

# A Method for Detecting Street Parking Using Dashboard Camera Videos

Akihiro Matsuda,<sup>1\*</sup> Tomokazu Matsui,<sup>1</sup> Yuki Matsuda,<sup>1,2</sup>  
Hirohiko Suwa,<sup>1,2</sup> and Keiichi Yasumoto<sup>1,2</sup>

<sup>1</sup>Graduate School of Institute Science, Nara Institute of Science and Technology,  
8916-5 Takayama-cho, Ikoma, Nara 630-0912, Japan

<sup>2</sup>RIKEN, Center for Advanced Intelligence Project, 1-4-1 Nihonbashi, Chuo-ku, Tokyo 103-0027, Japan

(Received July 29, 2020; accepted October 21, 2020)

**Keywords:** street parking, dashboard camera, city sensing, object detection, machine learning

In recent years, street parking in prohibited areas has become a social problem, especially in urban and tourist areas. In addition, because street parking can cause traffic congestion and accidents, real-time detection is required. The detection of street parking has been previously implemented on the basis of comparisons of videos recorded by fixed-point cameras. However, this approach has a limited detection area and low accuracy. To overcome these problems, this study aims towards a real-time street parking detection system that uses dashboard camera videos. We propose a machine learning method based on the characteristics of on-street parked vehicles derived by transforming images into text. The object detection model YOLOv3 was used to analyze videos. We created a dataset based on the coordinate information of 1765 vehicles and the recording vehicle information. We also created a model using random forest and logistic regression algorithms and evaluated it using the holdout and stratified 5-fold validation methods. F-measure values of up to 92% and 89% were obtained for the two types of model, respectively. These results confirm the effectiveness of the proposed street parking detection method based on bounding boxes and recording vehicle data.

## 1. Introduction

In recent years, street parking in prohibited areas has become a social problem. According to a survey on street parking by the Traffic Bureau of the National Police Agency (Japan) in 2019, the number of instantaneous on-street parked vehicles in prohibited areas in Tokyo's special wards was about 52,700.<sup>(1)</sup> Illegal street parking is not only a traffic problem in itself, but can also be the cause of various other traffic problems including traffic jams and rear-end collisions.

To deal with street parking and its related problems, the extension of restricted parking zones and the implementation of time-limited zones are being considered. However, these measures are not a fundamental solution and may complicate existing problems. In addition, they require large-scale cooperation from the government and administrative agencies, which is very costly and time-consuming.

---

\*Corresponding author: e-mail: matsuda.akihiro.lr2@is.naist.jp  
<https://doi.org/10.18494/SAM.2021.2998>

A system that uses multiple fixed-point cameras installed in the city to detect street parking was previously proposed.<sup>(2,3)</sup> However, fixed-point cameras cover only a limited area.

In this study, we use videos from dashboard cameras, which are commonly installed in vehicles, for real-time street parking detection.<sup>(4)</sup> Dashboard camera videos contain much information. In addition, they are not limited to a certain area and can thus collect traffic information more efficiently than conventional methods. However, uploading dashboard camera videos to a remote server for analysis requires a robust communication network, which increases costs and makes real-time processing difficult. In addition, unlike fixed-point camera videos, dashboard camera videos are influenced by vehicle movement and the surrounding environment. Therefore, to reduce the communication load and cost, a device installed in the vehicle is used to detect street parking in real time, and only the results of street parking location information are uploaded. We display street parking based on the location information.

In this study, we aim to develop a method for realizing the real-time detection of street parking, wherein a device with low computational capacity mounted in a vehicle can be used even in a complicated environment. Specifically, at the first step, we use an existing object detection model for dashboard camera videos to detect street parking. The coordinate information of the bounding box (BB) given at the recording time is used as text data, and the detected vehicle frame-in to frame-out is stored. In addition, the vehicle information (e.g., velocity and acceleration) at the time of detection is also stored. Then, based on the basis of the constructed data and the videos with BBs, we label a vehicle as being either an on-street parked vehicle or a moving vehicle. Also, with the labeled data, a machine learning method is used to detect street parking.

To verify the effectiveness of the proposed method, we constructed a text dataset of 1765 vehicles. In addition, we constructed an imbalanced dataset, a sampled dataset, and an additional validation dataset from this dataset. We trained the model with random forest (RF) and logistic regression (LR) machine learning algorithms using the dataset. We verified the trained model and confirmed the maximum value of recognition accuracy in terms of the F-measure score. For the imbalanced dataset, LR holdout testing showed 59% accuracy, for the balanced dataset, RF 5-fold testing showed 92% accuracy, and for the additional validation dataset, LR-sampled testing showed 83% accuracy.

The rest of this paper is organized as follows. In Sect. 2, related research on the detection of street parking is reviewed and the problems of existing methods are summarized. The flow of the proposed detection method for street parking is described in Sect. 3. In Sect. 4, evaluation experiments for street parking detection using a machine learning method are described. In Sect. 5, the results of the evaluation experiment are presented. In Sect. 6, the results are discussed. Finally, in Sect. 7, the conclusions of this study and future prospects are given.

## 2. Related Research

In this section, we review related research and existing technologies regarding street parking. Although many studies on street parking exist, there are few studies on detecting street parking.<sup>(5–9)</sup> Kyutoku *et al.* detected the vehicle in front using the difference from past on-board camera

images.<sup>(10)</sup> They used a single-lane road as their detection area to ensure a high correlation between past and present time-matched frames. The area to the right of the center line was excluded from the study, and a comparison was made for each window. As an indicator of the difference between corresponding windows, the horizontal edge information of the vehicle was used. With this method, the precision score and recall score were up to 82 and 71% in a cloudy environment, and up to 53 and 63% in a sunny environment, respectively. However, the average precision score and recall score were around only 50% and the recognition accuracy was greatly affected by weather conditions. In addition, there are many limitations, such as the detection of only objects in a certain lane area.

Kyutoku *et al.* also proposed a method for the detection of unspecified obstacles using spatiotemporal differences from past on-board camera images.<sup>(11)</sup> In the frames corresponding to the present and the past, the detection area is limited to the road surface, which is assumed to be a plane surface considering the gap between the photographic viewpoints. The angle between the RGB vectors, which is robust to changes in luminance, is used as an indicator of the difference to improve detection accuracy. However, there are problems regarding the alignment of the road surface in terms of time and accuracy, limited detection area, and travel lanes. In addition, it is difficult to verify the difference between the past images and the present images from the on-board camera when there are obstacles, such as on-street parked vehicles in the past images.

Xie *et al.* proposed a detection system for illegal parking using the Single Shot MultiBox Detector (SSD) algorithm.<sup>(12)</sup> The SSD is used to detect and track vehicles in the region of interest and count the number of stopped objects. However, this system uses fixed-point camera images and the detection area is limited. In addition, it is difficult to judge whether a vehicle is on-street parked or waiting at a light (all stopped vehicles are counted). The above methods have limited detection areas and detection conditions, making them suboptimal because, in practical applications, on-street parked vehicles are spread out throughout the city and conditions vary. In this paper, we propose a method for lane- and past-information-independent detection that uses dashboard camera videos.

### 3. Proposed Street Parking Detection Method

In this section, we describe the proposed method based on text data for detecting whether a vehicle is parked on the street.

#### 3.1 Outline of proposed method

In this work, we develop a model for detecting street parking in an urban environment. The proposed method takes advantage of the different characteristics between on-street parked vehicles and moving vehicles. The determination of on-street parking is made using machine learning based on the acquired datasets that contain vehicle information and recording vehicle information.

Figure 1 shows the flow of the analysis of dashboard camera videos and the detection model for street parking. The inputs are dashboard camera videos and the corresponding data of the

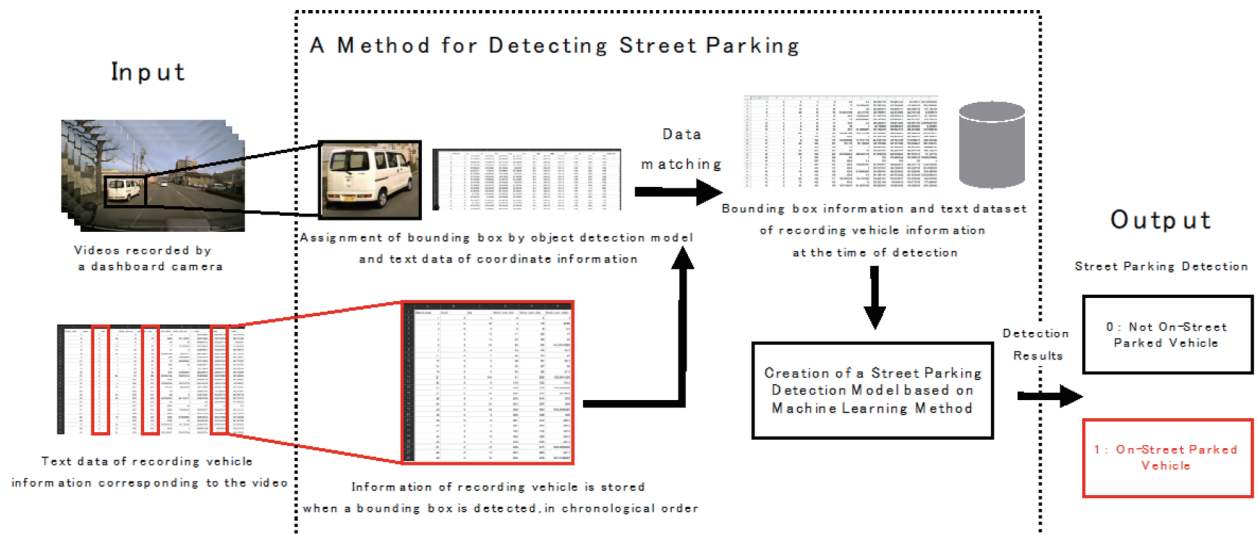


Fig. 1. (Color online) Flow of constructing a street parking detection model.

recording vehicle. The outputs are the decision result for a detected vehicle. In this paper, the BB information of vehicles is acquired from videos in time series and text data are generated for each video. In addition, the vehicle information is combined with the BB information based on the time series. A machine-learning-based approach is applied to the combined dataset to determine whether a vehicle is parked on the street.

If the effectiveness of the proposed method is confirmed, we will develop a system using lightweight, on-board devices such as Raspberry Pi and Jetson for real-time detection. We aim to display the on-street parking position by acquiring the location information of a vehicle detected as being street parked and uploading only the results of the location information.

### 3.2 Target data

As the target data, we used dashboard camera videos and recording vehicle data. The dashboard camera videos were provided by a taxi company in Kyoto Prefecture. The video resolution was  $1280 \times 720$  pixels and the frame rate was 14 frames per second (fps). The videos were recorded by seven taxis. The angle of view of the dashboard camera was constant. Recording started when the engine was turned on and ended when it was turned off. The taken videos are divided into 2 min files and stored. The recording area is mainly in the urban area of Kyoto Prefecture, and includes sightseeing spots and streets in Kyoto City. The videos were taken during daylight hours between March 15 and 27, 2019. The weather conditions were sunny or cloudy. The video data for one to three lanes in each direction were used.

A total of 12 videos (total length: about 24 min) taken by three taxis were analyzed. In addition, to reduce the processing load, the video resolution was changed from  $1280 \times 720$  pixels to  $512 \times 288$  pixels (the 16:9 aspect ratio was maintained).

We used the data of the velocity and acceleration of the vehicle during the recording. Acceleration in the left-right ( $X$ ), front-back ( $Y$ ), and vertical ( $Z$ ) directions relative to the

dashboard camera mounted inside the taxi was positive toward the right, front, and top, respectively. Because the sampling rate of the vehicle data was high, resampling was conducted to match the video. Velocity and acceleration in the  $X$ ,  $Y$ , and  $Z$  directions were selected as features.

### 3.3 Feature extraction

In this section, we describe the features to be extracted from the target data.

#### 3.3.1 BB information

Using the object detection model, a BB is assigned to the videos and the coordinate information is acquired. Figure 2 shows an image with BB information. Figure 3 shows the coordinate information to be obtained. The detected vehicles are numbered in the order of their detection. Tracking is conducted up to a certain range. The vehicles enclosed by BBs in Fig. 2 are the 98th and 112th vehicles. Each of these numbers (i.e., 98th and 112th) holds coordinate information corresponding to the frame. The coordinate information is tracked for a single vehicle and acquired from frame-in to frame-out of the video. The coordinate information is the  $x$ - and  $y$ -coordinates of the upper left and lower right of the BB when viewed from the front. In addition to these four coordinates, the number of detected frames was selected as a feature.

We selected the coordinate information as a feature for several reasons. The most significant reason is that on-street parked vehicles have different characteristics from those of moving vehicles. Regarding the coordinate information of the BBs, the BBs of on-street parked vehicles will increase in size as the recording vehicle approaches. In the frame before frame-out, the size of the BB is maximized. The BB is framed out in the next frame. In contrast, for moving vehicles, the size of the BB varies depending on the situation. We focused on this change in the coordinate information and used it as a feature. The characteristics of on-street parked vehicles were defined in detail in a previous study.<sup>(13)</sup>

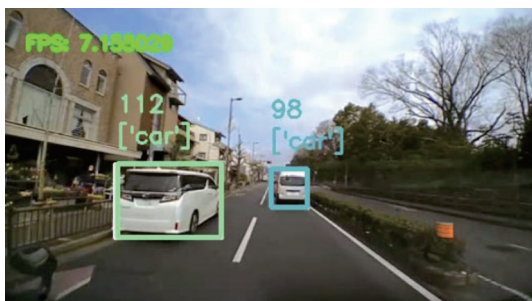


Fig. 2. (Color online) Vehicles with corresponding BBs.

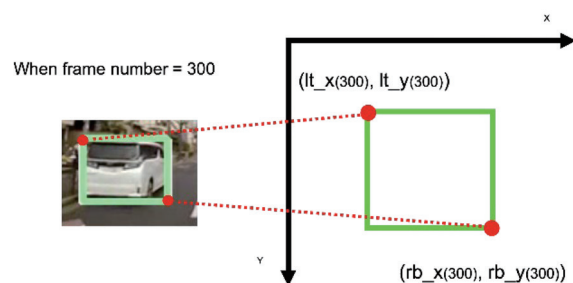


Fig. 3. (Color online) Coordinate information acquired from a BB.

### 3.3.2 Information of recording vehicle

The date and constant values were removed from the vehicle data and classified. Velocity and acceleration in the  $X$ ,  $Y$ , and  $Z$  directions were selected as features. These features were selected because we focused on the coordinate information of the BBs and the amount of the increase in the coordinate information is considered to be caused by the velocity and acceleration of the vehicle. In addition, to correlate the coordinate information with the recording vehicle data in a time series, the frame number of the coordinate information of a BB and the frame number of the recording vehicle data at that time were used.

### 3.3.3 Additional features

The width, height, and area of the BB were also calculated and added as features to the four coordinates. Table 1 shows the features and the statistics (average and standard deviation) of each feature used in this study. The maximum, minimum, average, and variance were calculated for the seven categories related to the BB and the four categories related to the recording vehicle data. The number of detected frames was also used as a feature, for a total of 45 features. These features are stored for each vehicle. Each feature is input as numerical data to the classification model.

## 3.4 Object detection model

A video is input to an object recognition model and the BB information is acquired. In this work, we selected YOLOv3 (You Only Look Once) as the object recognition model.<sup>(14,15)</sup> YOLO is widely used in the field of object recognition as a low-cost algorithm with high recognition accuracy and low computational cost. YOLO divides the whole image into grids and finds the object class and BB for each segmented region. Therefore, the whole image is recognized only once without searching for region candidates. All processes, from object detection to classification, are performed by a convolutional neural network, which reduces the computation time and enables real-time detection. In this paper, we used the COCO dataset as training data.<sup>(16)</sup>

In addition, the Kalman filter is used for tracking the recognized objects. The intersection over union between the BB of a vehicle detected by YOLO and that of the same vehicle in the next frame calculated by the Kalman filter was calculated. Certain thresholds were set to correspond to the movements of the recording and detected vehicles.

## 3.5 Video analysis environment

Table 2 shows the analysis environment used for the videos. For video analysis, we used the GPU environment of Google Colaboratory. For real-time detection, it is necessary to shorten the processing time of the video analysis used to judge whether a vehicle is on-street parked. To confirm the effectiveness, we analyzed the video only in the device environment (i.e., without Google Colaboratory), as shown in Table 2; the processing speed was about 1.5 fps. A similar



Table 1  
Features and statistics.

Domain	Category	Features	Street Parking (Ave, SD)	Moving (Ave, SD)	
BB	Frame	Number of frames	Size	26.76 (15.21)	28.47 (85.63)
			Min	−9.78 (56.08)	257.15 (131.03)
	Top left X-coordinate	Max	173.83 (49.70)	309.41 (128.17)	
		Mean	104.94 (49.52)	281.00 (123.03)	
		Variance	4217.88 (2300.95)	689.53 (1352.33)	
		Min	128.78 (7.45)	134.30 (13.30)	
	Top left Y-coordinate	Max	139.61 (11.92)	140.44 (14.23)	
		Mean	132.94 (8.03)	136.96 (13.19)	
		Variance	12.16 (19.46)	7.90 (21.86)	
		Min	110.28 (40.32)	304.00 (121.07)	
	Bottom right X-coordinate	Max	202.58 (36.47)	370.48 (143.67)	
		Mean	168.49 (37.41)	332.69 (124.09)	
		Variance	941.87 (549.28)	1147.44 (1906.91)	
		Min	151.38 (12.86)	157.42 (28.82)	
	Bottom right Y-coordinate	Max	211.88 (32.47)	180.87 (42.55)	
		Mean	172.73 (18.11)	166.73 (31.86)	
		Variance	498.42 (513.28)	166.69 (358.71)	
		Min	27.64 (17.35)	38.13 (68.18)	
	Width	Max	126.20 (43.36)	71.69 (79.81)	
		Mean	63.55 (22.09)	51.73 (70.83)	
		Variance	1353.64 (1132.87)	325.97 (646.67)	
		Min	18.93 (10.48)	21.06 (23.66)	
	Height	Max	75.97 (29.91)	42.67 (38.07)	
		Mean	39.79 (15.73)	29.77 (27.39)	
		Variance	461.99 (529.15)	145.62 (337.22)	
		Min	683.98 (743.92)	2129.84 (8715.11)	
	Area	Max	10642.57 (6920.13)	5430.04 (11078.68)	
		Mean	3566.82 (2434.79)	3259.73 (9463.90)	
Variance		16771503.79 (24863966.62)	4585890.60 (13482441.80)		
Min		37.51 (12.07)	28.72 (16.63)		
Speed	Max	40.45 (8.89)	31.33 (15.91)		
	Mean	38.99 (10.23)	29.90 (16.26)		
	Variance	3.32 (10.29)	3.48 (17.40)		
	Min	0.83 (5.29)	1.42 (13.40)		
Acceleration X-direction	Max	0.01 (0.04)	0.00 (0.03)		
	Mean	−0.02 (0.02)	−0.02 (0.03)		
	Variance	0.00 (0.00)	0.00 (0.00)		
	Min	−0.10 (0.05)	−0.08 (0.07)		
Acceleration Y-direction	Max	0.04 (0.05)	0.02 (0.07)		
	Mean	−0.03 (0.04)	−0.03 (0.05)		
	Variance	0.00 (0.00)	0.00 (0.00)		
	Min	0.93 (0.06)	0.95 (0.07)		
Acceleration Z-direction	Max	1.10 (0.05)	1.07 (0.07)		
	Mean	1.01 (0.02)	1.01 (0.05)		
	Variance	0.00 (0.00)	0.00 (0.00)		
	Min	0.00 (0.00)	0.00 (0.00)		

Table 2  
Video analysis environment.

Name	Description	Name	Description
Device	MacBook Pro 2018	OS (Colaboratory)	Ubuntu 18.04.3 LTS
OS	macOS Mojave (version 10.14.6)	CPU (Colaboratory)	Intel Xeon (2.3 GHz)
CPU	Intel Core i5 (2.3 GHz)	GPU (Colaboratory)	NVIDIA-SMI 440.82
GPU	Intel Iris Plus Graphics 655 (1536 MB)	Memory (Colaboratory)	11–13 GB of RAM
Memory	8 GB of 2133 MHz LPDDR3	—	—

video was analyzed in the GPU environment of Google Colaboratory at 15 fps. This analysis speed will be used as a benchmark for future real-time analysis.

### 3.6 Street parking detection model

We selected two machine learning algorithms for comparison, namely, RF and LR. RF was selected because it has been adopted in a wide range of fields for its high accuracy, expressive power, and generalization performance.<sup>(17,18)</sup> LR was selected because it has been widely adopted, mainly for binary classification problems.<sup>(19)</sup> The training model was built using scikit-learn, a machine learning library in Python. The parameters of the two machine learning algorithms were the default settings of scikit-learn.<sup>(20)</sup>

## 4. Evaluation Experiment

In this section, we describe the dataset we built for the evaluation experiments, the machine learning method, and the evaluation validation method.

### 4.1 Dataset

A total of 1765 vehicles were identified in the 12 videos (total length: about 24 min). Of these vehicles, 56 were detected as on-street parked vehicles, giving a ratio of on-street parked vehicles to moving vehicles of 1 to 30. This ratio varies depending on the data to be analyzed in the videos, the area recorded in the videos, the time of year, and the time of day.

In general, if a model is created for imbalanced data, it is likely to result in a discriminator that only fits a class with a large number of samples. Therefore, in this experiment, we used imbalanced data (collected data) and random sampling (undersampling) of nonparked vehicles from the imbalanced data to evaluate 51 nonparked vehicles and 51 on-street parked vehicles with balanced data.

To validate the model's performance under real-world usage scenarios, additional videos were analyzed under the same conditions. The dataset for the additional validation consisted of a total of 100 vehicles (95 nonparked vehicles and 5 on-street parked vehicles).

### 4.2 Evaluation and validation method

We used the holdout and stratified  $k$ -fold validation methods to partition and validate the dataset. The holdout method combines training and test data in a dataset at a constant ratio and the verification method. In this paper, the ratio was set to 7:3. The general  $k$ -fold cross-validation method simply divides the dataset into  $k$  subsets, whereas in the stratified  $k$ -fold validation method, the class ratio within a subset is divided so that the ratio of classes within the subset is the same as the overall ratio. The remaining  $k - 1$  subsets are then used for recognition for one subset. This method is repeated  $k$  times and evaluated. This improves the generalization performance. In this paper, the value of  $k$  is set to 5.



## 5. Results of Street Parking Detection

In this section, we describe the results of applying machine learning to labeled combined data. A machine learning method is applied to a dataset that combines the data from the analysis of each video.

### 5.1 Results for imbalanced data

The confusion matrices of recognition results obtained using the holdout method for imbalanced data and using the stratified 5-fold validation method are shown in Fig. 4. Here, the class label of on-street parked vehicles is 1 and that of moving vehicles is 0. Because the purpose of this test was to detect the number of on-street parked vehicles, the accuracy, precision, recall, and F-measure were calculated. The accuracy is for all vehicle data. The results are shown in Table 3. For the imbalanced data, a maximum F-measure of 59% was obtained for LR holdout. Because of the imbalanced data, the reproducibility rate is low.

### 5.2 Results for sampled data

The confusion matrices of recognition results obtained using the holdout method for the sampled data and using the stratified 5-fold validation method are shown in Fig. 5. We calculated the accuracy, precision, recall, and F-measure for street parking. The accuracy is for all vehicle data. The results are shown in Table 4. For the sampled data, a maximum F-measure of 92% was found for RF 5-fold. Compared with the imbalanced data, high recognition accuracy was confirmed for most of the indicators. However, some samples were misclassified.

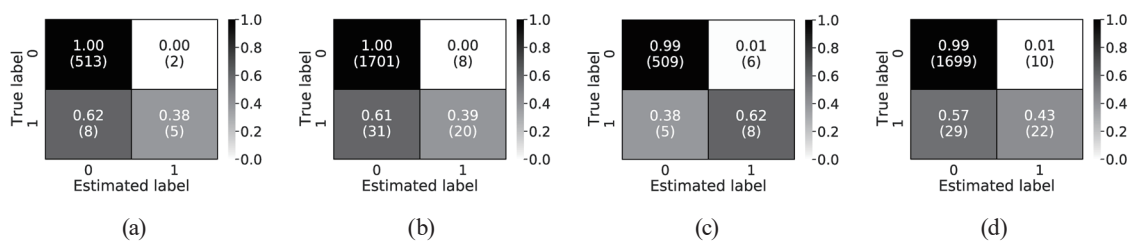


Fig. 4. Confusion matrices of detection results for imbalanced dataset. (a) RF holdout, (b) RF 5-fold, (c) LR holdout, and (d) LR 5-fold.

Table 3  
Results for imbalanced dataset.

Method	Accuracy	Precision	Recall	F-measure
RF holdout	0.98	0.71	0.38	0.50
RF 5-fold	0.98	0.71	0.39	0.51
LR holdout	0.98	0.57	0.62	0.59
LR 5-fold	0.96	0.69	0.43	0.53

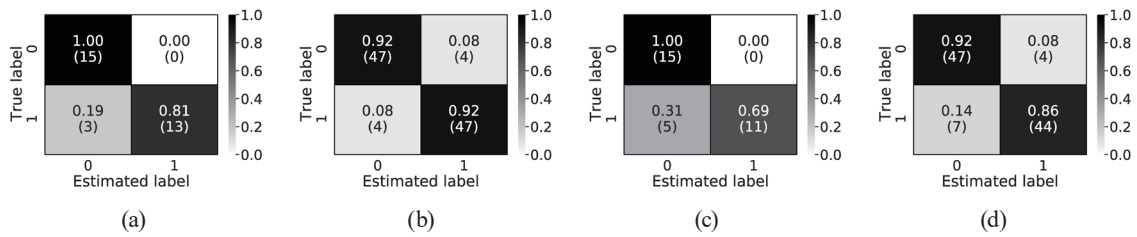


Fig. 5. Confusion matrices of detection results for sampled dataset. (a) RF holdout, (b) RF 5-fold, (c) LR holdout, and (d) LR 5-fold.

Table 4  
Results for sampled dataset.

Method	Accuracy	Precision	Recall	F-measure
RF holdout	0.90	1.00	0.81	0.90
RF 5-fold	0.92	0.92	0.92	0.92
LR holdout	0.84	1.00	0.69	0.81
LR 5-fold	0.89	0.92	0.86	0.89

### 5.3 Results for additional validation data

In order to evaluate the accuracy of recognition on datasets that more closely resemble real-world situations, a total of 100 vehicles that were additionally collected were validated with a model that was trained on the sampled data. Therefore, the high-accuracy model trained on a sampled dataset was used for validation. The confusion matrices of the recognition results obtained using the sampled data for the additional validation dataset are shown in Fig. 6. Table 5 shows the calculated accuracy, precision, recall, and F-measure for street parking. The accuracy is for all vehicle data. For the additional validation data, a maximum F-measure of 83% was obtained. Recall was 100% and the two methods were able to recognize all street parking, but precision was slightly low. In the future, the precision will be improved by increasing the training data.

## 6. Discussion

In this section, we discuss the results presented in Sect. 5. In particular, we discuss the importance of features and misclassified samples for the model with high recognition accuracy. Also, we discuss the effectiveness of the system in the real world.

### 6.1 Effectiveness of detection model

Overall, the accuracy was higher for imbalanced data. As mentioned, this model is considered to be suitable for the moving vehicle class with a large sample size. The maximum values of precision, recall, and F-measure were 71%, 62%, and 59%, respectively. These values can be improved by increasing the sample size of the street parking class and by improving the learning method.

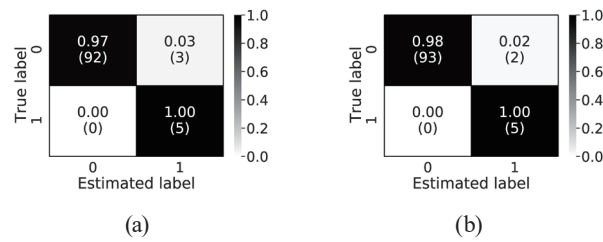


Fig. 6. Confusion matrices of detection results for additional validation dataset. (a) RF sampled and (b) LR sampled.

Table 5  
Results for additional validation dataset.

Method	Accuracy	Precision	Recall	F-measure
RF sampled	0.97	0.62	1.00	0.77
LR sampled	0.98	0.71	1.00	0.83

For sampled data, the precision, recall, and F-measure were all higher than those for the imbalanced data. The maximum values of precision, recall, and F-measure were 100%, 92%, and 92%, respectively. We consider that this high-accuracy model can be applied to future real-time recognition methods using in-vehicle devices. We need to further analyze the videos and increase the number of samples to improve accuracy. More effective machine learning methods may also be considered.

For additional validation data, the classification of binary classes was highly accurate despite the small number of data points. This confirms the validity of the model. Training with samples of the street parking class in various situations can be used to construct a model with high generalization performance.

Furthermore, in this study, we did not use the positional information such as latitude and longitude. As even if the positional information is the same, if the direction of the vehicle is different, the taken videos show very different scenes, and the situation of on-street parking varies significantly depending on the season and time of day, even at the same location. However, there are areas and times of day when street parking is more likely to occur. Therefore, the detection accuracy may be improved using the positional information combined with car direction and time of day information. In the future, we will verify whether we can construct a more accurate model by adding these features.

## 6.2 Importance of features

In order to detect on-street parking, we used information of BBs and the recording vehicle. In general, on-street parking is closer to the edge of the road, and it is considered that coordinate information is important. To confirm this, we visualized which features are important in Fig. 7. The importance of the features was calculated from the RF holdout that was trained on sampled data with high recognition accuracy. As shown in Fig. 7, the importance of information related

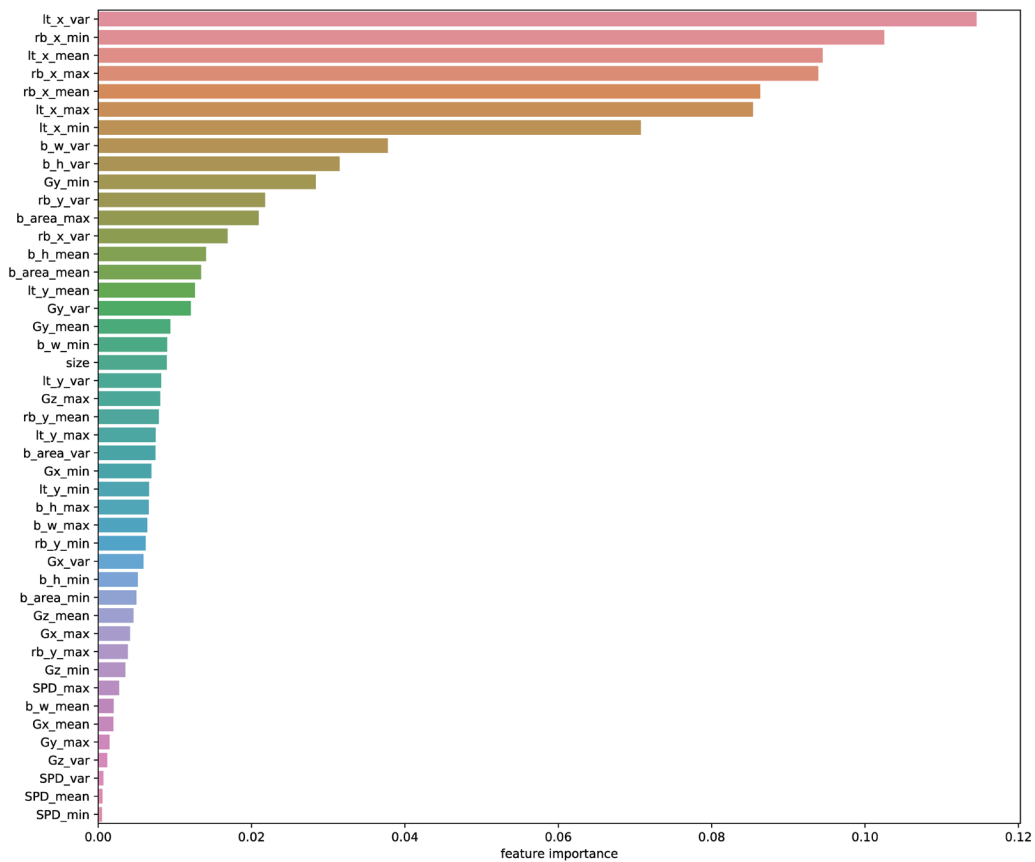


Fig. 7. (Color online) Importance of features learned by RF holdout of the sampled data.

to the BBs is higher. This indicates that BB information is effective for detecting on-street parking.

In addition, to investigate the influence of information related to BBs, we reduced the number of features and conducted an ablation study. We eliminated seven categories of BB information and conducted training using 17 features. Table 6 and Fig. 8 show the results of the RF holdout that was trained on the sampled data. The results show that the F-measure decreased to 69%, which confirms that BB information contributes to the detection of on-street parking. Also, Fig. 9 shows the importance of the features of ablation study. In the absence of BB information in the features, we can confirm that size (number of frames) is important for the detection of on-street parking. It is necessary to increase the number of samples and to further investigate which features are important. Also, to improve detection accuracy, it is necessary to change the weights to the features.

### 6.3 Misclassification

As mentioned, although the recognition accuracy was high, there were some cases of misclassification. We consider the case of misclassification with RF with a high F-measure (see the confusion matrix of recognition results for the sampled dataset in Fig. 5).

Table 6  
Results of ablation study.

Method	Accuracy	Precision	Recall	F-measure
RF holdout sampled	0.65	0.63	0.75	0.69

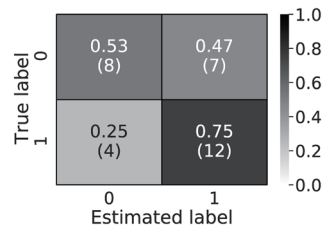


Fig. 8. Confusion matrix of detection results of ablation study in RF holdout sampled.

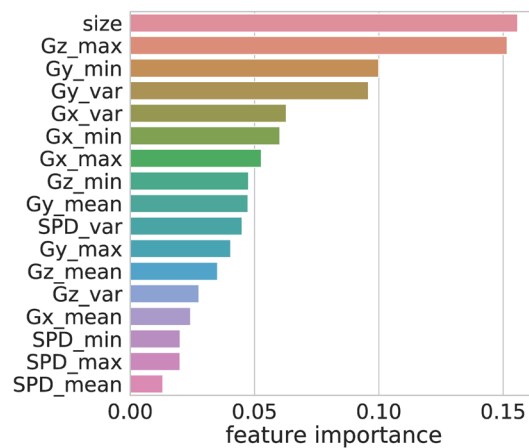


Fig. 9. (Color online) Importance of features of ablation study trained by RF holdout of the sampled data.

For RF holdout in Fig. 5, there are two samples in which a vehicle in the street parking class was misclassified into the moving vehicle class. Figures 10(a) and 10(b) show these two samples. In Fig. 10(a), vehicle number 238 was misclassified as a moving vehicle even though it was on-street parked. We consider that this was due to the delay in detecting vehicle number 238 as a result of a silver vehicle being in front of the recording vehicle. This silver vehicle made it difficult to detect the gradual spread of the BBs, which is a characteristic of on-street parked vehicles. In addition, the number of frames to be detected was reduced and sufficient information was not obtained.

In Fig. 10(b), on-street parked vehicle number 46 was misclassified as a moving vehicle. In this case, as in Fig. 10(a), a vehicle moving alongside and slightly in front of the white vehicle is considered to be the cause of the misclassification. We consider that the number of frames used to detect street parking is small. Therefore, increasing the number of samples of street parking in various situations could reduce the number of misclassifications. To improve labeling accuracy, it is also necessary to quantify and clarify the conditions suitable for labeling.



Fig. 10. (Color online) Two cases of on-street parked vehicle misclassified as moving vehicle.

For RF 5-fold in Fig. 5, there are four samples for which vehicles in the moving vehicle class were misclassified into the street parking class. Figures 11(a)–11(d) show these separate cases. A common characteristic among these cases is that a vehicle in the lane opposite to that of the recording vehicle is misclassified as an on-street parked vehicle. Vehicles in Figs. 11(a), 11(c), and 11(d) were misclassified into the street parking class. This is because the rate of change of the gradual widening of the BB, a characteristic of on-street parked vehicles, is similar to that of vehicles in the opposite lane, as described above. To deal with this, the detection model can be improved by changing the weights of the contribution rates of the features used for training. It is also considered that filtering based on the position of the BB and fine-tuning of the object detection model can reduce the number of misclassifications. Figure 11(b) shows an on-street parked vehicle located in the opposite lane. Under the current labeling conditions, the vehicle is recognized as a nonparked vehicle, but it was misclassified as on-street parked. However, because the vehicle is actually on-street parked, we will change the labeling conditions to detect the vehicle in the opposite lane to increase the versatility of the model.

#### 6.4 A system for the real world

In this section, we describe the effectiveness of our system in the real world. We will develop an on-vehicle device for the real-time detection of street parking. To input the required information to a system directly, we attach a camera and sensors to the device to construct a system that imitates a dashboard camera. Figure 12 shows the explanation of this system. The conversion of the image to text data will allow processing on the device. Then, only the results processed on the device will be uploaded to the cloud and displayed. This will also help reduce the communication load.

In this paper, much street parking data was obtained from videos of taxis that were not driving to look for on-street parking. Existing detection methods are not capable of collecting much data, however, this system can be installed in many vehicles. Therefore, we can detect many vehicles not only in a specific area but in the entire city. Furthermore, we believe that participatory sensing and gamification systems are effective for sensing roads with light traffic.<sup>(21,22)</sup> Matsuda *et al.* proposed a participatory sensing system that collects data using the participants as sensors.<sup>(21)</sup> In addition, Kawanaka *et al.* proposed a method of collecting data in



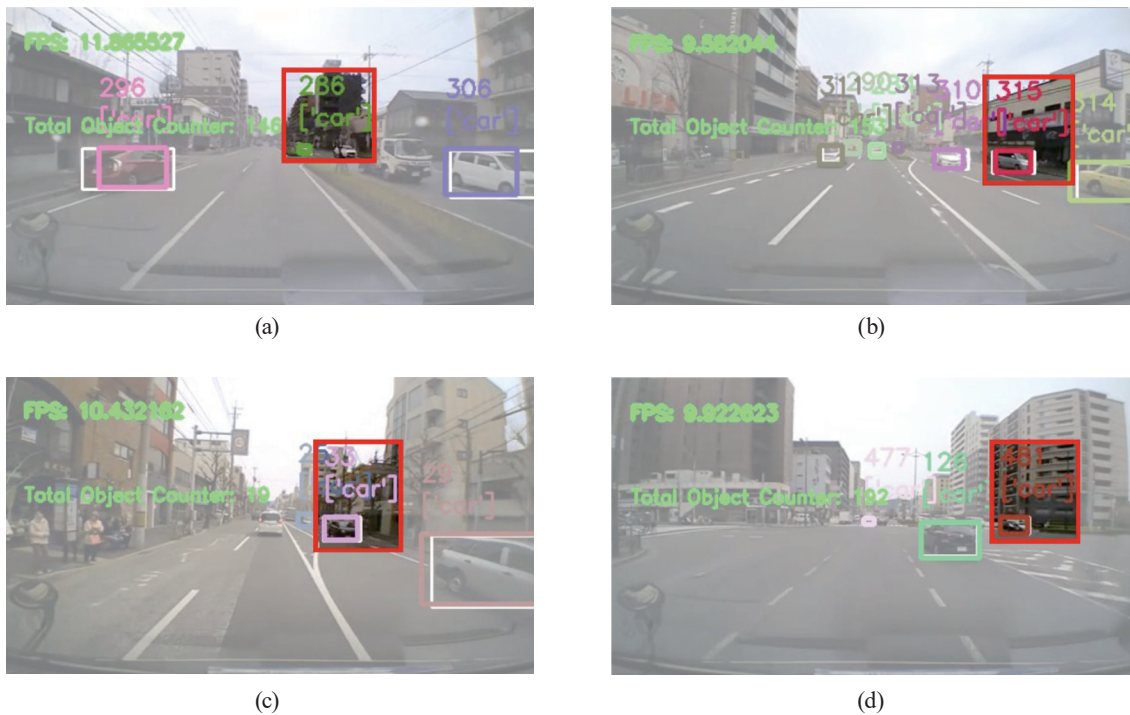


Fig. 11. (Color online) Four cases of moving vehicle misclassified as an on-street parked vehicle.

Real-time Street Parking Detection System using the Proposed Method

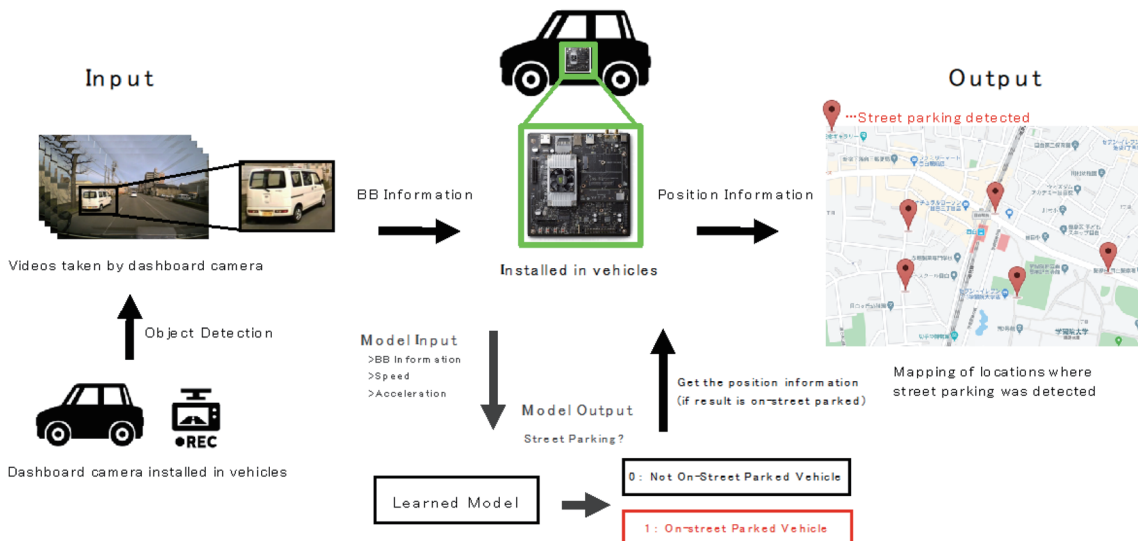


Fig. 12. (Color online) Real-time street parking detection system using the proposed method.

areas where data are difficult to collect, using gamification.<sup>(22)</sup> By combining these systems and the proposed model, we can construct a street parking vehicle detection system for real-world applications.

## 7. Conclusion

In this paper, we proposed a machine learning approach for the real-time recognition of street parking. We created a dataset based on BB information and recorded vehicle information by analyzing dashboard camera videos using an object detection model. We constructed a machine learning model for street parking using the dataset. The F-measure values calculated from the recognition results were up to 59, 92, and 83% for the imbalanced, sampled, and additional validation data, respectively. To further improve recognition accuracy, we analyzed misclassified vehicles and identified the problem that led to misclassification.

In the future, we will increase the number of videos analyzed and create detection models based on samples of street parking classes in various situations. In addition, we consider that it is necessary to modify the labeling conditions and carry out fine-tuning to further reduce the number of misclassifications. As shown in Fig. 12, the model created in this paper will be processed on a device. In the future, we consider that working with the police and other public agencies will help to reduce further problems. In this manner, we aim to create a system that efficiently detects street parking in real time and addresses the misclassification problem.

## Acknowledgments

The videos used in this study were provided by DENSO TEN Limited.

## References

- 1 Japan Parking Association: <http://japan-pa.or.jp/20200228/5382> (accessed July 2020) (in Japanese).
- 2 E. Bernal, Z. Fan, Y. Rong Wang, R. P. Loce, N. W. Zeck, and G. S. Pennington: US Patent No. 13/441,294 (2013). <https://patentimages.storage.googleapis.com/9b/f6/27/4f3a7027591d3a/US20130265423A1.pdf>
- 3 Y. Rong Wang, Z. Fan, G. S. Pennington, and H. Ding: US Patent No. 13/461,191 (2014). <https://patentimages.storage.googleapis.com/25/09/c7/805e54297fd14c/US8682036.pdf>
- 4 JEITA: <https://www.jeita.or.jp/japanese/stat/drive/2019/> (accessed July 2020) (in Japanese).
- 5 S. Ono, K. Hirahara, M. Kagesawa, and K. Ikeuchi: IEICE Trans. **88** (2005) 247 (in Japanese). <https://www.cvl.iis.u-tokyo.ac.jp/papers/all/712.pdf>
- 6 B. Li, T. Wu, and S. Chun Zhu: Proc. 2014 Springer Eur. Conf. Computer Vision (2014) 652–667. [https://doi.org/10.1007/978-3-319-10599-4\\_42](https://doi.org/10.1007/978-3-319-10599-4_42)
- 7 X. Sevilano, E. Märmol, and V. Fernandez Arguedas: Proc. 2014 IEEE Int. Conf. Information Fusion (IEEE, 2014) 1–8. <https://ieeexplore.ieee.org/abstract/document/6916135/authors#authors>
- 8 K. Malecki: J. Comput. Sci. **28** (2018) 32. <https://doi.org/10.1016/j.joics.2018.07.005>
- 9 W. Xiao, B. Vallet, K. Schindler, and N. Paparoditis: ISPRS J. Photogramm. Remote Sens. **114** (2016) 166. <https://doi.org/10.1016/j.isprsjprs.2016.02.007>
- 10 R. Kyutoku, T. Takahashi, I. Ide, T. Mekada, and H. Murase: Proc. 2008 ISPRS FIT (ISPRS, 2008) H-038 (in Japanese). <http://www.murase.is.i.nagoya-u.ac.jp/~ide/res/paper/J08-taikai-kyutoku-1.pdf>
- 11 R. Kyutoku, D. Deguchi, T. Takahashi, T. Mekada, I. Ide, and H. Murase: Proc. 2012 PRMU MIRU (PRMU, 2012) (in Japanese). <http://www.murase.is.i.nagoya-u.ac.jp/~ide/res/paper/J12-symposium-kyutoku-1pub.pdf>
- 12 X. Xie, C. Wang, S. Chen, G. Shi, and Z. Zhao: Proc. 2017 ACM Int. Conf. Deep Learning Technologies (ACM, 2017) 23–27. <https://doi.org/10.1145/3094243.3094261>
- 13 A. Matsuda, Y. Matsuda, H. Suwa, and K. Yasumoto: Proc. 2019 IPSJ Kansai branch (IPJS, 2019) (in Japanese).
- 14 J. Redmon, S. Divvala, R. Girshick, and A. Farhadi: Proc. IEEE Conf. Computer Vision and Pattern Recognition (IEEE, 2016) 779–788. [https://www.cv-foundation.org/openaccess/content\\_cvpr\\_2016/papers/Redmon\\_You\\_Only\\_Look\\_CVPR\\_2016\\_paper.pdf](https://www.cv-foundation.org/openaccess/content_cvpr_2016/papers/Redmon_You_Only_Look_CVPR_2016_paper.pdf)

- 15 J. Redmon and A. Farhadi: arXiv preprint arXiv:1804.02767 (2018). <https://arxiv.org/pdf/1804.02767.pdf>
- 16 T. Y. Lin, M. Maire, S. Belongie, J. Hays, P. Perona, D. Ramanan, P. Dollár, and C. Lawrence Zitnick: Proc. 2014 Springer Eur. Conf. Computer Vision (2014) 740–755. [https://doi.org/10.1007/978-3-319-10602-1\\_48](https://doi.org/10.1007/978-3-319-10602-1_48)
- 17 H. Habe: CVIM **31** (2012) 1 (in Japanese). [https://ipsj.ixsq.nii.ac.jp/ej/?action=repository\\_uri&item\\_id=82592&file\\_id=1&file\\_no=1](https://ipsj.ixsq.nii.ac.jp/ej/?action=repository_uri&item_id=82592&file_id=1&file_no=1)
- 18 H. Habe: J. Inst. Image Inform. TV. Engrn. **70** (2016) 788 (in Japanese). <https://doi.org/10.3169/itej.70.788>
- 19 H. Nakazawa: J. Nippon Med. Sch. **10** (2014) 186 (in Japanese). <https://doi.org/10.1272/manms.10.186>
- 20 F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Pretternhof, R. Wesis, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay: J. Mach. Learn. Res. **12** (2011) 2825. <http://www.jmlr.org/papers/volume12/pedregosa11a/pedregosa11a.pdf>
- 21 Y. Matsuda, Y. Arakawa, and K. Yasumoto: Proc. 2016 ACM Int. Conf. MobiSys (ACM, 2016) 57. <https://doi.org/10.1145/2938559.2948836>
- 22 S. Kawanaka, Y. Matsuda, H. Suwa, M. Fujimoto, Y. Arakawa, and K. Yasumoto: Smart Cities **3** (2020) 736. <https://doi.org/10.3390/smartcities3030037>

