

**Indian Institute of Information Technology,
Allahabad
Department of Information Technology**



**Image and Video Processing Course 2021
Progress Report on
Intelligent Traffic Management System**

as part of C1 assessment

by

Nidhi Kamewar - IIT2019189

Vanshika Garg - IIT2019216

Anshuman Bhardwaj - IIT2019227

Anirudh Gupta - IIT2019228

Vishwam Shriram Mundada - IIT2019235

Ishneet Sethi - IIT2019237

Chandramani Kumar - IIT2019238

1 Abstract

Traffic monitoring and traffic control have always been challenging tasks. Intelligent Transportation Systems (ITS) based on wide range of technologies have certain practical challenges in their application and implementation. Video surveillance has proven advantageous over traditional systems based on inductive loops sensors and detectors for traffic monitoring. Accurate traffic density estimation which is basic to tackling traffic congestions requires detection of vehicles, assessing their speed, and tracking vehicles passing through surveillance zones. Image processing techniques require processing of large number of image frames for real-time applications in traffic management. More efficient and less costly image processing techniques for accurate vehicle detection and density determination are required for developing more effective traffic management systems. There is a need for developing algorithms with robust performance under heavy traffic loads and varied environmental conditions. Developments in artificial intelligence offer new vistas in image processing for regulation and management of traffic by signal control mechanisms and creation of neural networks for unhindered traffic flow.

Index Terms: Image processing techniques, Intelligent traffic management, Traffic monitoring, Vehicle detection

2 Introduction

In modern life we have to face with many problems one of which is traffic congestion becoming more serious day after day. It is said that the high tome of vehicles, the scanty infrastructure and the irrational distribution of the development are main reasons for augmented traffic jam. The major cause leading to traffic jam is the high number of vehicle which was caused by the population and the development of economy. To unravel this problem, the government should encourage people to use public transport or vehicles with small size such as Vietnam, the local authorities passed law limiting to the number of vehicles for each family. The methods mentioned above is really efficient in fact. That the inadequate infrastructure cannot handle the issue of traffic is also a decisive reason. the public conveyance is available and its quality is very bad, mostly in the establishing countries. Besides, the highway and roads are incapable of meeting the requirement of increasing number of vehicles. Instead of working on the roads to accommodate the growing traffic various techniques have been devised to control the traffic on roads like embedded controllers that are installed at the junction. These techniques are briefly described in next section.

2.1 Standard Traffic Control Systems :

2.1.1 Manual Controlling

Manual Controlling the name instance it require man power to control the traffic. Depending on the countries and states the traffic polices are allotted for a required area or city to control traffic. The traffic polices will carry sign board, sign light and whistle to control the traffic. They will be instructed to wear specific uniforms in order to control the traffic.

2.1.2 Automatic Controlling

Automatic traffic light is controlled by timers and electrical sensors. In traffic light each phase a constant numerical value loaded in the timer. The lights are automatically getting ON and OFF depending on the timer value changes. While using electrical sensors it will capture the availability of the vehicle and signals on each phase, depending on the signal the lights automatically switch ON and OFF.

2.2 Drawbacks :

In the manual controlling system we need more man power. As we have poor strength of traffic police we cannot control the traffic. ON the other side, automatic traffic controlling a traffic light uses timer for every phase. Using electronic sensors is another way in order to detect vehicles and produce signal that to this method the time is being wasted by a green light on an empty road. Traffic congestion also occurs while using the electronic sensors for controlling the traffic. All these drawbacks are supposed to be eliminated by using image processing.

2.3 Image Processing in Traffic Light Control :

We propose a system for controlling the traffic light by image processing. The vehicles are detected by the system through images instead of using electronic sensors embedded in the pavement. A camera will be placed alongside the traffic light. It will capture image sequences. Image processing is a better technique to control the state change of the traffic light. It shows that it can decrease the traffic congestion and avoids the time being wasted by a green light on an empty road. It is also more reliable in estimating vehicle presence because it uses actual traffic images. It visualises the practicality, so it functions much better than those systems that rely on the detection of the vehicles' metal content.

2.4 Introduction to Image Processing :

Image Processing is a technique to enhance raw images received from cameras/sensors placed on space probes, aircrafts and satellites or pictures taken in normal day-to-day life for various applications. An image is rectangular graphical object. Image processing involves issues related to image representation, compression techniques and various complex operations, which can be carried out on the image data. The operations that come under image processing are image enhancements operations such as sharpening, blurring, brightening, edge enhancement, etc. Image processing is any form of signal processing for which the input is an image, such as photographs or frames of video; the output of the image processing can be either an image or a set of characteristics or parameters related to the image. Most image-processing techniques involve treating the image as a two-dimensional signal and applying standard signal-processing techniques to it. Image processing usually refers to digital image processing, but optical and analog image processing are also possible.

3 Methodology being used

This section presents the essential steps in project implementation, starting from a image/video of a traffic scene and ending in vehicle detection and counting. First, the image/video data of the traffic scene are entered. Then, the road surface area is extracted and divided. The YOLO deep learning object detection method is used to detect the vehicle object in the highway traffic scene.

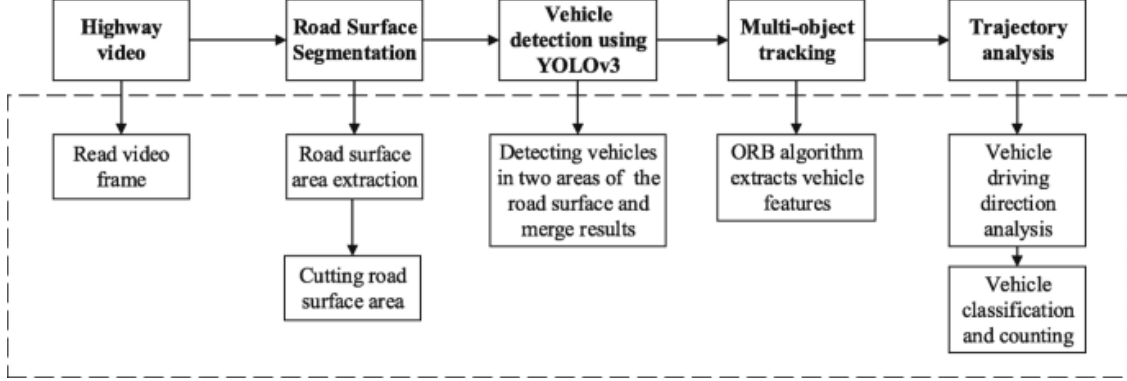


Fig.1: Vision Based vehicle detection using deep learning in highway scenes

3.1 Vehicle Classification

We will be classifying vehicles into different categories. Classification is done by categorizing the vehicles into three classes according to the size of the vehicles, namely, large, medium and small. Since it is easy to find the length of vectors, the length has been taken as the parameter to classify vehicles according to the defined sizes. The following table shows the classification used in this work.

Large	Containers, Lorries, Buses
Medium	Cars, Vans, Pickups
Small	Two wheelers, Three wheelers

Table 1: Classification of vehicles

For each new vehicle that enters into the line on the region of interest, the length of vector has been calculated and subjected to the minimum length requirement. Classification was carried out for those vehicles that pass the length requirement.

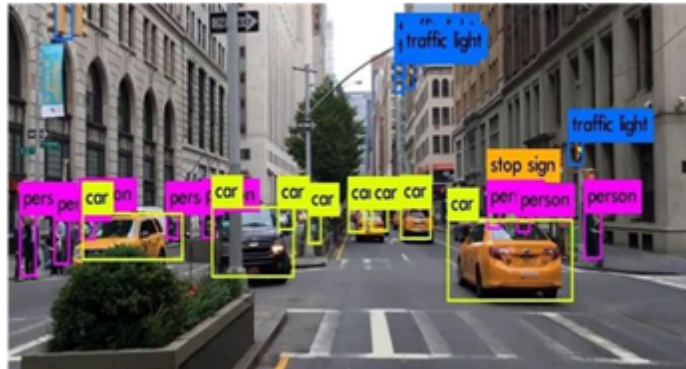


Fig.2 : Vehicle classification

3.2 Vehicle detection

First we detect vehicle candidates using different models. These model creates a common sketch from a small number of training images. Since we only focus on vehicles with more than three wheels, training images are captured from the top and front views. Different cars, vans and trucks or buses have more common structures in this view. When a vehicle is detected as a candidate by this algorithm, we extract it within its bounding box and check its reflection symmetry by calculating the correlation value.

3.2.1 *Convolution Neural Network (CNN)*

CNN is widely used neural network architecture for computer vision related tasks. Advantage of CNN is that it automatically performs feature extraction on images i.e. important features are detected by the network itself. CNN is made up of three important components called Convolutional Layer, Pooling layer, fully connected Layer. Considering a gray scale image of size 32×32 would have 1024 nodes in multi-layer approach. This process of flattening pixels loses spatial positions of the image.

3.2.2 *Region-based Convolution Neural Networks (RCNN)*

The Region-based Convolutional Network method (RCNN) is a combination of region proposals with Convolution Neural Networks (CNNs). R-CNN helps in localising objects with a deep network and training a high-capacity model with only a small quantity of annotated detection data. It achieves excellent object detection accuracy by using a deep ConvNet to classify object proposals. R-CNN has the capability to scale to thousands of object classes without resorting to approximate techniques, including hashing. The fig 4. shows that Regional based convolutional neural network.

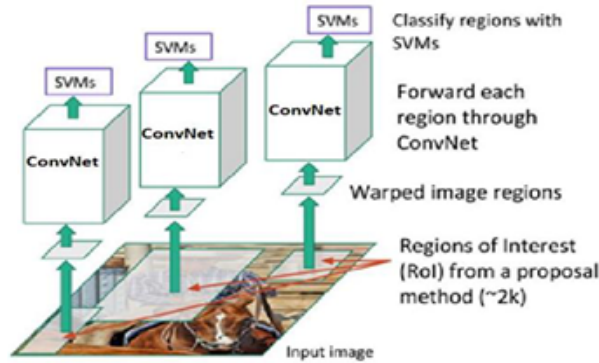


Fig.3 : Regional based CNN

3.2.3 *Single Shot MultiBox Detector*

Single Shot Detector (SSD) is a method for detecting objects in images using a single deep neural network. The Single Shot Detector network combines predictions from multiple feature

maps with different resolutions to naturally handle objects of various sizes. Fig 5. shows Single Shot MultiBox Detector.

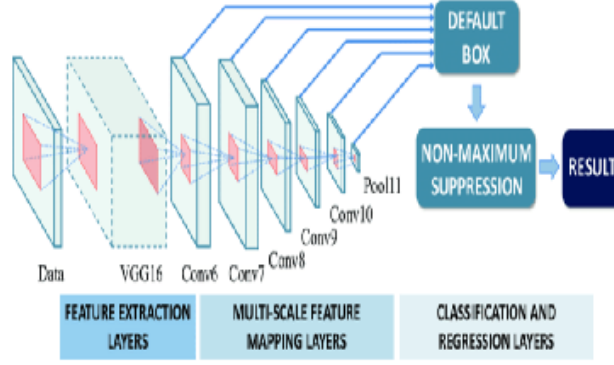


Fig.4 : Single Shot Multiplex Detector

3.2.4 YOLO Algorithm

One way for traffic controlling is to control the signals in real time and operate them automatically according to demand in traffic.

This section describes the object detection methods used in this study. The implementation of the highway vehicle detection framework used the YOLOv3 network. The YOLOv3 algorithm continues the basic idea of the first two generations of YOLO algorithms. The convolutional neural network is used to extract the features of the input image. According to the size of the feature map, such as 13×13 , the input image is divided into 13×13 grids. The centre of the object label box is in a grid unit, and the grid unit is responsible for predicting the object. The network structure adopted by YOLOv3 is called Darknet-53. This structure adopts the full convolution method and replaces the previous version of the direct-connected convolutional neural network with the residual structure. The branch is used to directly connect the input to the deep layer of the network. Direct learning of residuals ensures the integrity of image feature information, simplifies the complexity of training, and improves the overall detection accuracy of the network. In YOLOv3, each grid unit will have three bounding boxes of different scales for one object. The candidate box that has the largest overlapping area with the annotated box will be the final prediction result. Additionally, the YOLOv3 network has three output scales, and the three scale branches are eventually merged. Shallow features are used to detect small objects, and deep features are used to detect large objects; the network can thus detect objects with scale changes. The detection speed is fast, and the detection accuracy is high. Traffic scenes taken by highway surveillance video have good adaptability to the YOLOv3 network. The network will finally output the coordinates, confidence, and category of the object.

When using YOLO detection, images are resized to the same size, such as 416×416 , when they are sent to the network. Since the image is segmented, the size of the remote road surface becomes deformed and larger. Therefore, more feature points of a small vehicle object can be acquired to avoid the loss of some object features due to the vehicle object being too small. The dataset presented in “Vehicle dataset” section is placed into the YOLOv3 network for training, and the vehicle object detection model is obtained. The vehicle object detection model can detect three types of vehicles: cars, buses, and trucks. Because there are

few motorcycles on the highway, they were not included in our detection. The remote area and proximal area of the road surface are sent to the network for detection. The detected vehicle box positions of the two areas are mapped back to the original image, and the correct object position is obtained in the original image. Using the vehicle object detection method for obtaining the category and location of the vehicle can provide necessary data for object tracking. The above information is sufficient for vehicle counting, and the vehicle detection method thus does not detect the specific characteristics of the vehicle or the condition of the vehicle.

4 Datasets being used

IDD (Indian Driving Dataset):

published by Indian Institute of Information Technology, Hyderabad

[dataset](#)

Google Image Dataset:

[dataset](#)

Vehicle Detection dataset:

We found this dataset on kaggle which contains more than 5000 images which will surely help train and test our model.

[dataset](#)

5 Results and discussion

Most of the studies done have dealt with highway and urban roads and the most popular techniques used in recent years were either background subtraction or feature-based techniques. Despite all these techniques, there are some issues that have negatively impacted such systems. Examples of such challenges include camera view and operating condition, which present additional limitations. ITSs face a number of difficulties, especially with urban roads traffic scenes and intersections in which heavy traffic, vehicle occlusion, and camera placement affect the performance of the systems. These challenging issues still require additional research and development, especially in the case of urban roads. In recent years, studies have focused on how to detect vehicles in complex scenes and how to detect vehicles under occlusion objects or in poor lighting situations. Researchers and also tried to use huge vehicles data sets in their efforts to achieve high accuracy.

6 Conclusion

Through this paper we have shown various algorithms and techniques of detecting vehicles image-based traffic surveillance and monitoring systems. Vehicle detection falls into two major branches: motion-based and appearance-based techniques. Both techniques can be applied to isolate vehicles from the background scene with varying computational complexity and detection accuracy. Following the descriptions of various existing techniques, it can be observed that most methods achieved high levels of accuracy. However, most of detection systems still have some limitations which negatively impact the accuracy levels. These limitations include poor lighting, weather changes, occlusion, and shadows.

As a future direction for further research on this topic of vehicle detection, we will try to focus on detecting vehicles and identify their licence number using Natural Language Processing Techniques using RCNN which will greatly improve our surveillance efficiency. There is further scope of research under occlusion and detection of the front and rear view of the vehicle under poor lighting conditions while removing the shadow from the scene.

7 Individual Contribution

Nidhi Kamewar (IIT2019189)

Explored the road surface segmentation method to detect traffic at highways.

Vanshika Garg (IIT2019216)

Explored ORB algorithm to extract features of the detected vehicles.

Anshuman Bhardwaj (IIT2019227)

Explored different versions of RCNN algorithm and tried to execute already available models based on Faster RCNN

Anirudh Gupta (IIT2019228)

Tried YOLO v3 for image detection on Google Colab and Jupyter notebook.

Vishwam Shriram Mundada (IIT2019235)

Tried YOLO v3 for image detection on Google Colab. Also explored different datasets available on internet regarding vehicle detection

Ishneet Sethi (IIT2019237)

Explored the algorithms R-CNN, YOLO to generate potential bounding boxes in an image and the different datasets available on internet.

Chandramani Kumar (IIT2019238)

Tried YOLO v4 for image detection on Google Colab and tested on different datasets available on internet regarding vehicle detection

8 References

- [1] M. Swathy, et al., "Survey on vehicle detection and tracking techniques in video surveillance," *International Journal of Computer Applications*, vol. 160, no. 7, pp. 22-25, 2017.
- [2] J. J. Antony, M. Suchetha, "Vision based vehicle detection: A literature review," *International Journal of Applied Engineering Research*, vol. 11, no. 5, pp. 3128–3133, 2016.
- [3] R. D. Sharma, S. K. Gupta, "Moving object detection and tracking based on background subtraction," *Object Detection, Classification, and Tracking Technologies*, vol. 2018, Oxford Journal of Intelligent Decision and Data Science, 2018, pp. 55–62.
- [4] S. Jeeva, M. Sivabalakrishna, "Survey on background modeling and foreground detection for real time video surveillance," *Procedia Computer Science*, vol. 50, pp. 566–571, 2015.
- [5] N. C. Joy, J. Prasad, "Survey on background modeling and foreground detection methods," *International Journal for Innovative Research in Science Technology*, vol. 3, no. 4, pp. 338–41, 2016.
- [6] M. Shehata, et al., "Vehicles detection based on background modeling," *International Journal of Engineering Trends and Technology*, vol. 66, no. 2, pp. 92-95, 2018. Available: [arXiv:1901.04077](https://arxiv.org/abs/1901.04077).
- [7] S. Joudaki, et al., "Background subtraction methods in video streams: a review," 2015 4th International Conference on Interactive Digital Media (ICIDM), IEEE, 2015, pp. 1–6.
- [8] P. M. Tank, H. A. Patel, "Survey on human detection techniques in real time video," *International Journal of Innovative Research in Science, Engineering and Technology*, vol. 7, no. 4, pp. 5852–5858, 2018.