VEHICLE DETECTION AND TRACKING FROM STATIONARY AND VIDEOS USING A HIGH RESOLUTION IMAGES BY MACHINE LEARNING AND COMPUTER VISION APPROACH.

Arpitha.K,
Department of Computer Science and Engineering,
Dr.Ambedkar Institute of Technology,
Bangalore-56,India

Dr.M.V.Vijaya Kumar,
Department of Information Science and Engineering,
Dr.Ambedkar Institute of Technology,
Bangalore-56,India

***Abstract*—In computer vision technology video surveillance seems to be one of the hot research topic where it exactly detects the objects, recognize the objects and track the objects from a given sequence of image(video).Apart from this it also aims at understanding and delivering of the objects behavior by replacing the old method of human being monitoring the cameras to detect and track the objects and also this technique is one among the important and challenging tasks of the computer vision which involves detection and tracking of the objects from a video sequence. To detect the vehicle from the dashboard video using machine learning techniques we use already existing datasets like GTI and KITTI vision. In this paper we use techniques of machine learning computer vision such as spatial binning, HOG, color histograms for extracting features and SVM is used as training classifier and other techniques such as sliding window, heat maps and many other techniques will be used to detect the object (vehicle) from given dashboard video and resultant is also saved in the video format. Apart from vehicle detection from the dash board video has to be tracked too this can be achieved with the combination of Kalman Filter and Optical Flow algorithm.**

Keywords*—Hog, Spatial binning, Support Vector Machines.

I. Introduction

For the robot to work in outdoor environment for example self-driving cars etc vehicle detection and tracking becomes their major tasks. As we have already described about object detection here object tracking means procedure to analyze the digital images optical flow is particularly used as a flexible representation of visual motion. By using the optical flow vectors it becomes easy to detect and track the moving single object in a given video and also we know that Kalman filter is used to remove the noise that will be affecting the subtracted background image and accurately predict the position of object. Whereas in our paper we try to represent the combination of these Kalman filter and optical flow for obtaining the accurate object detection and tracking system. While vehicle detection is taken care by the techniques of machine learning computer vision such as spatial binning, HOG, color histograms for extracting features and SVM is used as training classifier and other techniques such as sliding window, heat maps are used to detect the object (vehicle) from the given dashboard video and resultant is also saved in the video format. Vehicle detection and tracking in the moving state can be categorized in to two types: 1.From the forward direction .2. From behind. Whereas tracking of object can be partitioned into two: 1. From each frame we should detect the object (vehicle) which is in moving state 2. After tracking the object analysis should be done and occlusion should also be handled. When the object movement is uniform or non-uniform state to monitor those we use Gaussian mixture model as a background subtraction procedure. We also now the optical flow use already but when the objects are occluded it becomes challenging task to track the multiple objects we use optical flow algorithm. Tracking execution can be based on the exact feature extracted and the moving objects position in the whole video, with the help of Kalman filter and optical flow we know that multiple objects can be tracked simultaneously. The paper proposes object tracking and detection and background subtraction especially for the applications of visual surveillance. While designing this system our aim is to provide timely service with robustness and low computational cost as well

II. Related works

Before you begin to format your paper, first write and save the content as a separate text file. Complete all content and organizational editing before formatting. Please note sections A-D below for more information on proofreading, spelling and grammar. The already existing system for detection and tracking of vehicle use to make fully use of the shadows casted by the vehicles. For all the vehicles around the common attributes it exhibits to detect the vehicle is the shadow that it produces on a particular area. When it comes to outdoors illumination depends basically on the outdoor weather (cloudy/sunny) , which in turn makes difficulty in setting the thresholds for that particular area. It is difficult to differentiate between the cast and self-shadow by the region segmentation and boundary detection algorithm in the pixel domain since they both possess similar characteristics. It is difficult to rule out the cast shadow based on the detected edge information .Hence in this paper we shall be using other techniques to detect and track the objects (vehicle).

ChapleManisha, Dr.VyasVibh and Prof.PaygudeS.S [1] et al in 2013 suggested in their paper stated that the movement of the centroid of the car over the frames is used to calculate the distance travelled by the car and the estimation of the speed of the vehicle. But in Dynamic environment, the tracking system were not accurate.

J.Priyadarshini and D.Sudha [7] et al in 2016 suggested in their paper stated the robust detection model efficiency which in turn benefits the dynamic environment as well.

Jing Li, Quan Pan, Stan Z Li, Tao Yang [2] et al in their paper suggested the method of object segmentation, spitting and merging and stated that objects in different conditions like indoor and outdoor along with changing background is detected in the real time by occlusion handling of long deviation, but failed in providing better accuracy since many techniques were included and was also limited for the videos of 15-20 fips only.

Nishu Singla [3] et al in their paper suggested the method of motion detection and frame difference and introduced a very new algorithm based on the frame difference in the static background to detect the moving objects but failed to detect multiple objects since they worked on motion based detection which doesn't hold good for real time detection.

Kshivsagar, Kanchan Doke, Archana Vilas [6] et al in their paper they introduced a template matching for P-N learning technology and did the comparison of other existing template with their current template by applying threshold , but failed for occlusion handling and as the object increased the performance decreased.

III. Proposed System

In this paper we design a system which will detect the vehicles and also track them in a given input video sequence with a stationary background and as well as dynamic background .We use different technique for object tracking, background subtraction and crowd flow estimation for visual surveillance application.

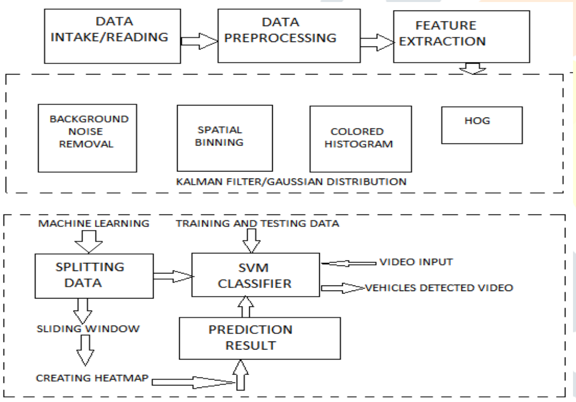


Fig 1: The Vehicle Detection and Tracking System Architecture.

To detect and track vehicles we follow the following steps:

1. Video to Image Conversion: We have to convert real time video into image to perform the process. Video is a collection of images. An image helps to process faster in the real-time video.

2. Detecting the Multiple objects in the image: Extracting the images from video, we have to preprocess the images (frames) to detect the multiple objects. Here we consider the human as objects.

3. Removing the unwanted objects in the image: Once human objects detected in the images we have to discard the unwanted objects in the images, may be its static or dynamic objects.

5. Object detection and Tracking in Different environment: Object detection in the different environment is very complex, we have to keep track on the environment, if we are detecting day time, and the process is different as well as if you detecting an object in night time the process will be different.

6. Occlusion Handling: It helps to keeps on the detected multiple objects separately in the real time dynamic environment.

IV. Methodology

A. *DATASET COLLECTION*

The most important thing for any machine learning problem is the labeled data set and here we need to have two sets of data: Vehicle and Non Vehicle Images. The images were taken from some already available datasets like GTI and KITTI Vision. The images are of size 64x64 and somewhat looked like below

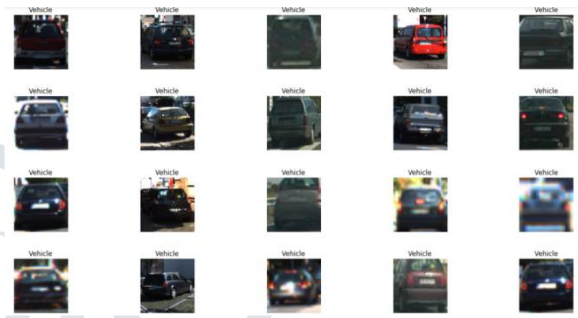


Fig 2 : Vehicle Images.

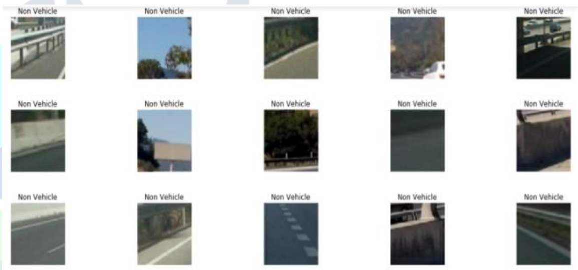


Fig 3 : Non-Vehicle Images

B. *PRE-PROCESSING OF DATA*

In this step we pre-process that has already been collected. This pre-process step is done to separate dataset into training and testing by performing shuffling and splitting on the dataset and data is been normalized and scaled too. In this process it is very important to note that we should not let our classifier to sneak peek in to our test data though we fit and transform all the data given.

C. *FEATURE EXTRACTION*

The following represents the methods of extracting feature.

1) Histograms of colors

The most simple and intuitive way is to extract the features from various color channels of the images. This can be done by plotting the histograms of various color channels and then collecting the data from the bins of the histogram. These bins give us useful information about the image and are really helpful in extracting good features.

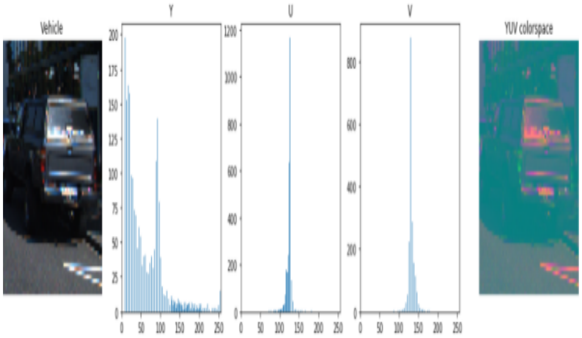


Fig 4 : Vehicle Image Histograms of Colors

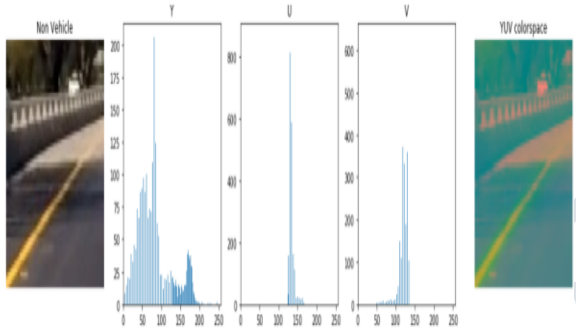


Fig 5 : Histogram of colors for Non-Vehicle Image

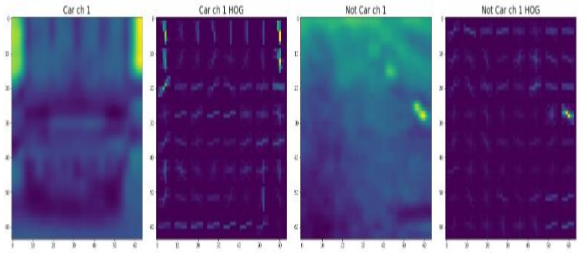
2) Spatial Binning

Color Histograms was certainly great, but if the features of an image are so much important then we take all the features using some sort of num.py function .We can extract all the information from the image by flattening it using numpy.ravel( ). But doing some calculation, image size is 64x64 and it is a 3 channel image so the total number of features extracted is 12,288!! Close to 12k features from a single image is not a good idea.So here Spatial Binning comes to picture. If we say, a 64x64 image gives the same information as 16x16 gives, Of course there is some loss of information but still we are able to extract good features out of the image! So if I apply numpy.ravel() to a 16x16 image, I would get only 768 features.

. **Fig 6 : Diagram of Spatial Biining**

3) Histogram of Oriented Gradients(HOG)

Feature extraction techniques discussed above are pretty cool but certainly not much powerful as compared to HOG. HOG actually takes an image, divides it into various blocks in which we have cells, in cells we observe the pixels and extract the feature vectors from them. Different orientations are made got by the a block. Here we are not counting the occurrence of a pixel in a particular orientation but instead we are interested in the magnitude of the pixel in that particular orientation. The Kalman filter keeps track of the estimated state of the system and the variance and the uncertainty of the estimate.



. **Fig 7 : HOG with 15 orientations.**

4) Background Noise Removal

The variance, the estimated state of the system and uncertainty of the estimated is kept on track by a Kalman Filter. The estimate is updated using a state transition model and measurements. X estimate of the system's state at time step k before the k-th measurement Yk has been taken into account; where P represents the corresponding uncertainty. The Kalman filter tries to estimate the state a ∈Rn of that system which is governed by the vector difference equation:

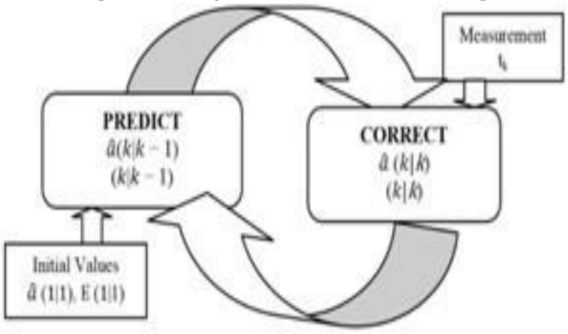


Fig 8 : The Kalman filter Predict/Correct model.

The prediction steps

- State prediction
- Error covariance prediction

The correction step

- Measurement prediction
- Residual
- Measurement prediction covariance
- Kalman gain
- State update
- Error covariance update

D. TRAINING CLASSIFIER

Support Vector Machines because they have good compatibility with HOG. Now in SVM we have SVC(Support Vector Classifier) and here also we have a choice with various kernels and different C and gamma values. We trained my classifier on both Linear and rbf kernel. To train with a test accuracy of 98.7% the Linear Kernel took around 1.8 seconds, whereas RBF kernel test accuracy was 98.3% and took 25 minutes. So we decided to use the Linear SVC along with the default parameters, since the time and accuracy was more compared to RBF kernel.

E. SLIDING WINDOW

Our classifier is now well trained and it will 99% of time will be able to predict vehicles and non-vehicles correctly .The next step is to apply the classifier to patches of your image in order to find where exactly in the image the car is.

But first you need to decide on various important parameters. The first thing is from where do you start searching the car from, obviously you should not search the car in the sky, hence you can ignore the top half of the image, so basically decide a horizon under which you will search your cars. The second important thing is what will be the window size you will look for and how much two windows should overlap. That depends on your input image length, since here it is 64x64 so we are going to start with base window size of 64x64 only. The next important thing and very important point here to note is that near the horizon we tend to search the small cars and as you move towards the dashboard camera you search for larger cars. This is because if the cars are near to horizon they are smaller as they are distant from your car and reverse is the case with the near cars.

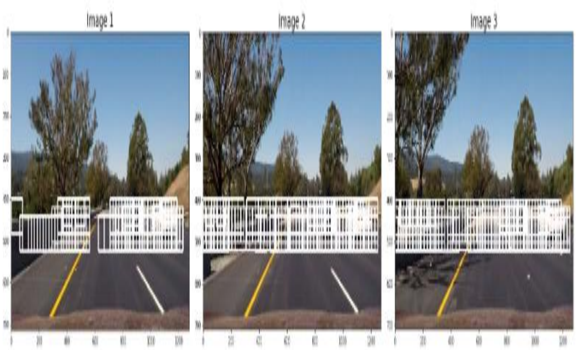


Fig 9 : Sliding Window

F. HEATMAP

To find the final bounding box where there are many widows overlapping with each other we use heat map. After creating a blank image which is same in size as that of the original image and will add the pixels values one by one for the whole region of the refined for all the refined windows that were identified .At last we apply threshold clip to the final images and finally get the final box co-ordinates.

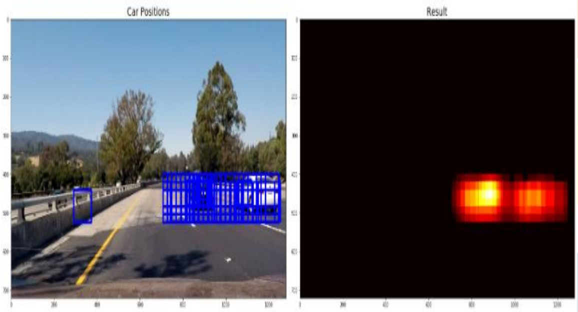


Fig 10 : Heat Map

G. OPTICAL FLOW ALGORITHM

The computer vision to characterize and quantify the motion of the object in a video stream we use optical flow estimation and also used often for motion based detection and tracking systems. This is nothing but the distribution of the apparent velocities of objects in an image and also the displacement field for each of the pixels in image. One can measure the velocities of the objects in the given video by using estimation of optical flow between the video frames. More apparent motion will be possessed by the vehicles that are closer to the camera rather than that of which are moving with the same speed. For instance we have assume there is no intensity change with each pixels moving on. In frame 2 the pixel is at ($x+\Delta x$, $y+\Delta y$) whereas in frame1 pixel is at location (x, y). Displacement vector with each pixel is associated with optical flow.

Video is taken as input (Real time Video)

1. With varied frame and difference delay we have to extract vertical components and horizontal components
2. Each frame mean is found.
3. To remove noise we have to apply median filter.
4. To each frame erosion operation and morphological close must be applied.
5. Vectors of optical flow are stored as complex numbers which in turn used to estimate the optical flow.
6. Threshold is calculated by computing their magnitude squares.
7. From the matrix of complex velocities, velocity threshold is calculated.
8. The image must be threshold.
9. To fill the holes in the blobs thinning of the objects must be applied.
10. The bounding box and the area of the blobs must be estimated.
11. Around the tracked/detected objects(vehicle) bounding boxes must be drawn.
12. Motion vectors are then calculated and drawn.
13. The resultant vehicle detected and tracked video must be displayed.

V. CONCLUSION

The Motion segmentation algorithms performance is improved using our proposed system. As said the goal was to design a system to provide timely service with robustness and low computational cost. Hence in processing video stream for event analysis in real time we have designed methods which are proactive. The ability to track the object in real adverse situation is the strength of the Kalman Filter. It can also assist the human operators by identifying the important events in videos given and responding with a timely manner to them. Among all the methods discussed above promising one is the improved optical flow algorithm. Since it gives more accuracy with less computational time. The future work of this is done by using novel tracking algorithm for the time buffering video and to achieve more all their performance is also compared accordingly.

REFERENCES

[1] Chaple Manisha, Dr.Vyas Vibha and Prof.Paygude S.S, “Vehicle Detection and Tracking using optical flow and background subtraction”, “Proc.of Int.Conf on Advances in Computer Science and Applications”.

[2] Stan Z.Li, Quan Pan, Jing Li and Tao Yang, “Real –time Multiple objects Tracking with Occulsion Handling in Dynamic scenes”,”National Laboratory of Pattern Recognition and National Natural Science Foundation”.

[3] Nishu Singla, “ Motion Detection Baesd on Frame Difference Method”.”International Journal of Information and amp;Computation Technology”,Volume 4, Number 5(2014).

[4] Kapildev Naina and Payal Ranipa,”Real-time Moving Object Tracking in Video Processing”,”International Journal of Engineering Research and General Science”,Volume 3, Issue1, January-February,2015.

[5] G.U.Kharat and Moiz a.Hussain,”Precise Human Motion Detection and Tracking in Dynamic Environment”, “International Journal of Signal Processing”.

[6] Kanchan Doker and Archana Vilas Kshisagar,” Multiple Object Detection using P-N learning Template Matching”,”International Journal of innovative Research in Computer and Communication Engineering.(An ISO 3927:2007 Certified Organisation)”,Vol.4,Issue 7,July 2016.

[7] J.Priyadarshini and D.Sudha,”Multiple Vehicle Detection and Tracking in Dynamic Environment”,”International Journal of signal Procesing.

[8] Girisha.R, Vasudev T and Chandrajit M ,” Multiple Object Tracking in Surveillance Video using color and Hu moments”,”An International Journal (SIPIT)”,Vol.7,No.3,June 2016.