International Journal of Early Childhood Special Education (INT-JECSE)
DOI:10.9756/INTJECSE/V14I5.79 ISSN: 1308-5581 Vol 14, Issue 05 2022

OBJECT DETECTION,TRACKING AND BEHAVIOURAL ANALYSIS FOR STATIC AND MOVING BACKGROUND

Dr Reeja S R
Professor- SCOPE,
Vellore Institute of Technology- Andhra Pradesh(VIT-AP)
reeja,sr@vitap.ac.in

Mr. Ishfaq yaseen
Professor,
shah.ishfaq@gmail.com
Sattam Bin Abdul Aziz University,Saudi Arabia

***Abstract*—This Detection of vehicles and tracking gives more promising way to manage and control the road traffic using the traffic surveillance system. In military and civilian application, vehicle detection and tracking applications play an important role. A challenging task of moving object segmentation in complex environment is more in Vehicle surveillance. As number of vehicles are increasing rapidly, controlling the traffic is a huge problem. Vehicle detection and tracking, calculating average speed of each individual vehicle, counting of vehicle in current frame can be done by vehicle detection process. Traffic analysis and vehicle categorizing objectives and may be implemented under different environments changes can be done on road using detection process. All this help us to handle the traffic in efficient way possible. We have detected and tracked the vehicle and also calculated behavior analysis of the vehicle whether it is moving towards forward or left or right. Also, the number of vehicles in frame is counted and also the static and dynamic mobility of the vehicle is analyzed.**

Keywords*— *object tracking, YOLO, static background object detection, and moving background.

I. INTRODUCTION

Object detection and tracking is challenging task which is important within the area of Computer Vision that try to detect, recognize and track objects over a sequence of images called video. It helps to understand, describe object behaviour instead of monitoring computer by human operators or intervention. It aims to locating moving objects in a video file or surveillance camera as frames. Object tracking is the process of locating an object or multiple objects using a single camera, multiple cameras or given video file which contains various object. The challenging factors are due to Invention of high quality of the imaging sensor, quality of the image and resolution of the image are improved, and the exponential increment in computation power is required to be created of new good algorithm and its application using object tracking in order to get better result. In Object Detection and Tracking we have to detect the target object and track that object in consecutive frames of a video file and show the counting in current frame.

II. LITERATURE SURVEY

[1] Lalita Gavit et.al demonstrated. "Object Tracking Using Multiple Cameras". This paper mainly highlights an algorithm where the usage of a single camera may not be proficient in covering huge spaces. Therefore, the concept of multiple cameras came into the picture which is usually placed in various parts of the spaces that are to be covered. These multiple cameras are positioned with an overlapping region between the field of video commonly called as FOV of different cameras thereby each camera captures the video of the field of video (FOV). By establishing communication among objects captured in multiple cameras the system is capable of tracking the people in multiple perspective imagery.

DOI:10.9756/INTJECSE/V14I5.79 ISSN: 1308-5581 Vol 14, Issue 05 2022

One of the methodologies used to track the object is Block Method Analysis that is dependent upon the principle of prediction. A search window for every object in the frame is collected through which we can find the trajectory of the object where every frame gives a track of the object. Labelling is being provided for each and every object in the frame for tracking multiple objects. The major problem that can be caused by using multiple cameras to capture images is that one object might overlap with the other in a particular frame this problem is stated as occlusion where one object is hidden behind another object in the FOV of any one of the cameras. Hence there is a need for resolving the problem of occlusion and removing it so that tracking of hidden objects might not be stopped. Finally, the method proposed in the paper plays a significant role in a surveillance system that can be a major key for security purposes in places and areas like parking, hotels.

[2] Dr J.L Mazher Iqbal et.al 2017 demonstrated “Object detection and tracking using Thermal camera”, Volume 114 No. 10. This paper mainly focuses on Object detection and tracking through the use of Thermal cameras and uses foreground detection, Kalman Filter algorithm for tracking. As we know Tracking Moving Objects is not easy in A Real-Time Environment, due to continuous change in the location Of Objects during movement. This Proposed System Can Achieve Good Result for Object Detection and Tracking through the use of Foreground detection and Kalman Filter algorithm for detection even in dark places using Thermal camera. The key aim of visual surveillance is to reduce human effort by developing computer vision algorithms. By using the normal camera, it is difficult to track in night vision. The proposed system uses the foreground detection method to detect the object and Kalman filter to track the object in real-time using a thermal camera.

One of the significant responsibilities in the field of computer vision and image processing is chiefly Foreground detection whose principal plan is to recognize variations in image sequences and it further divides the foreground from that of the background which is primarily based on these transformations taking place in the foreground. The Algorithm which is widely used to track moving objects, allows us to estimate the velocity along with the acceleration of an object and measurements of its location is Kalman Filter Algorithm. The steps used in the processing of this algorithm are that it takes video input, foreground method to detect the objects through the use of thermal cameras to capture the images for analysing, because it can capture the images and make processing even in dark places, and at unconditional weather conditions.

[3] Mukesh Tiwari et.al demonstrated “Improved Algorithm for Object Tracking in Video Camera Network”. This paper aims to develop a completely unique approach for tracking the thing in multiple camera environments. The situation gets worse if the object hid beside any other object, or if the object moves from the range of the camera span. The optical flow estimation technique is being used for tracking the object where the magnitude with the gradient direction is recognized. The threshold magnitude comparison of the pictures detects the specified target. In this work, background detection is adopted due to its better ease of implementation than other motion detection systems. The background subtraction approach is useful during a fixed camera arrangement system thanks to the easily available static background information.

In addition to that, it produces sufficient information of the moving blobs for its fast processing that produces more easy information than corners and edges. Such an approach requires less time and memory complexities than the other existing method for moving object segmentation. In this method, the work is aimed to use background frames for object detection by accompanying the background updating scheme. Detection parameters within the proposed methods like SIMILARITY, F1, PRECISION, RECALL are better as compared with the prevailing methods in viz RADCT, MSDE, SDE, and SSD. The proposed algorithm can easily detect an object even if it is either hidden or it moves to the frame of camera 2, or any other camera installed within the same camera network. The proposed methods tracked and detected the suspected object target for the appliance of safety and security system more accurately and with better simple implementation than other existing motion detection systems such as RADCT, MSDE, SDE, and SSD.

DOI:10.9756/INTJECSE/V14I5.79 ISSN: 1308-5581 Vol 14, Issue 05 2022

[4] Sayanan Sivaraman et.al 2013 demonstrated "Looking at Vehicles on the Road: A Survey of Vision-Based Vehicle Detection, Tracking, and Behaviour Analysis". The paper mainly focuses on vehicle detection, tracking, and behaviour analysis which is road vision-based. There is a lot of advancement made in the field of vision-based analysis when compared to their initial advancements. This paper explains on latest advancements in vehicle detection, stereo vision, active sensor-vision, and also handling on monocular vision. The main concept involved in the paper is vehicle-based tracking of the vehicles through the use of domains such as stereo, monocular vision as well as estimation, models include dynamical model and finally analysing filtering. As we know that behavioural analysis of tracking vehicles is a most emerging research area in recent times. The prerequisite for analysing the on-road behaviour of other vehicles include robust vehicle tracking and its detection. Therefore this paper places vision-based vehicle detection and tracking which is based on sensor on-road context and also the comparisons with the techniques particularly radar and lidar technology. [6] shows the review paper about multicamera object detection.

III. ALGORITHM DESCRIPTION

The Neural networks are used for detection and in Neural Network there are various algorithms like for object detection like the CNN and Fast R-CNN. This algorithm uses a method called Selective Search but there are huge computations which can be reduced using more advanced algorithm. Like YOLO which stands for you look only once, They scan the image only once and predict the Bounding Boxes so they are much faster compare to the other methods.YOLO V3 is an improvement version of YOLO detection networks, it features multi-scale detection, stronger feature extractor network, and just like other single-shot detectors, YOLO V3 also runs quite fast and makes real-time inference possible on GPU devices.They are more accurate and fast that YOLO V2.YOLO uses convolution neural network for object detection, it can detect multiple objects on a single image.

Apart from predicting classes of object, YOLO also detects locations of this object on the image.YOLO divides image into regions and probabilities for every region.YOLO original consists of 53 convolutional layers that are also called Darknet-53 but for detection, original architecture stacked with 53 more layers that gives us 106 layers of architecture. Each convolutional layer followed by batch normalization layer and ReLU activation function.There are no pooling layers but additional convolutional layers with stride 2 are used to down sample feature maps prevents loss of low-level features.capturing low-level features helped to improve ability for detection of small objects. YOLO version 3 makes detections at three different scales and at three separate places in the Network and these are layers 82, 94 and 106.These layers number are called stride of the network and that they show how the output at three separate places in the Network is smaller than input given to the Network.

For instance, if we consider stride 32 and input network size 416 by 416, then it will give us the output of size 13 by 13. Consequently, for the stride 16 the output will be 26 by 26 and for the stride 8 the output will be 52 by 52. 13 by 13 is responsible for detecting large objects,26 by 26is responsible for detecting medium objects and 52 by 52 is responsible for detecting small objects.The next topic to focus on is detection kernels.

To produce output YOLO version 3 applies 1 by 1 detection kernels at these three separate places in the Network. 1 by 1 convolutions applied to down sampled input images: 13 by 13, 26 by 26 and 52 by 52. Resulted feature maps will have the identical spatial dimensions. The shape of detection kernel also has its depth that is calculated by following equation. "b" here represents number of bounding boxes that each cell of the produced feature map can predict. YOLO version 3 predicts 3 bounding boxes for every cell of these feature maps. That is why, "b" is equal to 3. Each bounding box has 5 + c attributes that describe following center coordinates of bounding box, width and height that are dimensions of bounding box, objectless score, and list of confidences for every class to which this bounding box might belong to. Each feature map produced by detection kernels at three separate places within in the Network, has

DOI:10.9756/INTJECSE/V14I5.79 ISSN: 1308-5581 Vol 14, Issue 05 2022

one more dimension depth that incorporates 255 attributes of bounding boxes for COCO dataset. And the shapes of these feature maps are as following: 13 by 13 by 255, 26 by 26 by 255 and 52 by 52 by 255.

Let's move now to the next topic of grid cells. We already know that YOLO version 3 predicts 3 bounding boxes for every cell of the feature map. Each cell, in turn, will predicts an object through one of its bounding box if the centre of the object belongs to the field of this cell. And this is the task of YOLO version 3 while training. Identify this cell that falls into the centre of the object. Again, this is one of the feature map's cell produced by detection kernels that we discussed before. When YOLO version 3 is training, it has one ground truth bounding box that is responsible for detecting one object.

Firstly, we need to define which cells this bounding box belongs to and to do that we need to consider first detection scale by using 32 as stride of the Network. The input image of 416 by 416 is down sampled into 13 by 13 grid of cells as we calculated few moments ago. This grid now represents produced output feature map. When all cells, that ground truth bounding box belongs to, are identified, the centre cell is assigned by YOLO version 3 to be responsible for predicting this object. And objectness score for this cell is equal to 1. To predict bounding boxes YOLO version 3 uses pre-defined default bounding boxes that are called anchors or priors. These anchors can later be used to calculate predicted bounding box's real width and height. In total, 9 anchor boxes are used. Three anchor boxes for each scale. The feature map can predict 3 bounding boxes by using 3 anchors for every grid cell at each scale. To calculate these anchors, k-means clustering is applied in YOLO version 3.

Let's consider graphical example on how one of the 3 anchor boxes is chosen to calculate later real width and real height of predicted bounding box. We have input image of shape 416 by 416 by 3. Image goes through YOLO version 3 deep CNN architecture till the first separate place that we discussed earlier and that has stride 32. Input image is down sampled by this factor to the dimension 13 by 13 and 255 depth of the feature map produced by detection kernels as we calculated earlier.

Since we have 3 anchor boxes, each cell then encodes information about 3 predicted bounding boxes. Each predicted bounding box has following attributes: centre coordinates, predicted width and predicted height, objectness score and list of confidences for every class this bounding box might belong to.

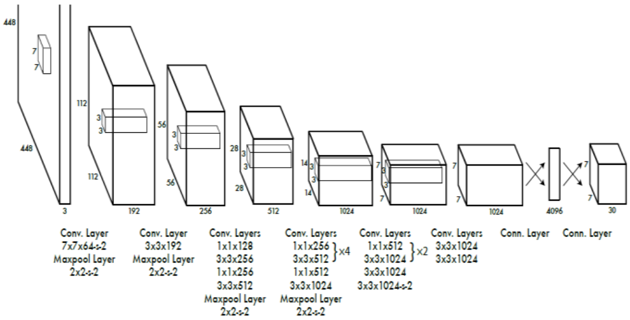

Figure 1. YOLO Architecture Diagram

IV. SYSTEM DESIGN

The below module provides various aspects of design that includes flow diagrams, static and movement background that details each and every aspect of the process. Activity diagram provides how the actual flow of the implementation happens.

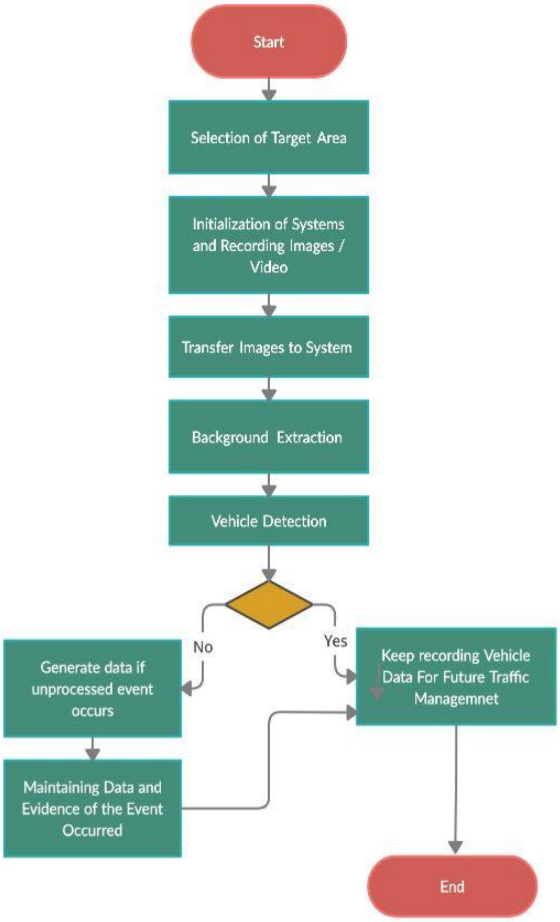

Figure 2. Activity Diagram

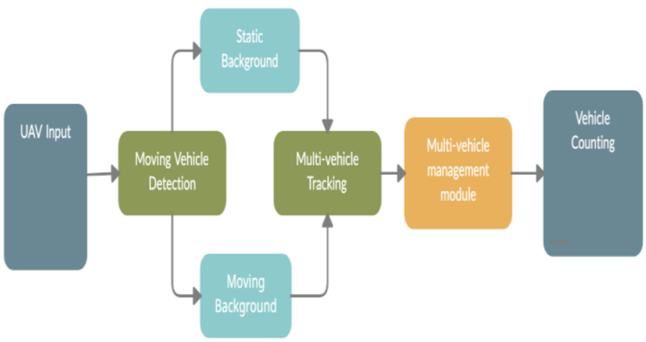

Figure 3. Design

The model proposed consists of vehicle detection, multiple vehicle management, multiple vehicle tracking, and vehicle counting. A visual sensor will be integrated within a UAV or camera. Detection of vehicles is carried out through the following situations that include static background and moving background and later tracking modules track the vehicles that are detected and finally analyze the tracker results and provide the count of the number of vehicles.

Vehicle Detection and Tracking:

Input:

The input provided possibly a video or real-time camera video for detection and tracking.

Fps:

The frames per second for reading the video are of an average of 30fps.

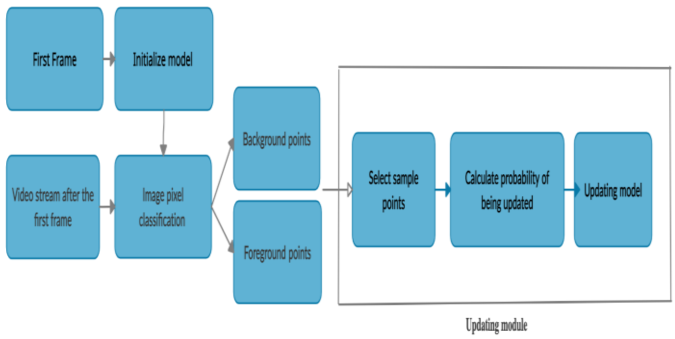

Figure 4. Static Background

Vehicle detection is a necessary process for vehicle counting.

Here we explain how does UAV works in hovering mode. While in the fixed background, background modeling is used to extract the moving vehicles.The viBe for vehicle detection is applied in our framework with the resulting benefits.

Firstly, the background model can be updated through the ViBe foreground detection algorithm, and after updating the background model, we can efficiently deal with the noise[5,6,8,9] and anomaly points that are occurred due to slight variations of the brightness and can be completely suppressed in images.

Secondly, this algorithm considers a certain part in the image for background modeling instead

DOI:10.9756/INTJECSE/V14I5.79 ISSN: 1308-5581 Vol 14, Issue 05 2022

of modeling the entire part of the image, which hugely reduces the computational load.

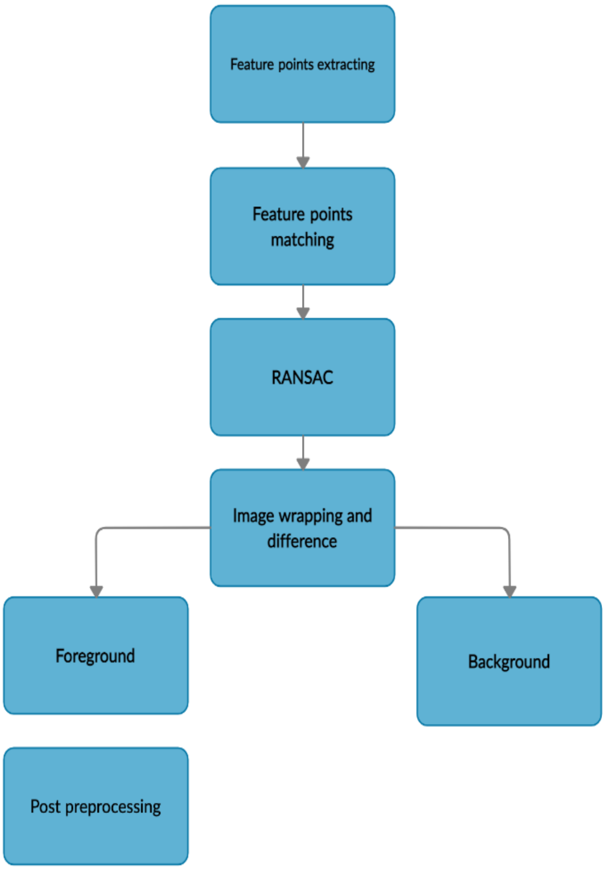

Figure 5. Moving Background

In this model firstly, extracting the SURF feature points that are employed to match the two frames. Secondly, the transformation between the frames is estimated using the RANSAC algorithm. Next, we transform the camera coordinates of nearby frames to a system namely the reference coordinate system. Then we extract the foreground using the image difference method. Finally, the results are processed through the morphological method.

V. Result And Analysis

The input may be an video or real-time camera video for detection and tracking. For moving background, image-registration is used to estimate the camera motion, which allows detecting vehicle in a reference frame.

Input:

Before Detection

Figure 6: Input video file

The commonly used vehicle counting method is based on the regional mark and the virtual test line. The former method is to count the number of connected areas, while the latter sets up a virtual test line on the road. We define an area that is used to count vehicles. Each tracker tracks an independent object with no interference between the objects, which ensures that the status information of each object is not confused. When the result of the tracker is unreliable, the detector

DOI:10.9756/INTJECSE/V14I5.79 ISSN: 1308-5581 Vol 14, Issue 05 2022

reinitializes the corresponding tracker. In terms of multiple trackers processing, we employ multi-threading technology, which can greatly reduce the computational load. The frames per second for reading the video are average of 30fps.

Output

After Detection

Figure 7: Output frame with FPS value, Current Vehicle Count, and Total Vehicle Count.

VI. Conclusion

By the experimental results and thesis, we were able to analyse the output where object detection was more precise and accurate and also individually identify the objects with the exact location of the object in the provided picture along the x, y-axis. The proposed method uses two major situations namely static background and dynamic background. Under static, we use a foreground detector to overcome small exceptions of an original picture by renewing the model. While in moving the background, we estimate the motion of the camera through the image registration that usually allows detection of a vehicle in a reference frame. Besides, to address the transformation of shape and scale of the vehicle in images, we employ an online- learning tracking method in our framework, which can renew the samples utilized for training and designing a
multiple object management module for effectively connecting the detector and tracker through the use of multi-threading technology and thus intelligently analyse the status of tracked Vehicle. The experimental outcomes of 16 aerial videos reveal that the proposed method yields higher than 90% and an accuracy of 85% accuracy on fixed-background videos and moving-background videos, respectively.

VII. Future Scope

In this proposed system, multiple objects are detected and tracked on various frames of the video. Hence further training the models on powerful GPUs and also increasing the count of images to evaluate the models on the different datasets. Further to modify the design if required to obtain the model more suitably and robustly for real-time applications.

References

[1] Lalita Gavit, Reenal Sanghavi, Manasi Parab, Prof. Mohini P. Sardey, "Object Tracking Using Multiple Cameras" IJTEEE Paper Published in 2014, Volume 2.

[2] J L Mazher Iqbal and K. Iswarya Lakshmi, "Object detection and tracking using Thermal camera" International Journal of Pure and

Applied Mathematics Volume 114 No. 10 2017, 87-95.

[3] Mukesh Tiwari and Rakesh Singhai "Improved Algorithm for Object Tracking in Video Camera Network", Indian Journal of Science and Technology, Vol 10(10).

[4] Sayanan Sivaraman and Mohan Manubhai Trivedi,"Looking at Vehicles on the Road: A Survey of Vision-Based Vehicle Detection, Tracking, and Behavior Analysis" IEEE transactions on intelligent transportation systems ,2013 IEEE.

[5] S. R. Reeja and N. P. Kavya, "Real time video denoising," *2012 IEEE International Conference on Engineering Education: Innovative Practices and Future Trends (AICERA)*, 2012, pp. 1-5, doi: 10.1109/AICERA.2012.6306745.

[6] Rino Cherian, Jothimani K. and Reeja S.R.. A Review on Object Tracking Across Real-World Multi Camera Environment. International Journal of Computer Applications 174(12):32-37, January 2021

[7] Dias N., Reeja S.R. (2020) An Improvement of Compelling Graphical Confirmation Plan and Cryptography for Upgrading the Information Security and Preventing Shoulder Surfing Assault. In: Arai K., Bhatia R., Kapoor S. (eds) Proceedings of the Future Technologies Conference (FTC) 2019. FTC 2019. Advances in Intelligent Systems and Computing, vol 1070. Springer, Cham. https://doi.org/10.1007/978-3-030-32523-7_30

[8] Soja Rani S., Reeja S.R. (2020) A Survey on Different Approaches for Malware Detection Using Machine Learning Techniques. In: Karrupusamy P., Chen J., Shi Y. (eds) Sustainable Communication Networks and Application. ICSCN 2019. Lecture Notes on Data Engineering and Communications Technologies, vol 39. Springer, Cham. https://doi.org/10.1007/978-3-030-34515-0_42

[9] Jose J.M., Reeja S.R. (2022) Anomaly Detection on System Generated Logs—A Survey Study. In: Shakya S., Bestak R., Palanisamy R., Kamel K.A. (eds) Mobile Computing and Sustainable Informatics. Lecture Notes on Data Engineering and Communications Technologies, vol 68. Springer, Singapore. https://doi.org/10.1007/978-981-16-1866-6_59