Vehicle counting via car parts detection from an in-vehicle camera image

Fumito Shinmura[1], Yasutomo Kawanishi[2], Daisuke Deguchi[3], Ichiro Ide[2], Hiroshi Murase[2]

[1]: Institutes of Innovations for Future Society (MIRAI), Nagoya University, Japan
[2]: Graduate School of Informatics, Nagoya University, Japan
[3]: Information Strategy Office, Nagoya University, Japan
E-mail: shinmuraf@murase.is.i.nagoya-u.ac.jp

Abstract

This paper proposes a method to count vehicles from an in-vehicle camera image by regression based on car parts detection. In the case of an in-vehicle camera image, since vehicles are frequently occluded by other vehicles in traffic congestion, it is difficult to accurately count vehicles. Therefore, we propose a method to count vehicles by regression based on the number of visible car parts. For this, we make an estimator by learning the relation between the number of visible car parts and that of vehicles by Support Vector Regression. We evaluated our method using in-vehicle camera images recorded in an actual environment, where the proposed method performed better than counting detected vehicles.

Keywords perception for vehicle, vehicle counting, in-vehicle camera, computer vision

1. INTRODUCTION

For safe driving assistance and autonomous driving, demand for the sensing and perception technologies using an in-vehicle camera is increasing. In order to avoid traffic accidents, it is important to detect pedestrians and vehicles from an in-vehicle camera. In addition, since driving behavior should be changed according to traffic condition, it is also important to grasp the traffic condition from an in-vehicle camera. Since the distance between vehicles becomes shorter in traffic congestion, the density of surrounding vehicles increases according to the degree of congestion. Accordingly, the information on the number of surrounding vehicles will help the system to grasp the degree of traffic congestion. Therefore, this paper proposes a method to count vehicles from an in-vehicle camera image.

In general, the object counting problem includes counting of objects passing a certain place and counting of objects existing in a scene. This paper focuses on counting objects in a scene since we aim to count vehicles from an in-vehicle camera image. The solution for this kind of object counting is mainly classified into three types: First, there are methods based on object detection. Object detection methods have been studied by many researchers [1, 2, 3]. Detection methods based on deep neural network have been widely studied in recent years [2, 3]. Focusing on the vehicle counting problem, Salvi proposes a counting method by extracting small regions with motion and detecting vehicle regions by clustering and tracking the regions [4]. Methods based on background subtraction have also been proposed largely [5, 6].

Second, there are methods based on clustering. A conventional approach is to cluster motion features such as optical flow [7, 8]. It counts the number of clusters since one cluster corresponds to one object. These methods are effective in a congested situation where object detection is difficult due to occlusion. However, background removal is required for clustering, so these methods are mainly used for counting from a fixed camera installed on the street.

Third, there are methods based on regression. These methods estimate the number ob objects by regression based on the relation between the number of objects and image features [9, 10, 11, 12]. They are also effective in a congested situation same as the method based on clustering. Since it is difficult to design a model for calculating the number of objects from image features, many researchers have tried to learn their relation by statistical learning theory. Conte et al. proposed a regression method that estimates the number of people from the number of feature points and density of SURF features using Support Vector Regression (SVR)[10]. Bansal et al. combined the estimation from SIFT features and head detection, and estimated the number of people by SVR [11].

For vehicle counting from an in-vehicle camera image, methods using background subtraction and motion features are not effective since camera moves and background changes variously. In contrast with these methods, counting methods based on detection are effective since they are not affected by camera movements. Therefore, methods such as those proposed in references [1, 2] are usually used for counting from an in-vehicle camera image. However, in the case of an in-vehicle camera image, a large part of a vehicle is often occluded by other vehicles in traffic congestion, as shown in Fig. 1. Thus, the detection accuracy decreases, and it is difficult to correctly count the number of vehicles.

In the proposed method, we combine detection and regression approaches. Even if a large part of a vehicle is occluded, some car parts are usually visible. In the case of an in-vehicle camera image, the camera usually records leading ve-

Fig. 1. Example of an in-vehicle camera image.

hicles and oncoming vehicles from similar view points. There is a correlation between the number of visible car parts and the number of vehicles. We consider that we can predict the number of vehicles based on these visible car parts. Therefore, the proposed method detects some kinds of car parts and estimates the number of vehicles by regression based on the number of the detected car parts. By learning the relation between the number of vehicles and the number of visible car parts using a statistical learning method, we will be able to estimate the number of vehicles by regression.

Our contributions are summarized as follows:

- We propose a method to count vehicles from an in-vehicle camera image based on car parts detection.
- We show the effectiveness of regressing the number of vehicles from the number of visible car parts.

2. VEHICLE COUNTING VIA CAR PARTS DETECTION

The proposed method counts vehicles in an in-vehicle image. As mentioned earlier, since the detection accuracy sometimes decreases because of large occlusion, we take a regression approach based on the number of visible car parts.

The process-flow of the method is outlined in Fig. 2. It mainly consists of the car parts detection process and the vehicle number estimation process using an estimator. The details of each process are described in the sections below.

2.1. Detection of car parts

The parts to be detected are better to be common for every vehicle and visible in traffic congestion as much as possible. Accordingly, we decided as detection targets, the following seven kinds of car parts; license plate, window (including windshield and rear window), tire, right headlight, left headlight, right tail-light, and left tail-light. In the following experiment, Redmon's object detection method (YOLOv3) [3] is used for the car parts detection. We train the network to detect each car part. In the car parts detection step, the constructed detector is applied to an input image. We then count the detected car parts and obtain the number of each car part in a scene.

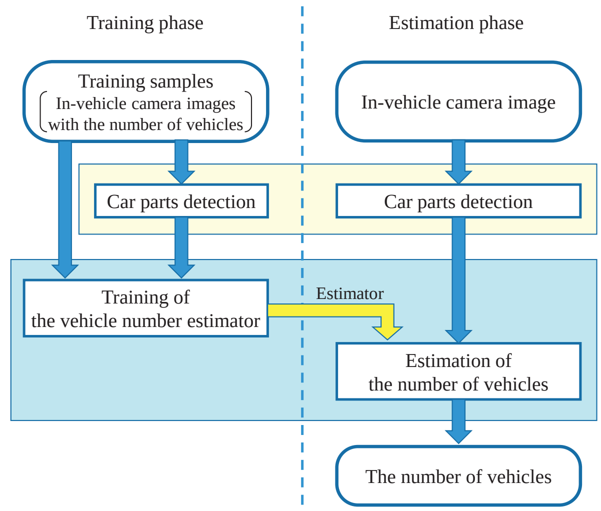

Fig. 2. Process-flow of the proposed method.

2.2. Estimation of the number of vehicles

The proposed method estimates the number of vehicles from the number of car parts. As shown in Fig. 3, sometimes the car parts are visible, but sometimes they are not. Since there are various combinations between the number of vehicles and that of car parts, we take a regression approach using Support Vector Regression (SVR).

In the training phase, we prepare in-vehicle camera images annotated with the number of vehicles as training samples. The car parts detector introduced in 2.1 is applied to these training samples, and the number of each car part is counted. Next, we construct a vehicle number estimator by learning the relation between the number of vehicles and that of car parts by SVR. We use the radial basis function (RBF) kernel as the kernel function of the SVR.

In the estimation phase, the proposed method detects car parts from an input in-vehicle camera image using the car parts detector. Next, the number of vehicles is estimated from the number of detected car parts using the constructed estimator.

3. EXPERIMENT

In order to evaluate the effectiveness of the proposed method, we conducted a vehicle counting experiment.

Name	# of
License plate	2
Window (rear)	3
Tire	6
Right tail-light	4
Left tail-light	1
Vehicle	4

Name	# of
License plate	4
Window (rear)	4
Tire	8
Right tail-light	4
Left tail-light	2
Vehicle	4

Fig. 3. Number of vehicles and car parts.

Fig. 4. Samples from the dataset.

3.1. Dataset

We collected in-vehicle camera videos by driving a vehicle on a public street in Nagoya, Japan. The vehicle for data collection was equipped with a commercially available video camera (Canon iVIS HF G20), and we drove it on the same route twice. The videos were recorded in a frame rate of 24 fps, and in a resolution of 1,920×1,080 pixels. We constructed a dataset by extracting one image for every 120 frames from the video. The dataset included 892 in-vehicle camera images in total. Samples from the dataset are shown in Fig. 4.

3.2. Experimental result

In the experiment, we used one sequence for training and another sequence for testing. We conducted experiments twice by exchanging training and test sequences, and evaluated the average accuracy (i.e. two-fold cross-validation). The evaluation criterion was the mean absolute error (MAE).

We used a counting method by vehicle detection using YOLOv3 [3] as a comparative method. For this, we constructed a vehicle detector by making the network of YOLOv3 learn features of vehicles.

Fig. 5. Example of car parts detection result.

Fig. 6. Example of vehicle detection result.

Table 1. Experimental result.

	Mean Absolute Error		
	Seq. 1	Seq. 2	Total
Proposed method	1.16	0.81	0.98
Comparative method	1.41	0.78	1.09

First, we show examples of the results of car parts detection and vehicle detection in Figs. 5 and 6, respectively. From these results, we can see that vehicles with large occlusions were difficult to be detected while detection of car parts was possible.

The experimental result is shown in Table 1. The proposed method performed better in total.

Although the proposed method is effective for large occlusions of vehicles, it has little advantage when there is no occlusion of vehicles. Actually, the vehicle detection based method sometimes performs better. We show the results when there are more than three vehicles and less than two vehicles in a scene in Tables 2 and 3, respectively. As mentioned above, the proposed method performs better when there are more than three vehicles in a scene, where the vehicles are frequently occluded. Therefore, we could confirm that the proposed method is effective against the occlusion of vehicles. On the other hand, the proposed method performs equally or

Table 2. Experimental result when there are more than three vehicles in an in-vehicle image.

	Mean Absolute Error		
	Seq. 1	Seq. 2	Total
Proposed method	2.64	1.52	2.08
Comparative method	3.40	2.05	2.72

Table 3. Experimental result when there are less than two vehicles in an in-vehicle image.

	Mean Absolute Error		
	Seq. 1	Seq. 2	Total
Proposed method	0.44	0.52	0.48
Comparative method	0.43	0.27	0.35

worse when there are less than two vehicles in a scene where the occlusion of vehicles hardly occurs.

4. CONCLUSION

We proposed a method to count vehicles in a scene based on car parts detection. Considering that the camera is not fixed and that occlusion is serious in case of in-vehicle camera images, we proposed a method that detects car parts and regresses the number of vehicles from the number of the detected car parts.

Future works include automatic selection of target car parts to detect, which is currently selected manually. We expect that the accuracy of counting can be increased by automatically selecting the most appropriate target car parts based on statistical learning theory. In addition, the performance improvement when no occlusion will be one of our future works. For this, we are considering to combine with the vehicle detection method.

5. REFERENCES

[1] P. F. Felzenszwalb, R. B. Girshick, D. McAllester and D. Ramanan, "Object Detection with Discriminatively Trained Part-Based Models", IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol.32, No.9, pp.1627–1645, 2010.

[2] S. Ren, K. He, R. Girshick and Jian Sun, "Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks", Advances in Neural Information Processing Systems 28, pp.91–99, 2015.

[3] J. Redmon and A. Farhadi, "YOLOv3: An Incremental Improvement", Computer Research Repository, arXiv:1804.02767, 2018.

[4] G. Salvi, "An Automated Vehicle Counting System Based on Blob Analysis for Traffic Surveillance", In Proceedings of 2012 International Conference on Image Processing, Computer Vision, and Pattern Recognition, pp.397–402, 2012.

[5] E. Bas, A.M. Tekalp and F.S. Salman, "Automatic Vehicle Counting from Video for Traffic Flow Analysis", In Proceedings of 2007 IEEE Intelligent Vehicles Symposium, pp.392–397, 2007.

[6] L. Unzueta, M. Nieto, A. Cortes, J. Barandiaran, O. Otaegui and P. Sanchez, "Adaptive Multicue Background Subtraction for Robust Vehicle Counting and Classification", IEEE Transactions on Intelligent Transportation Systems, Vol. 13, No. 2, pp.527–540, 2012.

[7] V. Rabaud and S. Belongie, "Counting Crowded Moving Objects", In Proceedings of 2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, pp.705–711, 2006.

[8] S. Fujisawa, G. Hasegawa, Y. Taniguchi and H. Nakano, "Pedestrian Counting in Video Sequences based on Optical Flow Clustering", International Journal of Image Processing, Vol. 7, No. 1, pp.1–16, 2013.

[9] A. Albiol, M.J. Silla, A. Albiol and J.M. Mossi, "Video Analysis using Corner Motion Statistics", In Proceedings of IEEE International Workshop on Performance Evaluation of Tracking and Surveillance 2009, pp.31–37, 2009.

[10] D. Conte, P. Foggia, G. Percannella and F. Tufano, "A Method for Counting People in Crowded Scenes", In Proceedings of the 7th IEEE International Conference on Advanced Video and Signal Based Surveillance, pp.225–232, 2010.

[11] A. Bansal and K.S. Venkatesh, "People Counting in High Density Crowds from Still Images", International Journal of Computer and Electrical Engineering, Vol. 7, No. 5, pp. 316–324, 2015.

[12] V.-Q. Pham, T. Kozakaya, O. Yamaguchi and R. Okada, "COUNT Forest: Co-voting Uncertain Number of Targets using Random Forest for Crowd Density Estimation", In Proceedings of 2015 IEEE International Conference on Computer Vision, pp.3253–3261, 2015.