RESEARCH ARTICLE OPEN ACCESS

Recurrent Convolutional Neural Network Based Vehicle Counting Framework for Intelligent Transportation Systems

M. Thachayani [1], G. Rubavani [2]

[1] Department of ECE, Puducherry Technological University - Puducherry
[2] Department of ECE, Puducherry Technological University - Puducherry

ABSTRACT

Counting the number of on road vehicles is one of the most important task in intelligent transportation system. This helps in monitoring traffic status and provide information for traffic control providing better driving routes bypassing the congested roads. Counting is rather challenging due to limited information about vehicles and large variance in vehicle numbers. This paper presents a highway vehicle counting method from images obtained from surveillance cameras.

Keywords: - Intelligent transportation system, vehicle counting, R-CNN, image processing.

I. INTRODUCTION

An intelligent transportation system (ITS) is an advanced application, which aims to provide innovative services relating to different modes of transport and traffic management. It helps the users to make safer, more coordinated, and smarter use of transport networks. Some of these technologies include calling for emergency services when an accident occurs and using cameras to enforce traffic laws. Intelligent transport systems encompass various technologies and systems such as car navigation; traffic signal control systems; container management systems; variable message signs and automatic number plate recognition. More advanced applications that integrate live data and feedback from a number of other sources, such as parking guidance and information systems also comes under ITS.

Counting on-road vehicles in the highway is fundamental for intelligent transportation management. Counting is rather challenging due to limited information about vehicles and large variance in vehicle numbers. This work includes counting of two wheelers in addition to four wheelers. This paper presents the highway vehicle counting method using videos captured from roadside surveillance cameras. This counting is based on image processing. Image processing methods perform some operations on an image in order to get an enhanced image or to extract some useful information from it. It is a type of signal processing in which input is an image and output may be image or characteristics/features associated with that image. Image processing basically involves importing the image via image acquisition tools, and analyzing and manipulating the image. Image recognition using Neural networks employs artificial intelligence technology to automatically identify objects, people, places and actions in images.

This input images are obtained using videos captured from roadside surveillance cameras. Compared to the system employing specialized sensors, the visual analysis based counting system is cost effective as it uses the already existing roadside surveillance cameras.

II. LITERATURE SURVEY

Comparison of different deep learning based vehicle detection methods based on surveillance video is presented in [1]. Vehicle detection using the neural network based object detection algorithm, YOLO with HSV color model-based segmentation is proposed in [2]. It is reported that the best accuracy performance is achieved in case of YOLOv4 with KCF tracker. Shibo Zhang and XiojieWang[3] proposed the tracking and detection by the histogram of oriented gradients features which is used to represent the edge information of images. The main challenge in this approach is the employment of suitable non-uniform perspective grid of points which concentrates more points into informative regions and less concentrated on less informative ones. Cong Zhang and HongshengLi[4] proposed the usage of deep convolutional neural network for cross scene crowd counting. This process requires background subtraction using shadow elimination techniques. In paper [5], an object segmentation method based on Gaussian mixture model and conditional random fields is proposed. The limitation of this scheme is multi label energy is not taken into account for semantic image. S.Zhang and J.M.F.Moura Zhang[6] had shown that domain transfer learning can be explored to enable the model with better robustness in understanding traffic density from large scale web camera data. M.Lian, et al., [7] proposed a novel algorithm for counting and classification of highway vehicles by regression analysis. This work involves background estimation with shadow elimination technique. Ning He, Jiaheng Cao and Lin Song[8] proposed scale space theory to detect human in still images by designing and integrating more sophisticated multiscale decomposition approach for human detection. Li Liu, Songyang Lao, and Paul W. Fieguth [9] proposed a novel descriptor for texture classification, the median robust extended low binary pattern for applications such as image patching and object recognition. Yingying Zhang et al., [10] proposed single-image crowd counting using multi-column convolutional neural network and observed that the model trained on a source domain could not be easily transferred to a target domain by fine-tuning only the last few layers of the trained model. The objective of this study is to propose an effective Highway Vehicle Counting framework exhibiting minimum counting error for a wider range of vehicle density and size. The proposed method is implemented using python and evaluated extensively on real highway surveillance videos.

III. PROPOSED FRAMEWORK

Input surveillance camera videos are converted to n number of frames. Frame rate varies from one video to the other. Hence, video file reader is used which is specially designed to convert the videos with variable framerates into frames.

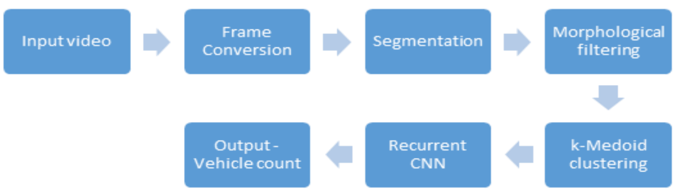

Fig. 1 Proposed Vehicle Counting Framework

Frame is segmented using the region based segmentation method known as Gaussian mixture model, which implements the region growing segmentation. Image segmentation technique is used to partition an image or grouping of pixels into meaningful parts having similar features and properties. Segmentation should provide a set of region having the connectivity and compactness, regularities of boundaries, homogeneity in terms of color and texture and differentiation from neighbor regions. Regions in an image are a group of connected pixels with similar properties. This process removes the background and segments the vehicle regions. Background subtraction is a commonly used class of techniques for segmenting out foreground objects from the background in a sequence of video frames. The name “background subtraction" comes from the simple procedure of subtracting the experimental image from the estimated image.

In the context of a traffic surveillance system, each background pixel is modeled using a mixture of three Gaussians corresponding to road, vehicle and shadows. This model is initialized using an expectation maximization EM algorithm. Then, the Gaussians are manually labeled in a heuristic manner as follows: the darkest component is labeled as shadow, in the remaining two components, the one with the largest variance is labeled as vehicle and the other one as road. For the foreground detection, each pixel is compared with each Gaussian and is classified according to it corresponding Gaussian. Segmented image is filtered for removing the unwanted noise using the morphological filtering.

K-Medoid partitional algorithm attempts to minimise the distance between points labelled to be in a cluster and a point designated as a center of that cluster. K-Medoid chooses data points as centers and can be used with arbitrary distances. K-medoid is a partitioning technique of clustering, which clusters the data set of n objects and into k clusters with the number of clusters assumed to be known, which implies that given value can be assessed with methods of clustering. It is more robust to noise and outliers are less as compared to k-means because it minimises a sum of dissimilarities instead of a sum of a squared Euclidean distances. A medoid can be defined as the object of a cluster whose average dissimilarities to all the objects in the cluster is minimal. Segmented and clustered image is processed by the recurrent convolutional neural network through the max pooling algorithm. Recurrent Convolutional neural network works by extracting features from images. This eliminates the need for manual feature extraction. The algorithm used is Fast RCNN. This takes the input image to ConvNet and returns feature maps of the image. From there it applies RPN(region proposal network)on these feature maps and the object proposals of the input image are obtained. It then resizes all proposals to the same size and passes it to the fully connected convolution layer in order to classify the boxes of the image. The fully connected layers combine the features together to create a model.

IV. RESULTS AND DISCUSSION

Number of training samples used is 50frames. The work shows that the proposed framework based on RCNN is promising for the real world video surveillance system based vehicle counting. The figures, 2 to 4 shows the vehicle detection output for sample frames with different composition. In Fig. 2, the frame contains only four wheelers in a lane setting and the count of the four wheeler in the frame is obtained in the output screen, correctly as 31.

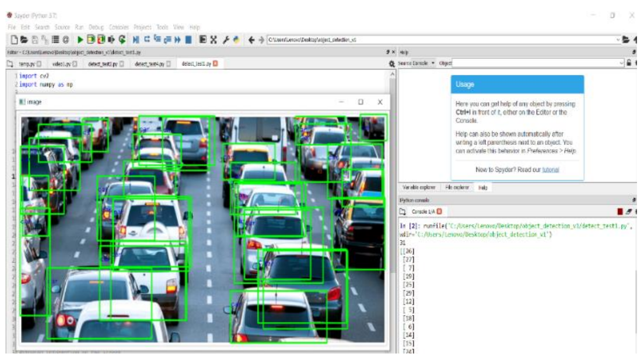

Fig. 2 Sample output for frame consisting of four wheelers in lane setting

The cars, which are only partially seen in the frame are also counted correctly. In Fig. 3, the frame consisted of assorted vehicles and the correct count of the vehicle, 13 is obtained for this case also.

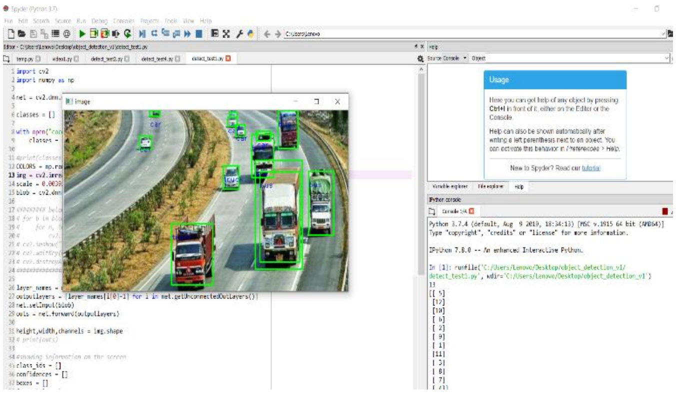

Fig. 3 Sample output for frame consisting of assorted vehicles

The ratio of correctly detected vehicles to the number of vehicles in the frame (ground truth) is measured and the average value obtained is 99.3%. The accuracy is in the range of 99 to 100 % for most of the frames. Only in frames having very dense and continuous stream of vehicles, the faraway vehicles could not be detected with 100 % accuracy. From the outputs obtained, it is ascertained that the proposed framework is very much effective in detecting and counting vehicles from frames extracted from surveillance camera videos.

V. CONCLUSIONS

An R-CNN based framework is proposed to provide vehicle count using frames obtained from surveillance videos. The system is coded in python and the accuracy performance is verified. The system achieved an average accuracy of 99.3%. This system providing automatic vehicle count can augment the traffic light control to achieve an intelligent traffic control and management system.

REFERENCES

[1] S. Venkatesh and B. S. Babu, "A Survey: Vehicle Detection and Counting," 2022 13th International Conference on Computing Communication and Networking Technologies (ICCCNT), Kharagpur, India, 2022, pp. 1-5.

[2] A. Stanley and R. Munir, "Vehicle Traffic Volume Counting in CCTV Video with YOLO Algorithm and Road HSV Color Model-based Segmentation System Development," 2021 15th International Conference on Telecommunication Systems, Services, and Applications (TSSA), Bali, Indonesia, 2021, pp. 1-5.

[3] Zilei Wang, Xu Liu, Jiashi Feng, Jian Yang,HongshengXi,“compressed domain highway vehicle counting by spatial and temporal regression,” *IEEE Transactions on Circuits and Systems for Video Technology*,Vol. 29, pp. 263-274, Issue.1, Jan. 2019.

[4] S. Zhang, G. Wu, J. P. Costeira, and J. M. F. Moura, “Understanding traffic density from large-scale web camera data,” in *Proc. IEEE Conf. Comput. Vis. Pattern Recogn.*, pp. 5898–5907,2017.

[5] Y. Zhou, L. Liu, L. Shao, and M. Mellor, “Dave: a unified framework for fast vehicle detection and annotation,” in *Proc. Eur. Conf. Comput. Vis.*, pp. 278–293,2016.

[6] S. Ren, K. He, R. Girshick, and J. Sun, “Faster r-cnn: Towards real-time object detection with region proposal networks,” in *Proc. Adv. Neural Inform.Process.Syst.*,pp.91–99,2015.

[7] S. S. Kruthiventi and R. V. Babu, “Crowd flow segmentation in com- pressed domain using CRF,” in *Proc. Int. Conf. Image Process.*, pp. 3417– 3421,2015.

[8] Shibo Zhang, XiaojieWang , “Human Detection and Object Tracking Based on Histograms of Oriented Gradients ,” in Proc. *IEEE Conf. Natural Computation*, pp. 1394-1353, Aug. 2013.

[9] H. T. Niknejad, A. Takeuchi, S. Mita, and D. McAllester, “On-road multivehicle tracking using deformable object model and particle filter with improved likelihood estimation,” *IEEE Trans. Intell. Transp. Syst.*, vol.13,pp.748–758,Jun.2012.

[10] H. Sabirin and M. Kim, “Moving object detection and tracking using a spatio-temporal graph in h. 264/avc bitstreams for video surveillance,” *IEEETrans.Multimedia*,vol.14,pp.657–668,Jun.2012.