International Journal of Computer Trends and Technology
ISSN: 2231-2803 / https://doi.org/10.14445/22312803/IJCTT-V67I3P102
Volume 67 Issue 3, 5- 8, March 2019
© 2019 Seventh Sense Research Group®

Original Article

Deep Learning-Based Vehicle Tracking in Traffic Management

Mohamed Shehata[1], Reda Abo-Alez[2], Farid Zaghlool[3], Mohamed Taha Abou-Kreisha[4]

[1,2,3,4] *Department of Systems & Computers Engineering, Faculty of Engineering, Al-Azhar University, Egypt*

Abstract - *The visual vehicle tracking system (VTS) has become a vital part of the Intelligent Transportation System (ITS). It is the cornerstone of vehicle behavior analysis. Our methodology for developing a VTS achieves video-based vehicle detection, classification, and tracking. A deep machine learning system based on a faster region conventional neural network (Faster-RCNN) detector is used in the detection process. A deep machine learning system based on a conventional neural network (CNN) is used in the classification process. The motion vector estimation (MVE) algorithm is used to determine vehicles' directions and positions in the video frames in the tracking process. Finally, a vehicle behavior understanding algorithm based on vehicle trajectory implementation and vehicle speed calculation is used to manage the traffic flow. After testing the developed VTS, the results show that 95% of the tested vehicles are precisely detected, 90% of the detected vehicles are successfully classified, and 92% of detected vehicle's tracks are well generated.*

Keywords - *Video Processing, Faster-RCNN, CNN, Motion Vector Estimation*.

I. INTRODUCTION

During the last decades, the total number of vehicles worldwide has grown enormously, increasing traffic density and resulting in more accidents, more congestion, and air pollution. More than 1.35 million people die each year on the world's roads, making road traffic injuries a leading cause of death globally for people aged from 5 to 29 years [1], with millions more sustaining serious injuries and living with long-term adverse health consequences. In addition, almost half of the victims are vulnerable road users such as pedestrians, bicycle users, and motorcyclists. Along with the grief caused by traffic accidents, it causes immense economic losses. The ITS becomes an important step in the fight to reduce road traffic deaths as a solution to various transport-related issues. Hence, continuous development in the ITS field must be considered to help stop the death and destruction on the world's roads. ITS mainly depends on video processing for road surveillance cameras. Traffic video processing has a great development in the past few years, and it aims to automatically detect, recognize and track vehicles from image sequences to understand and

Describe dynamics and interactions among them and provide important statistics and provide early warning, which contributes to enhancing traffic performance. Vehicle tracking is used to predict vehicle positions in subsequent frames, match vehicles between adjacent frames, and ultimately obtain the trajectory and location for each frame in the camera field of view of the vehicle [2]. The tracking method holds the vehicle trajectory by identifying motion dynamic attributes and characteristics to locate its position in every frame. The paper is organized as follows: Section II provides a briefing of related work. The overview of the overall developed system is explained in Section III. Section IV deals with tested results. And the paper is concluded with Section V.

II. RELATED WORK

Over the last few decades, many researchers have demonstrated vehicle tracking systems. Vehicle tracking can be classified into region-based and feature-based tracking [3]. At the same time, the region-based tracking techniques are characterized by complex computations and fail in crowded situations. At the same time, the feature-based tracking approaches are suitable for tracking vehicles with low pixels in the image by representing parts of a vehicle, and some related works are as follows:

Kiran Kale [4] developed an object detection and tracking algorithm which use optical flow in conjunction with motion vector estimation for object detection and tracking in a sequence of frames. The optical flow gives valuable information about the object movement, and the motion vector estimation technique estimates object position from consecutive frames. A median filter used a more robust algorithm in the presence of noise.

Nicholas A. Mandellos [5] developed an algorithm that can be applied in an existing traffic surveillance system by reconstructing the actual background based on statistical color sampling per pixel over time, followed by a background subtraction algorithm. The background reconstruction algorithm managed to accurately reconstruct the actual background in various harsh conditions, including heavy congestion and changes in the lighting, then used the mean shift algorithm and template matching algorithms for the tracking process.

Wei Zhang [6] developed a multilevel algorithm to detect and handle vehicle occlusion. It consists of the intraframe, inter-frame, and tracking levels. On the intraframe level, occlusion is detected by evaluating vehicles' compactness ratio and interior distance ratio. The detected occlusion is handled by removing a cutting region of the occluded vehicles. On the inter-frame level, occlusion is detected by performing subtractive clustering on the motion vectors of vehicles, and the occluded vehicles are separated according to the binary classification of motion vectors. On the tracking level, occlusion layer images are adaptively constructed and maintained. The detected vehicles are tracked in the captured images and the occlusion layer images by performing a bidirectional occlusion reasoning algorithm.

Norbert Buch [7] developed a tracking system for vehicles in urban traffic scenes. The task of automatic video analysis for existing CCTV infrastructure is of increasing interest due to the benefits of behavior analysis for traffic control. Based on 3D wireframe models, a combined detector and classifier are used to locate the ground plane positions of vehicles. The system uses a Kalman filter with variable sample time to track vehicles on the ground plane.

III. VEHICLE TRACKING DEVELOPED SYSTEM

The proposed system consists of vehicle detection, classification, tracking, and behavior understanding. The framework of the developed system is illustrated in Fig.1.

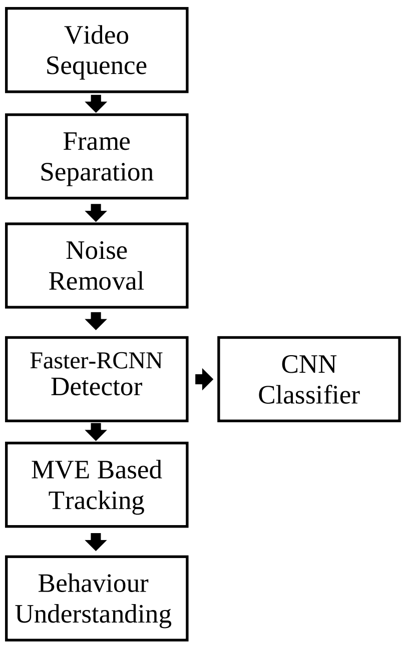

Fig. 1 Developed system framework

A. *Vehicle detection*

Vehicle detection is the key step in vehicle tracking and behavior understanding systems. After the video is separated into frames, some noise cancellation pre-processes are applied to remove natural noise like rain, dust, and fog in the open air; in addition, to handle some defects caused by the camera like blurred and noise [8], we used deep learning techniques to achieve a robust vehicle detection process, where the recent research indicates that deep learning techniques have achieved very promising results in computer vision filed [9].

We have used a vehicle dataset for different types and sizes of vehicles to train a Faster R-CNN object detector [10] for detecting vehicles. It uses CNN as a featured network to generate features from the input image. Then a region proposal network (RPN) is added to generate region proposals directly in the network. The RPN consists of a simple network with three convolutional layers; one common layer feeds into two layers, one for classification and the other for bounding box regression. RPN generates several bounding boxes called Region of Interests (ROIs) that have a high probability of containing any vehicles.

At last, an RCNN takes input from both the CNN feature network and RPN and generates the final class and bounding box. RCNN comprises four fully connected, two of them stacked common layers shared by a classification layer and a bounding box regression layer. To help it classify only the inside of the bounding boxes, the features are cropped according to the bounding boxes to generate the detected vehicle bounding boxes (DVBBS).

B. *Vehicle classification*

The vehicle classification algorithm aims to recognize various vehicle types, either small vehicles like cars and motorbikes or various large size vehicles like buses and trucks. A CNN [11] is used to train a dataset for different types of vehicles in terms of size. The CNN first layer defined the type and size of the detected vehicle; next, we defined the middle layers of the network. The middle layers are made up of repeated blocks of convolution layers, which are responsible for performing the image manipulation processes using the convolution operations to manipulate and extract the vehicle image features. Each convolution layer is followed by a cross-channel normalization layer, a rectified linear unit ReLU, and a pooling layer to transform the convolution operation results and make the CNN invariant to image translation and illumination. These layers form the core building blocks of CNN. The final layers are the classification layers which have four classes to represent different types of vehicles in terms of size; as shown in Fig.2, for the classification task, two fully connected neural network layers are used, but first, we have to resize the detected vehicle image to be equal to the unified size of the training images. These layers combine all the features learned by the previous layers across the image to identify the larger patterns.

Fig. 2 Samples of the training data set classes.

C. Vehicle Tracking

Vehicle tracking predicts vehicle positions in subsequent frames, match vehicles between adjacent frames, and ultimately obtains the trajectory and location for each frame in the camera field of view of the vehicle. The tracking method holds the vehicle trajectory by identifying motion dynamic attributes and characteristics to locate its position in every frame. The detected vehicles and their correspondence are jointly estimated by updating location iteratively using information obtained from previous frames. In our system, the vehicle detection process is performed in every frame.

After obtaining the DVBBs, a tracking algorithm is applied. When a new DVBB is discovered, it is placed in a three-frame buffer to dismiss static objects and noise and avoid the parked vehicle's problem [12]. If the same DVBB is successfully tracked over three frames, it must be removed from the buffer and entered into the active tracking list.

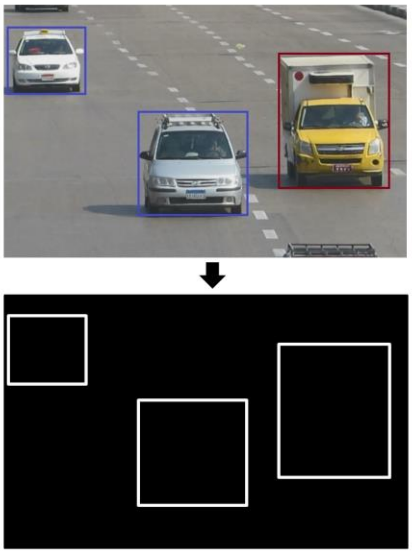

Fig. 3 DVBBs binary transfer process

First, we transfer all DVBBs to a binary image because the operations such as edge detection, noise and dilation removal, and object labeling are suitable in a binary platform, as shown in Fig.3. After labeling each DVBB and extracting relevant features pertaining to each DVBB, such as its center coordinates, height, width, and the pixels contained in the original frame (vehicle image).

For each DVBB in the current frame, a certain range around its centroid is considered a centroid search area CSA, where its boundaries are calculated as a 25% percent from its DVBB size. Fig.4 shows CSA. In the next frame, a 2D log MVE search algorithm, which depends on pixel-based motion representation [13], is applied, determining DVBBs centroids and checking if any of them are located in any of the CSAs.

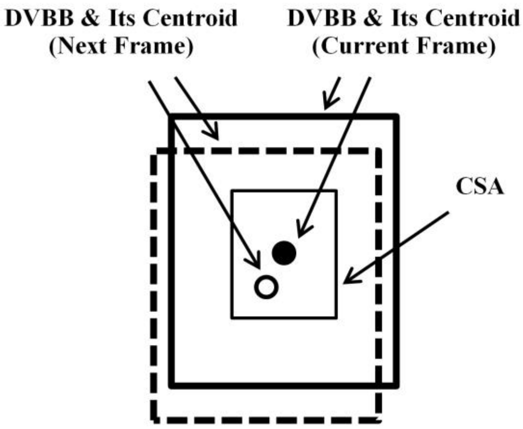

Fig. 4 Centroid search area

The search algorithm uses a logarithmic search method that starts from the position corresponding to zero displacements and tests five search points arranged in a diamond shape. It selects the block that yields minimum error and uses it to form a new search region around it. If the best matching point is the center point, it proceeds with a new search region with an offset of half the amount of the previous offset; otherwise, the offset remains the same. The process is repeated for successively smaller ranges until the offset equals 1.

We considered any DVBB in the next frame whose centroid falls within the CSA a possible candidate. If only one possible candidate is detected, we consider it a match and update the tracking data set for that DVBB with the new coordinates of an existing vehicle. If there is more than one possible candidate, we then compare object sizes, and the closest size to DVBB is considered a match. Finally, if a DVBB cannot be matched up at any given frame, then we predict the new DVBB position from the previous frames MVE, as we predict the new vehicle position in the next five frames. If the vehicle is still not detected, it has to be removed from the active tracking list. However, its history remains in the master object tracking database, which records each DVBB class, position, velocity, and shape.

D. Vehicle Behaviour Understanding

Vehicle behavior understanding is very important to automatically detect anomalies and traffic violations to serve traffic management by improving the flow of vehicle traffic and safety, where the vehicle behavior understanding depends on vehicle trajectories and vehicle speed.

A motion trajectory is identified by tracking a vehicle from one frame to the next and linking its positions in consecutive frames. And the speed of the vehicle in each frame is calculated using the vehicle's position in each frame. The DVBB centroid is important to understand the distance of the vehicle moving in adjacent frames. Therefore, the speed calculation becomes possible as the frame rate of captured moves is known.

After obtaining the vehicle trajectory and speed, it will be possible to detect anomalies in traffic such as illegal stop vehicles, converse driving, jaywalk, and illegal lane changing.

IV. SYSTEM RESULTS

The developed algorithm is applied to various vehicle datasets collected from real stationary road monitor cameras. Experiments have been performed to test the detection, classification, and tracking processes involved in the developed vehicle tracking system and measure its accuracy. The systems are designed in MATLAB R2017, the test videos were taken under various illumination and mobile conditions, and the results are as follows:

- 95% of the tested vehicles are precisely detected.
- 90% of the detected vehicles are successfully classified
- 92% of detected vehicles tracks are well generated

V. CONCLUSION

We developed a robust system to visually detect, classify and track vehicles based on deep machine learning methods combined with motion vector estimation technique. To recognize and track various vehicles, either small vehicles or various large size vehicles. The video frames are firstly separated, and then some noise cancellation pre-processes are applied. After that, a vehicle's data set is loaded to train a Faster R-CNN detector and CNN classifier to detect and classify vehicles. A vehicle tracking algorithm using the MVE technique determines the vehicle's directions and positions in the video frames. Finally, we demonstrate that traffic management can be performed by implementing the trajectories of the vehicles and the vehicles' speed calculations to understand the vehicle behavior.

REFERENCES

[1] World Health Organization, “Global status report on road safety 2018,” World Health Organization, 2018.
[2] B. Tian, et al., “Hierarchical and networked vehicle surveillance in ITS: a survey,” IEEE Transactions on Intelligent Transportation Systems, vol. 16, no. 2, pp. 557-580, 2015.
[3] R. Chandran, and N. Raman, “A review on video-based techniques for vehicle detection, tracking and behavior understanding,” International Journal of Advances in Computer and Electronics Engineering, vol. 2, no.5, pp. 07-13, 2017.
[4] K. Kale, S. Pawar, and P. Dhulekar, “Moving object tracking using optical flow and motion vector estimation,” in Proc. 2015 4th International Conference on Reliability, Infocom Technologies and Optimization (ICRITO), (Trends and Future Directions). IEEE, 2015.
[5] N. A. Mandellos, I. Keramitsoglou, and C. T. Kiranoudis, “A background subtraction algorithm for detecting and tracking vehicles,” Expert Systems with Applications, vol.38, no. 3, pp. 1619-1631, 2011.
[6] W. Zhang, et al., “Multilevel framework to detect and handle vehicle occlusion,” IEEE Transactions on Intelligent Transportation Systems, vol. 9, no. 1, pp. 161-174, 2008.
[7] N. Buch, et al., “Urban vehicle tracking using a combined 3D model detector and classifier,” in proc. International Conference on Knowledge-Based and Intelligent Information and Engineering Systems, Springer, Berlin, Heidelberg, 2009.
[8] 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.
[9] V. Bui, and L. Chang. “Deep learning architectures for hard character classification,” in Proc. on the International Conference on Artificial Intelligence (ICAI). The Steering Committee of The World Congress in Computer Science, Computer Engineering and Applied Computing (WorldCom), 2016.
[10] S. Ren, et al., “Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks,” *IEEE Transactions of Pattern Analysis and Machine Intelligence*. Vol. 39, no. 6, pp. 1137-1149, 2017.
[11] Y. LeCun, Y. Bengio, and G. Hinton. “Deep learning,” Nature 521.7553 (2015): 436.
[12] O. Marques, Practical image and video processing using MATLAB. John Wiley & Sons, 2011.
[13] M. Jakubowski, and G. Pastuszak. “Block-based motion estimation algorithms a survey,” Opto-Electronics Review, vol. 21, no. 1 pp. 86-102, 2013.
[14]