Improving On-Road Vehicle Detection Performance by Combining Detection and Tracking Techniques

Aryuanto Soetedjo
Department of Electrical Engineering
National Institute of Technology (ITN)
Malang, Indonesia
e-mail: aryuanto@gmail.com

I Komang Somawirata
Department of Electrical Engineering
National Institute of Technology (ITN)
Malang, Indonesia
e-mail: kmgsomawirata@yahoo.com

***Abstract*—This paper presents the combination method of detection and tracking to improve the recall of on-road vehicle detection. The vehicle detection method is a kind of the object detection where the vehicle in front of a car is detected using a camera installed on the car. In the proposed method, the Viola Jones detection and the support vector machine technique are combined complementary. Further the Lucas Kanade tracking technique is introduced to gain the true positive detection when both detection techniques fail to detect the vehicle. The experimental results show that the proposed method achieves the recall of 0.97, the precision of 0.96 and the frame rate of 25.54 fps (frames per second).**

Keywords-vehicle detection; Viola Jones; support vector machine; Lukas Kanade tracking

I. INTRODUCTION

Driver Assistance System (DAS) is one component in the Intelligent Transportation Systems (ITS). The DAS provides some information, such as: a) traffic sign recognition; b) driver fatigue detection; c) vehicle detection; d) pedestrian detection; e) road lane marking detection; etc. By providing such information, the DAS assists the driver to navigate the car safety.

This paper deals with the vehicle detection system that detects the vehicle in front of using a vision system [1, 2]. Generally the detection step consists of two stages [1], i.e. Hypothesis Generation (HG) and Hypothesis Verification (HV). In the HG, the area of vehicle is searched in an image. The HV is used to verify the presence of vehicle in the detected area. Some approaches combine both steps for detecting the vehicle [3-11].

Artificial neural networks were employed to detect vehicle [3, 4]. In [3], the deep neural network was trained with the images of rear view vehicles in the highway and urban areas. In [4], the Haar-like feature was employed as the image descriptor in the multilayer perceptron neural network. They showed that the performance of vehicle detection is affected by the number of neuron in the hidden layer.

The popular Viola Jones method (VJ) was proposed by [5, 6] to detect vehicle. It employed the Haar-like feature and the Adaboost classifier. In [5], they examined the vehicle detection performance by varying the window size and the scaling factor of the classifier. The experiments were conducted to detect the rear view vehicle in front of a car. In [6], the VJ method was employed to detect and count the vehicles from a CCTV camera installed in the top of traffic signal.

Histogram of oriented gradient (HOG) feature and support vector machine (SVM) was employed to detect vehicle [7, 12-14]. To reduce the computation time, the perspective window was employed to scan the image [7]. The SVM is applied to the sub-images inside the windows only. In [12, 13], the shadow under the rear vehicle was used in the HG step. The SVM was used for vehicle verification in the HV step. The method [14] proposed a region of interest in the sliding window for SVM classifier by considering the correlation between the vehicle size and distance.

The hybrid method of VJ and HOG-SVM was proposed by [8]. They proposed a switching method to select the VJ or HOG-SVM based on the detection speed for increasing the detection performance. The method was employed for detecting the vehicle from the unmanned aerial vehicle (UAV) images.

In [15], the VJ method was employed in the first step to find the region of interest of the vehicle candidate. In the second step, the HOG-SVM method was applied to find the precise vehicle in the region of interest defined previously. The detection accuracy and the computation time are improved by combining both methods.

In [16], both the Haar and HOG features were trained by the Adaboost classifier to detect the vehicle. The method was tested in many freeways and achieved the high true positive rate and the low false positive rate. The method was implemented on a Digital Signal Prosessor (DSP) module and achieved the fast computation time for real-time application.

The combination of VJ method and the optical flow algorithm was employed in [9] to detect the vehicle. The VJ is used to detect the vehicle that travels in the same direction. While the optical flow is used to detect the coming vehicle. The temporal and spatial context is employed to reduce the false positive.

The mean shift tracking was employed to track the detected vehicle [10]. At first the vehicle is detected using the SVM classifier. After detected, the vehicle is tracked for five seconds and verified it. When the verificaion is failed, the detection process is started, otherwise the tracking process is repeated.

The combination of Kalman filter and mean shift tracking was proposed in [11]. The mean shift algorithm is introduced to overcome the problem of detection error which sometimes occurs in the Kalman filter tracking. In addition, the algorithm uses the tracking list by comparing the tracked vehicles and the newly added vehicles.

As described previously, the combination method increases the detection performance. However selecting the best combination scheme is a challenging task, since the different combination schemes yield the different results. The combination of VJ and SVM proposed in [15] may fail to detect the vehicle when the vehicle is mis-detected in the first step (VJ method).

The combination of detection and tracking methods reduces the computation time effectively. However since the verification process, i.e. the detection is conducted in a specific time interval, e.g. five seconds [10], it may cause the error detection.

To overcome the above problems, we propose an approach to combine the VJ and HOG-SVM complementary. In each frame, at first the VJ is applied. When it fails to detect the vehicle, the HOG-SVM is then applied. Afterwards the Lukas Kanade tracking method (LKT) [17] is employed. The approach offers the advantage that the true positive detection will be increased by means of three methods, i.e. the VJ, the SVM, and LKT.

The rest of paper is organized as follows. The proposed method is presented in section 2. Section 3 discusses the experimental results. The conclusion is covered in Section 4.

II. Proposed Method

The flowchart of proposed method is depicted in Fig. 1. Instead of applying the VJ and the HOG-SVM in a cascade configuration [15], our proposed method applies them complimentary. In the sense that only one of them will be applied in each frame. Since the VJ method is faster than the HOG-SVM method, we apply the VJ method as the first attempt in each frame as show in Fig. 1.

When the VJ detects the vehicle, the *VJ_flag* is set to 1, otherwise set to 0. The *VJ_flag* is used in the LKT to decide the detected vehicle as described later. The detected vehicle will be shown in the image and the position is saved as the tracked points for LKT operation in the next frame.

When the vehicle is not detected by the VJ, the HOG-SVM method will be applied. Using this scheme, we may ensure that the mis-detected vehicle will be reduced or avoided. When the HOG-SVM detects the vehicle, the *SVM_flag* is set to 1, otherwise set to 0. Then the detected vehicle will be shown in the image and the position is saved as the tracked points, similar to the previous ones in the VJ method.

Although the detection process is provided by two detection methods, the vehicle may not be detected in several frames due to the blurred image that is caused by the vibration of the video camera during the image acquisition. Therefore we propose to employ the tracking method (LKT) to overcome such problem.

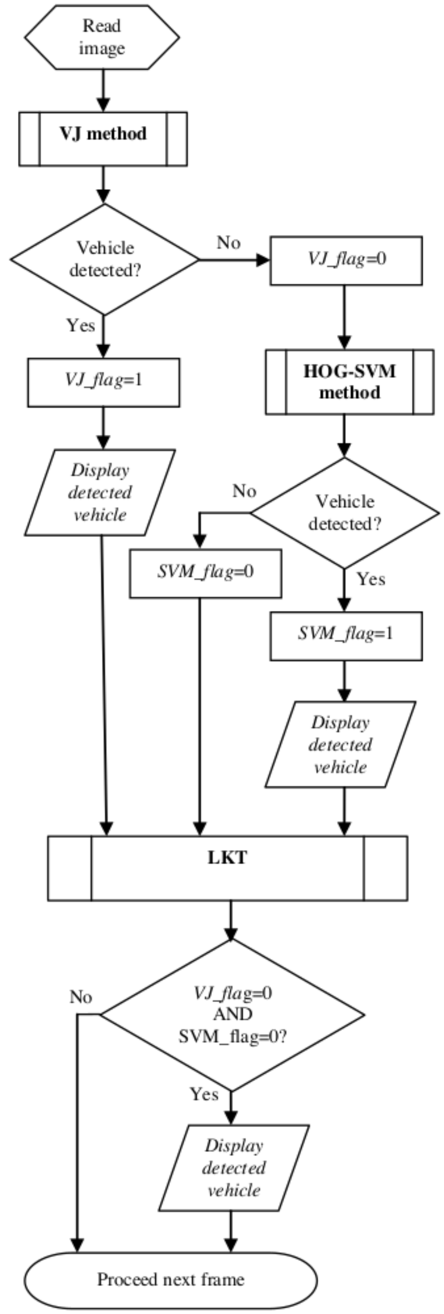

Figure 1. Flowchart of proposed method.

The LKT method [17] is used to track the position of detected vehicle from frame to next frame. The LKT method works as follows. Once the vehicle is detected (by the VJ or the HOG-SVM) in the current frame, the position of detected vehicle is used as the tracked points. In this work, the tracked points are defined as two points inside the rear body of vehicle which are defined as

$$PT1x=BULx+(0.25 \text{ x } WB) \quad (1)$$

$$PT1y=BULy+(0.25 \text{ x } HB) \quad (2)$$

$$PT2x=BULx+(0.75 \text{ x } WB) \quad (3)$$

$$PT2y=BULy+(0.75 \text{ x } HB) \quad (4)$$

where $PT1x$ and $PT1y$ are the x-coordinate and the y-coordinate of the first point respectively, $PT2x$ and $PT2y$ are the x-coordinate and the y-coordinate of the second point respectively, $BULx$ and $BULy$ are the x-coordinate and the y-coordinate of the upper left of detected vehicle's bounding box, WB and HB are the width and the height of detected vehicle's bounding box respectively.

The LKT will find the corresponding points in the current frame based on the position of tracked points in the previous frame. When the corresponding points are found, they are considered as the detected vehicle's bounding box if both the VJ and the HOG-SVM fail to detect the vehicle, i.e. when the *VJ_flag*=0 and the *SVM_flag*=0.

It is noted here that since the LKT requires the tracked points in the previous frame, the LKT is applied in each frame even though the VJ or the HOG-SVM detects the vehicle successfully on a particular frame. Fortunately the execution time of LKT method is very fast, thus it does not burden the computation time.

III. Experimental Results

The proposed algorithm is implemented on a PC Intel Core i7 3.4 Ghz using the C++ software language and OpenCV library. Several experiments are conducted to validate the proposed approach. There are ten video datasets taken from urban roads and expressway in Malang city, Indonesia. The videos are taken using a video camera installed on the dashboard of a car.

To compare our proposed method from the existing ones, we examine six methods, i.e. a) VJ only [5] (called as **VJ**); b) HOG-SVM only [12[(called as **SVM**); c) VJ and HOG-SVM [15] (called as **VJ+SVM**); d) VJ and LKT (called as **VJ+LKT**); e) HOG-SVM and LKT (called as **SVM+LKT**); f) Our proposed method (called as **PROP**).

In the experiments, the recall and the precision are computed which are expressed as

$$Recall=TP/(TP+FN) \quad (5)$$

$$Precision=TP/(TP+FP) \quad (6)$$

where TP is true positive, FN is false negative and FP is false positive. To examine the computation time, the frame rate is also calculated.

The experimental results are given in Table I. From the table it is obtained that the recall and precision of proposed method is the best among the other methods. The results show that by introducing the LKT, the recall is increased significantly, while the precision remains the same. It means that the vehicles which are not detected by the VJ or the SVM in several frames could be detected by the LKT. Thus the true positive (TP) will increase. Since $TP+FN$ in the experiment is fixed, thus according to (5) the recall will increase. Furthermore since the LKT tracks the points which are detected by the VJ or SVM, the false positive (FP) will increase. Thus according to (6), the precision does not change significantly.

TABLE I. Experimental Results

Method	Recall	Precision	Frame rate (fps)
VJ [5]	0.71	0,96	51.53
SVM [12]	0.50	0.89	9.73
VJ+SVM [15]	0.30	1.00	45.47
VJ+LKT	0.96	0.96	42.96
SVM+LKT	0.72	0.76	9.38
PROP	0.97	0.96	25.54

The recall of (**VJ+SVM**) is very low due to the fact that the recall of VJ is low and the cascaded scheme could not recover the missing vehicle as discussed previously. However since detected vehicle is further verified by the SVM, the precision of **VJ+SVM** is very high.

By examining the proposed method (**PROP**), the contribution of SVM method in the increment of recall is very small. It is caused by the fact that for the tested videos used in the experiments, the recall of **SVM** is low. However, the small increment of the recall of **PROP** from the one of **VJ+LKT** indicates that in general the proposed combination approach will provide a higher value of the recall.

The results of frame rates show that the **VJ** achieves the highest frame rate, while the **SVM** achieves the lowest one. The frame rate of **PROP** is 25.54 fps which is lower than **VJ+LKT**, however the value is enough for the real-time implementation.

Sample images of the detection results are illustrated in Fig. 2 and Fig. 3, where the pink, green, and blue rectangles represent the detected vehicle using the VJ, the SVM and the LKT respectively. As described previously, since the position of tracked points are inside the rear body of vehicle, the area of detected vehicle using LKT is smaller than the others.

Fig. 2 shows the detection results in the video dataset-1. As shown in the figures, the vehicle is detected by the SVM in frame-362. However both the VJ and SVM fail to detect it in the consecutive frame (frame-363). Fortunately the LKT is able to detect it in frame-363. In frame-364, the VJ detects the vehicle properly.

Fig. 3 shows the detection results in the video dataset-10. As shown in the figure, the VJ fails to detect the vehicle in frame-199. Fortunately the SVM detect it properly. In frame-200 and frame-201, the VJ and the SVM fail to detect the vehicles, however they are detected by the LKT.

(a) Frame=362

(b) Frame=363

(c) Frame=364

Figure 2. Detection results in video dataset-1: a) Frame=362; b) Frame=363; c) Frame-364.

(a) Frame=199

(b) Frame=200

(c) Frame=201

Figure 3. Detection results in video dataset-10: a) Frame=199; b) Frame=200; c) Frame-201.

IV. CONCLUSION

The combination method of the VJ, the SVM and the LKT is effective to increase the true positive detection in the on-road vehicle detection. The effectiveness of the method is indicated by the ability to detect the vehicles in the image sequences captured by a video camera which sometimes degrade in the quality. From the experiments using ten video datasets, the combination of the VJ and the LKT contributes the signification result. While the contribution of the HOG-SVM is very small in these particular datasets.

In future some improvements will be made in the detection and tracking techniques. Further the real-time implementation will be realized.

ACKNOWLEDGMENT

This work is supported by the Research Grant, Competence-Based Research scheme from Directorate General of Higher Education, Ministry of Research and Technology and Higher Education, Republic of Indonesia (SP DIPA. 042.06.1.401516/2018). Authors would like to thank people at Control System Laboratory, National Institute of Technology Malang, Indonesia for preparing the video data.

REFERENCES

[1] Zehang Sun, G. Bebis and R. Miller, "On-road vehicle detection: a review," IEEE Transactions on Pattern Analysis and Machine

Intelligence, vol. 28, no. 5, May 2006, pp. 694-711, doi: 10.1109/TPAMI.2006.104.

[2] S. Sivaraman and M. M. Trivedi, "Looking at Vehicles on the Road: A Survey of Vision-Based Vehicle Detection, Tracking, and Behavior Analysis," IEEE Transactions on Intelligent Transportation Systems, vol. 14, no. 4, Dec. 2013, pp. 1773-1795, doi: 10.1109/TITS.2013.2266661.

[3] Z. Zhang, C. Xu and W. Feng, "Road vehicle detection and classification based on Deep Neural Network," Proc. 7th IEEE International Conference on Software Engineering and Service Science (ICSESS), Aug. 2016, pp. 675-678, doi: 10.1109/ICSESS.2016.7883158.

[4] A. Mohamed, A. Issam, B. Mohamed and B. Abdellatif, "Real-time Detection of Vehicles Using the Haar-like Features and Artificial Neuron Networks," Procedia Computer Science, vol. 73, Dec. 2015, pp. 24-31, doi: 10.1016/j.procs.2015.12.044.

[5] M. Oliveira and V. Santos, "Automatic Detection of Cars in Real Roads using Haar-like Features," Proc. 8th Portuguese Conference on Automatic Control (CONTROLO2008), Jul. 2008.

[6] S. Choudhury, S. P. Chattopadhyay and T. K. Hazra, "Vehicle detection and counting using haar feature-based classifier," Proc. 8th Annual Industrial Automation and Electromechanical Engineering Conference (IEMECON), Aug. 2017, pp. 106-109, doi: 10.1109/IEMECON.2017.8079571.

[7] S.G.S. Filho, R.Z. Freire, L.S. Coelho, "Feature Extraction for On-Road Vehicle Detection Based on Support Vector Machine," Proc. European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning, Apr. 2017, pp. 65-70.

[8] Y. Xu, G. Yu, Y. Wang, X. Wu and Y. Ma, "A Hybrid Vehicle Detection Method Based on Viola-Jones and HOG + SVM from UAV Images," Sensors, vol. 16, no. 8, Aug. 2016, pp. 1-23, doi:10.3390/s16081325.

[9] J. Choi, "Realtime On-Road Vehicle Detection with Optical Flows and Haar-Like Feature Detectors," unpublished.

[10] W. Xu, M. Qiu, Z. Chen and H. Su, "Intelligent Vehicle Detection and Tracking for Highway Driving," Proc. IEEE International Conference on Multimedia and Expo Workshops, Jul. 2012, pp. 67-72, doi: 10.1109/ICMEW.2012.19.

[11] J. W. Baek, B. G. Han, H. Kang, Y. Chung and S. I. Lee, "Fast and reliable tracking algorithm for on-road vehicle detection systems," Proc. Eighth International Conference on Ubiquitous and Future Networks (ICUFN), Jul. 2016, pp. 70-72, doi: 10.1109/ICUFN.2016.7536983.

[12] X. Li and X. Guo, "A HOG Feature and SVM Based Method for Forward Vehicle Detection with Single Camera," Proc. 5th International Conference on Intelligent Human-Machine Systems and Cybernetics, Aug. 2013, pp. 263-266, doi: 10.1109/IHMSC.2013.69.

[13] T. Y. Lee, J. S. Oh and J. H. Kim, "On-road vehicle detection based on appearance features for autonomous vehicles," Proc. 15th International Conference on Control, Automation and Systems (ICCAS), Oct. 2015, pp. 1720-1723, doi: 10.1109/ICCAS.2015.7364641.

[14] J. Kim, J. Baek, D. Y. Kim and E. Kim, "On-road vehicle detection based on effective hypothesis generation," Proc. IEEE RO-MAN, Gyeongju, Aug. 2013, pp. 252-257, doi: 10.1109/ROMAN.2013.6628455.

[15] Y. Wei, Q. Tian, J. Guo, W. Huang and J. Cao, "Multi-vehicle detection algorithm through combining Harr and HOG features," Mathematics and Computers in Simulation, Feb. 2018, In press, doi:10.1016/j.matcom.2017.12.011.

[16] H. Wang and Y. Cai, "Monocular based road vehicle detection with feature fusion and cascaded Adaboost algorithm," Optik-International Journal for Light and Electron Optics, vol. 126, issue 22, Nov. 2015, pp. 3329-3334, doi: 10.1016/j.ijleo.2015.08.018.

[17] B.D. Lucas and Kanade, "An Iterative Image Registration Technique with an Application to Stereo Vision, Proc. DARPA Image Understanding Workshop, Aug. 1981, pp. 121-130.

Aryu_Proceeding_2018-2.pdf

ORIGINALITY REPORT

4%	5%	4%	%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	www.ijicic.org Internet Source	3%
2	Aryuanto Soetedjo, I Komang Somawirata. "Implementation of face detection and tracking on a low cost embedded system using fusion technique", 2016 11th International Conference on Computer Science & Education (ICCSE), 2016 Publication	2%

Exclude quotes	On	Exclude matches	< 2%
Exclude bibliography	On		