provide ease of living with its applications spread in all aspects of our day-to-day life. A part of machine learning is Image Processing, another growing field in which images and videos are used to provide information. Use of Image Processing is widespread and one of its applications is for security purposes. Image Processing not only reduces the human interference in video surveillance but also provides a more accurate system. This paper is prepared to present the survey of various algorithms and models used in carrying out the task of detection and tracking an object in a public area. Here, we have compared two methods of detection and presented the superior one which can be used for real world implementation. Some of the Object tracking methods have been carefully analyzed and the efficiency of each has also been discussed. And the implementation is done using an FPGA module.*

Key Words: Background subtraction, CNN, SSD, Tracking algorithms, Detection, TO-MHT, Point tracking, FPGA

I. INTRODUCTION

- Comparison has been done between background subtraction and Convolutional neural network-based models in the detection process of objects in a live video stream. Various Object tracking methods have also been reviewed. Tracking methods are also presented that will help in tracking the detected object across the frames of the video. Various hardware components are also identified which can be used for implementing for real world applications.

- *Motivation:*
  There are many instances in security surveillance where manual security checks can be cumbersome and require a lot of man-power as well as funds to support that man-power. At such times the advancement of technology is a big advantage. Now we can let the computer and software do all the hard work. Keeping this in mind, we decided to poke around in this field and came around the topic of Computer Vision and detection of objects using cameras and machine learning models.

Objective:

- Comparing different object detection and tracking models.

- To provide a system that can detect and track an object across frames of a video stream.
- Present a reliable and cost-efficient hardware to implement the model for real world applications.
- To reduce human effort in the security surveillance system and provide better accuracy.

II. LITRATURE SURVEY

A: Background Subtraction

Background Subtraction, a method of object detection that involves two initial steps. The first is to capture a still image of the background which will act as the default image which we will refer for detection of any foreign body that will appear on it. The next is to actually capture an image of the same background but with objects in front of it. What this algorithm does is, it first compares the two captured images and substacts them, which results in the detection of the object present in the second image. Hence the name background subtraction where it subtracts the background from the image itself. As it is clear from

this, if there is any change in lighting or ambient lighting of the background after taking the first image, there will be a lot of noise during the detection phase and then this is one of the major drawbacks of this model, that is a slight change in lighting or ambient temperature or any other factors massively affect the accuracy percentage of this method.

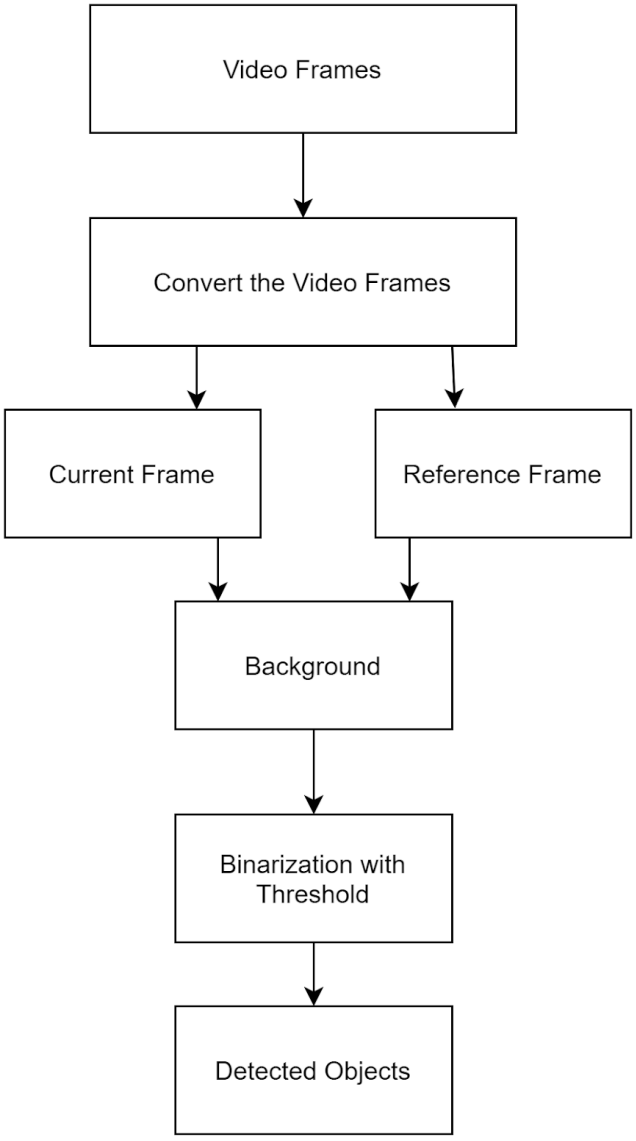


Fig 1: Background subtraction

Initially,we take the video as input as input from the camera, as shown in Fig. 1 [3], where we implement background subtraction for detection of an object. Since objects are tracked in real time, frame by frame video input is taken directly from the camera. Initially the camera sets the reference frame for the improved background subtraction process. The subtraction is carried out between each frame's pixels. Grayscale conversion is performed on later reference frames and current frames. Morphological operations have been used to eliminate the noise material from all images independently, as well as smoothing. And, pixel by pixel, two frames subtracted each other. If a variation is discovered, the object's existence in the current frame is verified. Foreground images are pixels that have a difference greater than the given threshold. The reference background image is dynamically transformed over time to scene content after applying post processing methods such as dilation, closure, and erosion to reduce the noise content in the foreground image..

B. CNN based model

CNN based models are widely used in detection of objects in images and videos for surveillance and security purposes. Surveillance is usually carried out in two types of environments- controlled (indoor) environments and uncontrolled(outdoor) environments. CNN has many object detection API which help in this process. Few of them are Fast-RCNN, Faster-RCNN, YOLO (You only look once) SSD (Single Shot Detector) [9]. Among these, depending on speed of detection, accuracy, the most suitable one is chosen. For real world implementation in public places SSD and YOLO are mostly used as they provide faster and accurate detection of objects [9]. SSD is more superior than YOLO as the speed of detection is comparatively greater than that of YOLO. Due to this factor we have also chosen SSD.

C. Object Tracking

Video Surveillance has been going through massive changes and consistent research has led to various techniques for the same. As object tracking is a fundamental part of surveillance systems it is necessary to study those methods, techniques and algorithms. In Object tracking what we do is generate a path and direction of an object in a 2-D plane and locate the coordinates of the object in the following frames. Object tracking has been divided into 3 parts viz. Point tracking, Silhouette based tracking & Kernel based object Tracking. Simply put all these methods and algorithms have their own limitations and features to be exploited further.

III. MODEL IMPLEMENTATION

- Background Subtraction
  - For implementing background subtractions, the first step is to capture a still background which is used as a frame of reference.
  - Next, when any object appears in front of that frame of reference the two frames, ie. the new frame and the first frame which consists only of the background, are subtracted from each other.
  - This results in the detection of the object in the frame of reference .

- CNN
  - From all the CNN based API we have chosen the SSD (single shot detector) as its accuracy and speed is well suited for application in the real world.
  - Vision based systems are mostly developed for security purposes, driverless cars, autonomous drones which require fast processing of the visual stream and provide quick output using video processing technologies [9].
  - CNN based models are very robust and they can adapt to the changing background with easy [9]. The input visual stream is passed through the network of CNN's model, which has been trained in images which results in generation of values which are later used to detect objects from other video streams or images. The video is given to the network by taking a frame from that video and detecting the objects from that frame and providing the required information about it.
  - Once the object is detected, a box is created around it with the tag and accuracy percentage. This box follows the object until it is in the video frame. This architecture of CNN has block boxes which offer limited tractability.
  - The softwares and models that we have used are the CNN based SSD mobilenet which is multi-box detector ie. it has the capabilities to detect multiple objects in a single video input as well as image input. Along with this, the use of tensorflow and OpenCV help to preprocess and train the models. Using images and tags we can create a model that is encoded to recognize those particular tags which is beneficial when the tags are less and thus it will require less processing power while training.

- **Tracking technology**

  - Point Tracking:

The object to be tracked is considered as a point and then to further detect those points in contiguous frames point tracking can be utilized [10].

The point tracking methods further has various algorithms such as Kalman Filtering, Particle filtering being Single object tracking algorithms, whereas, Joint probability Data Association filter (JPDAF) and Track-Oriented multiple hypothesis tracking (TO-MHT) algorithms are multiple object tracking algorithms.

  - Kalman Filter:Algorithm:

Kalman Filtering Algorithm

No. Of Object tracked: Single

Occlusion handling: No

Accuracy: Moderate

  - MHT (Multiple Hypothesis Tracking):

Algorithm: MHT algorithm

No. Of Object tracked: Multiple

Occlusion handling: Yes

Accuracy: Low to Moderate

  - Particle Filter:

Algorithm: Improved particle filtering algorithm

No. Of Object tracked: Single

Occlusion handling: Yes

Accuracy: High

Among these technologies, MHT technique being able to track multiple objects is the one of best tracking methods. The algorithm being used is named

as TO-MHT (Track-Oriented Multiple hypothesis Tracking) algorithm.

TO-MHT receives the measurements and determines the relation between required measurements and tracks. MHT also helps in multiple target tracking in a cluttered video frame by merging the process of initiating and deleting tracks. At every moment the track being generated follows the bottom-up approach of MHT and improves the global hypothesis. Fig. 2 will show the design of TO- MHT algorithm [12].

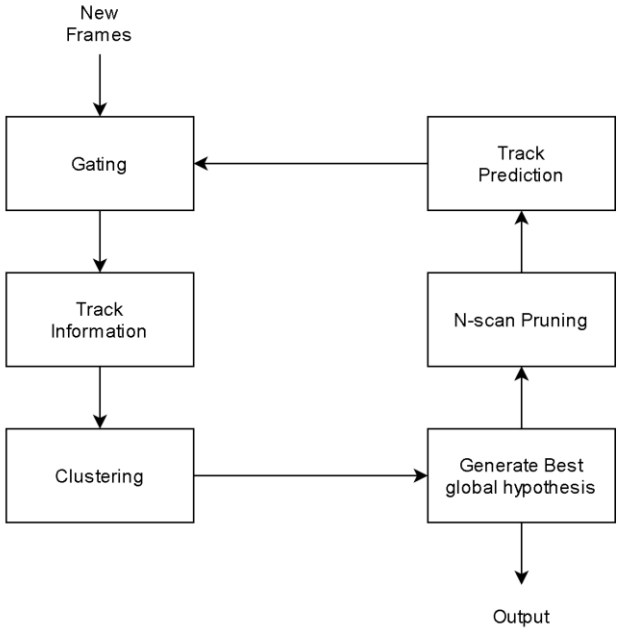


Fig.2: Design of TO-MHT algorithm

This Algorithm performs better with frames having various occlusions but with higher complexity. However, to Lessen the Computation there are various strategies among those Track confirmation and Deletion is easy to understand thus can be readily used.

- Confirm and delete track:

In this strategy we implement SPRT or sequential hypothesis test in Track Confirmation and Deletion which uses the track scores to compare the lower and upper thresholds. These comparisons are then used to build the following logic

1. Track Confirmation

2. Continue the Verification of tracks for future measurement.

3. Track Deletion

Further we use track score to check these tracks and as per sequential probability ratio test (SPRT) by comparing the upper and lower threshold [12]. The thresholds are defined as follows

$T_2 = \ln 1-, T_1 = \ln 1-$ ...(1)

Where α and β stands for probability of false tracks confirmed and true tracks deleted respectively.

- **Hardware technology**

- FPGA MODULE

The system uses Zynq XC7Z020 FPGA board for the video surveillance using modified background subtraction. This board consists of the OV7670 camera which has a resolution of 0.3 megapixel and is operated at speed of 30 frames per second. The camera control module controls the camera with the help of I2C(I squared C) communication and some other register set values, these values determine the pixel and resolution data representation. like Y'UV, RGB(Red Green Blue) etc. COM 7 register of the OVA camera module is programmed to directly retrieve the data in the YUV(Y-luma component, U-blue projection, V-red projection) format which uses 24bits(as 8bits each for Y, U and V) to show each and every pixel of an image.[6]

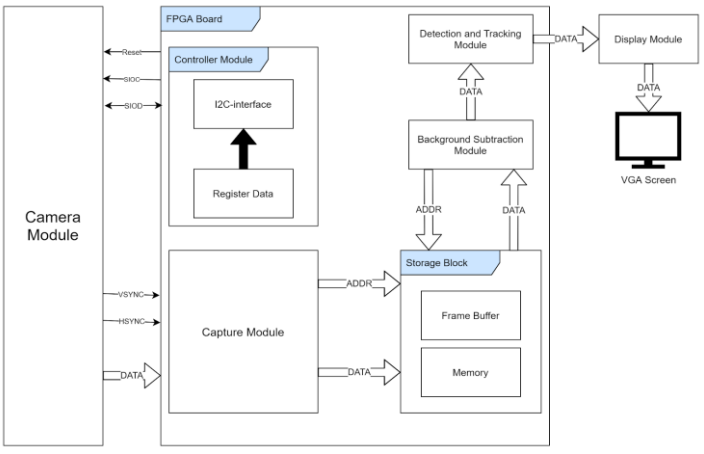


Fig.3. Block diagram of object tracking in real time

The module is divided into sub-module

- CAMERA CAPTURE MODULE

As the name suggests, it is used for capturing the video from the camera connected with the Field Programmable Gate Arrays(FPGA) board. Here OVA camera is used, this camera operates at the speed of 30 frames per second and operates at a speed of 50Mz.

- STORAGE MODULE

In the storage module, the data sent by the capture module for each and every pixel is received which is 8 bits each pixel. the background memory block stores the information of pixels of the background frame and the frame buffer block stores the data of the frames received from the video, these both data information are needed.In total there are 640*480pixels in each frame, as each pixel is of 8 bits therefore the storage size of each frame comes out to be 640*480*8bits that's equal to 307.2kb of data storage

- DISPLAY MODULE

The display module connects the display screen with the FPGA board. In this the video frame which is captured by the capture module(OVA camera) is processed using various methods like detection and tracking. Then this data is sent to the VGA module for display screen

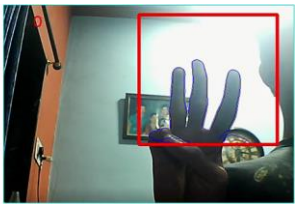
- COMMUNICATION MODULE

In the communication module, it establishes the communication between FPGA board and the camera, this module uses I2C (I Squared C) standard which is a synchronous, mult-imaster, multi-salve, serial computer bus. In the technical area it is often used for connecting low speed peripheral ICs to microcontrollers and processors.

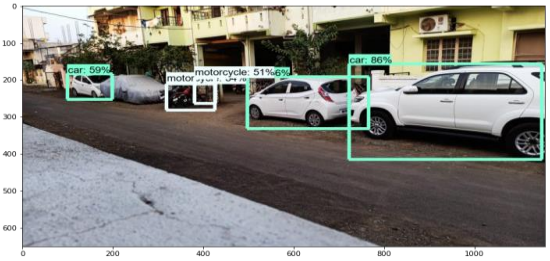
IV. RESULTS

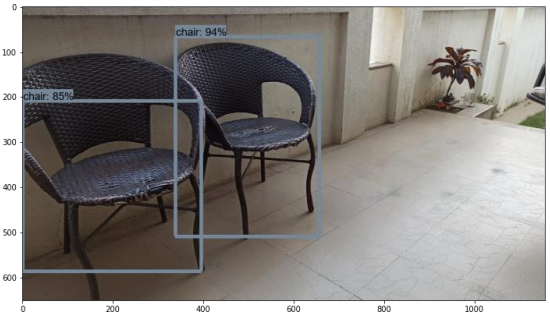
For detecting objects in a video stream we have used 2 methods. From the implementation we have come to certain conclusions that will help in selecting the suitable model based on our requirements.

- Background Subtraction
  - During implementing this method the most important past was to capture a still frame of reference.
  - The drawback of this method is, even a small change in ambient lighting or surroundings will affect the accuracy of detection.
  - This algorithm works at 15-26 fps.



- Single Shot Detector
  - This method is best when there is momentary change in ambient lighting and surrounding.
  - The MultiShot SSD model is able to detect and track multiple objects in one video output.
  - The output of this model is 96% accuracy at 25fps.





	Background Subtraction	SSD
FPS	15-25	~25
Accuracy	70-85%	85-96%

V. APPLICATION

1. Video Surveillance system
2. Weapon detection
3. Face detection & recognition
4. Human activity recognition
5. Detection of unattended and stolen objects
6. Path Finding For vehicles in traffic

VI. CONCLUSION

To conclude, we can say that the SSD MobileNet model is more suited for environments where there is an ever changing surrounding and ambient lighting. It has a better accuracy percent at higher frames per second. On the other hand background subtraction can be chosen when the camera is still or when there is no change in surrounding.

VII. FUTURE SCOPE

Facial recognition can be added to detect the person in the frame for security reasons. This will help the authorities to provide better security in public places. Implementation of infrared or thermal cameras that check the detected object for any illegal objects in places like airports or railway stations. Detection of objects like firearms, knives in security-sensitive areas can be more useful for better security purposes.

VIII. REFERENCES

[1] Lakhan H. Jadhav, Bashir Ahmed F. Momin, "Detection and Identification of Unattended/Removed Objects in Video Surveillance", IEEE International Conference on Recent Trends in Electronics Information Communication Technology, May 20-21, 2016, India

[2] Jaya S. Kulchandani, Kruti J. Dangarwala, "Moving Object Detection: Review of Recent Research Trends", IEEE 2015 International Conference on Pervasive Computing (ICPC) - Pune, India (2015)

[3] Mohana, H. V. Ravish Aradhya," Elegant and Efficient Algorithms for Real Time Object Detection, Counting and Classification for Video Surveillance Applications from single fixed camera", IEEE 2016 International Conference on Circuits, Controls, Communications and Computing (I4C) - Bangalore, India (2016)

[4] Guillermo Botela, Carlos Gracia and uwe Meyer-Base, "Hardware implementation of machine vision system: image and video processing", EURASIP journal on Advances in Signal Processing, 2019

[5] Ahinus H, Gopalakrishna my M E, "A New Object Detection and Tracking Using FPGA", IOSR Journal of Electronics & Communication Engineering (2017)

[6] Shreekant Sajjanar, Suraj K Mankani, Prasad R Dongrekar, Naman S.Kumar, Mohana, H. V. Ravish Aradhya," Implementation of Real Time Moving Object Detection and Tracking on FPGA for Video Surveillance Applications Department of Electronics & Communication Engineering R.V. College of Engineering India (2016)

[7] Kang Hao Cheong (Member, IEEE), Sandra Poeschmann, Joel Weijia Lai, Jin Ming Koh, U. Rajendra Acharya, Simon Ching Man Yu and Kenneth Jian Wei Tang "Practical Automated Video Analytics for Crowd Monitoring and Counting" Digital Object Identifier 10.1109/ACCESS.2017.DOI

[8] Prerna Dewan, Rakesh Kumar, "Detection of Object in Motion Using Improvised Background Subtraction Algorithm" International Conference on Trends in Electronics and Informatics ICEI 2017

[9] Bozhao Qi, Wei Zhao, Haiping Zhang, Zhihong Jin, Xiaohan Wang, Troy Runge, "Automated Traffic Volume Analytics at Road Intersections Using Computer Vision Techniques" The 5th International Conference on Transportation Information and Safety, July14 – July 17, 2019, Liverpool, UK

[10] Rakesh Chandra Joshi, Mayank Joshi, Adithya Gaurav Singh, Sanjay Mathur, "Object Detection, Classification and Tracking Methods for Video Surveillance: A Review", IEEE 2018 4th International

Conference on Computing Communication and Automation (ICCCA)

[11] Akshay Mangawati, Mohana, Mohammed Leesan and H. V. Ravish Aradhya, "Object Tracking Algorithms for Video Surveillance Applications", IEEE 2018 International Conference on Communication and Signal Processing, April 3-5, 2018, India

[12] ShengSen Pan, Qinglong Bao, Zengping Chen, "An Efficient TO-MHT Algorithm for Multi-Target Tracking in Cluttered Environment", 978-1-4673-8979-2/17/$31.00 ©2017 IEEE

[13] B. Maga, Mr. K. Jayasakthi Velmurgan, "AN EFFICIENT APPROACH FOR OBJECT DETECTION AND TRACKING", IEEE 2017 Third International Conference on Science Technology Engineering & Management (ICONSTEM)