

See discussions, stats, and author profiles for this publication at: <https://www.researchgate.net/publication/328640756>

Semantic segmentation-based parking space detection with standalone around view monitoring system

Article in Machine Vision and Applications · October 2018

DOI: 10.1007/s00138-018-0986-z

CITATIONS

0

READS

388

2 authors:



[Chulhoon Jang](#)

Hanyang University

11 PUBLICATIONS 195 CITATIONS

[SEE PROFILE](#)



[Myoungcho Sunwoo](#)

Hanyang University

202 PUBLICATIONS 2,018 CITATIONS

[SEE PROFILE](#)

Some of the authors of this publication are also working on these related projects:



Vision applications [View project](#)



Autonomous Vehicle [View project](#)



Semantic segmentation-based parking space detection with standalone around view monitoring system

Chulhoon Jang¹ · Myoungho Sunwoo¹

Received: 2 November 2017 / Revised: 21 July 2018 / Accepted: 9 October 2018
© Springer-Verlag GmbH Germany, part of Springer Nature 2018

Abstract

An auto-parking system is one of the promising technologies to reduce accidents and enhance driver convenience in parking lots. To accomplish collision-free parking, precise and robust parking space detection is required. However, harsh conditions such as varied illumination in outdoor parking lots and high reflection in indoor parking lots degrade the reliability of parking space detection. In this paper, we propose a unified structure for parking space detection to detect parking slot markings and static obstacles. A fully convolutional network for semantic segmentation can immediately identify free spaces, slot markings, vehicles, and other objects without using a range sensor or 3D reconstruction algorithm. Furthermore, a vertical grid encoding method can simultaneously detect unoccupied slots identified by parking slot markings and empty spaces created by surrounding static objects without sensor fusion. Experimental results show the robustness of the proposed method in various different parking scenarios. Even in challenging conditions such as dark shaded or high-glare areas, the detection performance maintains a precision rate of 96.81% and recall rate of 97.80%.

Keywords Automatic parking system · Parking space detection · Semantic segmentation · Around view monitor system · Computer vision · Fully convolutional network

1 Introduction

A parking lot is a commonly used place for people, and vehicles need to be parked at the completion of drives. Because parking requires positioning a car in a small space without collision, many drivers suffer damages. In fact, 40% of all car accidents happen in the parking lot, which causes physical damage and human injury; thus, parking accidents must be prevented [2]. In the automotive field, parking-aid sensors are mounted on commercial vehicles that can help drivers avoid such accidents. An ultrasonic sensor generates an audible alarm to prevent collisions when the vehicle is close to any object. A camera is used to display a rear view or a 360° surround view of the vehicle. Because the driver can check a blind area through a camera, collisions during parking can be avoided. Recently, a parking-assist system has employed

an ultrasonic sensor to perform parking space search and automatically parks the vehicle. The demand for automated parking systems is expected to increase in the future.

In general, an automated parking system is a type of intelligent functionality in the automotive field that enables a vehicle to drive itself and park in a desired location without any accident. An automated parking system consists of three main functions: parking space detection, parking-path generation, and vehicle control. The parking space detection recognizes available parking spaces using a range sensor or camera and maps the destination. The parking-path generation finds an optimal path to reach the destination. The vehicle control automatically manipulates the steering and acceleration to move the vehicle to the destination. Among them, parking space detection is one of the primary components of an automated parking system because it is required to reliably provide the position of a parking space in various parking scenarios.

Conventionally, parking space means an unoccupied area that is identified by parking slot markings or an empty space that is created by adjacent static objects such as vehicles, pillars, or walls. Thus, we should be able to simultaneously detect unoccupied slots and empty spaces for automated

✉ Myoungho Sunwoo
msunwoo@hanyang.ac.kr

Chulhoon Jang
chulhoonjang@gmail.com

¹ Hanyang University, 222 Wangsimni-ro, Seongdong-gu, Seoul 04763, Republic of Korea

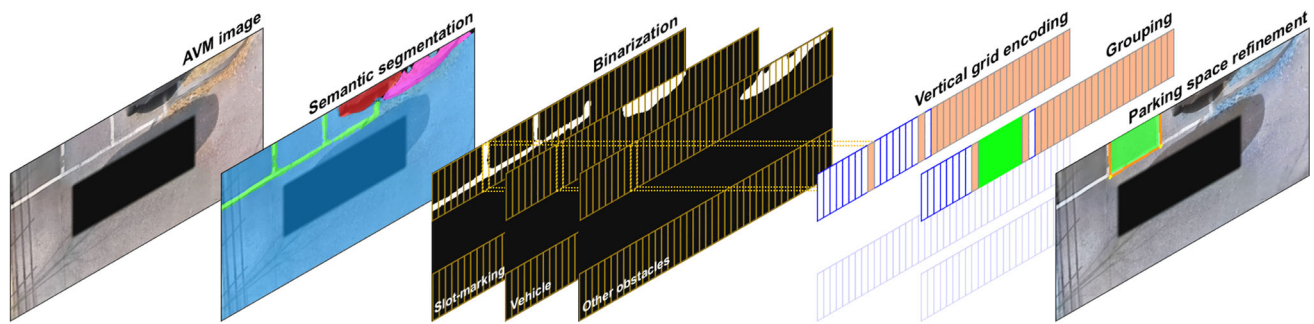


Fig. 1 Procedure of parking space detection using semantic segmentation and vertical grid encoding

parking. Camera-based methods are useful in recognizing parking slot markings, whereas range-sensor-based methods are effective in detecting static objects. Although parking slot markings are provided in most parking lots, deteriorated parking slot markings or faded lines may exist in some places. In such environments, we need to detect empty spaces formed using surrounding static obstacles. To overcome various challenging environments, recent research has integrated two methodologies to improve detection performance.

For general purposes, parking space detection should be possible in indoor and outdoor parking lots. In particular, modern cities have become densely populated, and spaces for outdoor parking lots are scarce. Thus, many parking lots are located inside buildings, and drivers have to perform frequent parking operations in indoor parking lots. However, indoor parking lots have narrow parking spaces to increase parking density. Therefore, high accuracy in parking space detection is required for automated parking. In addition, camera-based parking space detection can be easily affected by reflected lights from the ground due to ceiling lights. Reflected lights can appear very similar to parking slot markings in terms of shape and color. Distinguishing between the parking slot markings and reflected lights is difficult. In contrast to outdoor parking lots, precise detection of pillars is required in indoor environments because pillars are usually located next to parking lots, and vehicles should be able to avoid collision with the pillars. In outdoor parking lots, a major challenge for camera-based parking space detection is adapting to various illumination conditions. In particular, shadows created by surrounding obstacles can degrade intensity differences between the parking slot markings and ground, or colored lights can discolor obstacles.

In the present paper, we propose a parking space detection method using semantic segmentation and vertical grid encoding as in Fig. 1. The proposed system has the following advantages:

- (1) The deep-learning-based semantic segmentation can provide robustly labeled data in various parking environments, including different illumination conditions outdoors and high-glare conditions indoors.
- (2) The proposed system can simultaneously detect parking slot markings and static objects in an Around View Monitor (AVM) image; thus, range sensor or stereo-matching technique is not required to determine parking space occupancy.
- (3) The vertical-grid-based parking space detection can quickly and efficiently detect parking spaces in various parking environments without sensor fusion.

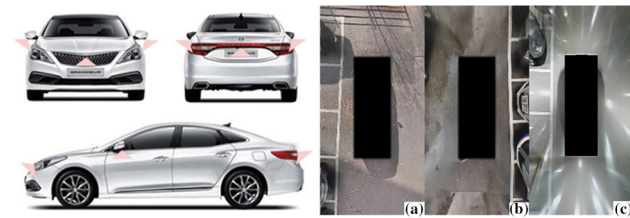
The proposed system consists of two stages: semantic segmentation and parking space detection. First, the deep neural network-based semantic segmentation classifies visual objects such as free space, parking slot marking, vehicle, and other objects in an AVM image. Second, the vertical-grid-based parking space detection encodes the labeled data in the lower and upper regions of interest (ROI) and identifies parking spaces that include an area formed by slot markings and an empty space surrounded by static objects. Parking space refinement estimates the accurate position and heading of parking spaces. Therefore, the proposed system can successfully detect perpendicular and parallel parking spaces indicated by parking slot markings or adjacent objects in both outdoor and indoor parking environments.

2 Previous research for AVM-based parking space detection

To detect vacant parking spaces using AVM system, two functionalities are required. First, we need to extract parking slot line and obtain geometrical information of parking space to locate the destination for automated parking. One of challenging parts for the line extraction is to satisfy robustness in various illumination conditions including outdoor/indoor and day/night because AVM system is a type of passive sensor that requires a light source. Second, we need to detect surrounding obstacles. The obstacles allow us to determine unoccupied slots and empty spaces between adjacent vehicles. Furthermore, the ego-vehicle can generate a feasible

Table 1 Slot marking filtering and line fitting algorithms for parking slot detection

Authors	Slot marking filtering	Line fitting
[12]	Canny edge	Radon transform
[5]	Pillbox kernel and morphological extrema filter	Radon transform
[10]	Sobel filter	RANSAC, chamfer matching
[7]	Top-hat operation	Directional density-based spatial clustering
[9]	Sobel filter	RANSAC
[14]	Line segment detection (LSD)	

**Fig. 2** Positions of four cameras of AVM system in the test vehicle and three sample images: **a** outdoor-clean, **b** outdoor-rainy, and **c** indoor

path to reach the destination and avoid collisions considering positions of the obstacles.

2.1 Parking slot detection

In previous research, most parking slot detection contains two procedures: filtering and fitting. In the filtering process, the intensity difference of parking slot markings against ground is emphasized by gradient-based filter, morphological operation, or customized filter. Based on the emphasized pixels, various fitting algorithms are applied to model lines of parking spaces. Table 1 summarizes filtering and fitting algorithms of AVM-based parking slot detection. One of challenging parts is the filtering process because it is hard to robustly extract parking slot markings in various illumination conditions as in Fig. 2a–c. Depending on the number of pixel candidates in the filtered image, the fitting algorithms might generate many false positives or fail to estimate the line due to the lack of pixel candidates. Therefore, both algorithms should be carefully tuned to satisfy all parking conditions.

2.2 Static obstacle detection

There are few previous researches for standalone AVM-based static obstacle detection. Wang et al. do not detect static obstacles directly. Instead, they perform the occupancy classification by counting edge pixels in parking slots [12]. Since this method assumes the parking slot has a non-texture surface, it might fail in parking spaces that have road surface markers such as woman-only or handicap. Han et al. [5] builds a height map of the surrounding vehicles

by dense motion stereo and utilizes the map fusion technique with parking slots to determine occupancies. This method works only when the ego-vehicle is moving. Other research utilizes ultrasonic sensors to detect static obstacles and conduct fusion techniques with parking slot detection for occupancy classification [9,10,14]. Thus, this method offers the possibility of extending the versatility of automated parking system in various parking scenarios such as in the absence of parking slot marking area or adjacent static objects. However, intensive effort might be required for sensor fusion, and data synchronization problem must be addressed. In addition, the system architecture should be concretely designed to realize synergy of the sensor fusion.

3 System overview

3.1 System architecture

The proposed system is composed of two steps, as shown in Fig. 1. First, semantic segmentation classifies four different objects in an AVM image, such as free space, slot marking, vehicle, and other objects. To understand the scene surrounding the ego-vehicle, a deep-learning-based encoder and decoder algorithm is employed. This method can immediately generate labeled data, which indicates the object classes for a full image. Second, a vertical grid efficiently encodes the labeled binarization data. The encoded data are utilized to classify occupied and unoccupied regions. Next, parking space refinement searches accurate parking spaces according to the unoccupied regions because accuracy of the parking destination is imperative for successful parking. As a result, precise position and heading to the parking spaces are obtained for parking-path planning.

3.2 Test vehicle

Figure 2 shows four camera locations in an AVM system: two cameras on the side mirrors, one camera at the front bumper, and one camera at the rear bumper. Each camera has

a 185 field of view; thus, stitching the four images generates an image that covers 360° that surrounds the ego-vehicle, which is called the birds eye view. This birds eye view is obtained by inverse perspective mapping [3], which assumes a flat ground. Therefore, if this assumption is not satisfied, the stitched image becomes incorrectly merged. Furthermore, extrinsic parameter calibration of each camera is important because if any camera position is changed, the birds eye view is also transformed. The AVM system can cover $12\text{ m} \times 6\text{ m}$ area, and the resolution of the image is 320×160 pixels. Figure 2a–c shows sample images of an AVM system under different lighting conditions: (a) outdoor-clean, (b) outdoor-rainy, and (c) indoor conditions. In the outdoor conditions, shadows cause varying illumination conditions that disrupt reliable marking extraction. Furthermore, if it is raining, rain-drops may cover the lens of the outside cameras; thus, the birds eye view could be irregularly transformed. In indoor conditions, high-glare conditions occur due to ground reflection. The glaring lights that appear similar to white markings can cause false positives.

4 Parking space detection using semantic segmentation and vertical grid

4.1 Semantic segmentation for AVM image

Inspired by the semantic segmentation [8,11], we formulate a scene interpretation problem for an auto-parking system using an AVM image. In parking scenarios, scene complexity can be simplified compared with an urban street scene. Figure 3 shows a common scene for parking. To proceed with auto-parking, the needed information includes free space, slot marking, vehicle, and other objects, including curbs, pillars, and walls. Free space indicates a movable area for parking-path planning and vehicle control. The slot marking guides the destination to the parking space. Adjacent vehicles represent occupancies in parking spaces. Finally, the other objects represent a sort of obstacles that the ego-vehicle cannot pass or should avoid during parking. Therefore, we consider these four classes, namely free space, slot marking, vehicle, and other objects, as outputs of the semantic seg-

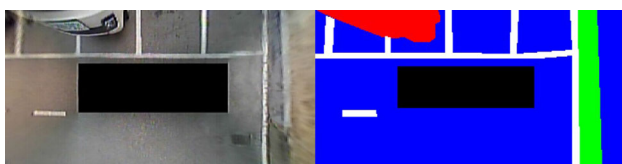


Fig. 3 Examples of AVM image and colored annotations for each class: blue for free space, white for the slot marking, red for the vehicle, and green for the wall. Black for ego-vehicle is excluded in training

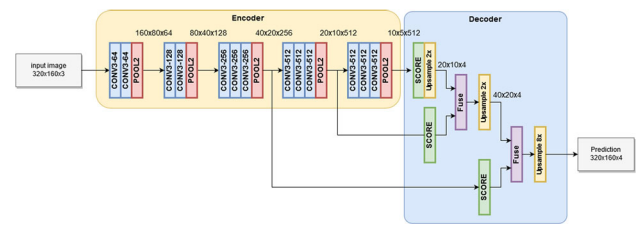


Fig. 4 Algorithm structure for semantic segmentation of AVM image. The input image is inserted as 320×160 pixels with RGB channels, and the prediction outputs a unified image with four channels. Each channel contains probability of a class inference

mentation, and an AVM image is used for the input, which contains RGB channels and 320×160 pixels.

Semantic segmentation consists of two modules, namely encoder and decoder, as shown in Fig. 4. The encoder is used in the abstraction of features from a low to high level in the input image. On the basis of a rich feature set, the decoder performs pixel-wise class inference by applying an upsampling method. The advantage of this structure is that end-to-end learning by backpropagation from pixel-wise loss becomes possible because the entire structure is composed of a fully convolutional neural network. Further, it can accommodate scalable input size because it only needs to perform convolution operation without a fully connected layer. Moreover, because structures and model parameters already proven in the classification can be applied to the encoder, much effort on training can be saved. The skip architecture applied to the decoder is effective in improving the low spatial precision in the upsampling. In the present study, we apply VGG 16-layer as an encoder. The VGG 16-layer is a proven model in ILSVRC14 with superior performance [13]. In the VGG 16-layer structure, 13 layers of convolutional neural network at the front end for feature extraction are implemented in the encoder because the fully connected layers at the back are responsible for classification. In the decoder, three intermediate layers are fused for refinement of the pixel-wise semantic segmentation. Bilinear interpolation parameters for upsampling are obtained through training. Finally, the decoder outputs an inference image where each channel contains probabilities for the four classes.

4.2 Vertical-grid-based parking space detection

On the basis of the semantic segmentation, we propose a vertical-grid-based parking space detection that consists of four steps: binarization, vertical encoding, grouping, and refinement as shown in Fig. 1. This architecture can provide refined parking space information, including unoccupied slots identified by parking slot markings and empty spaces formed by adjacent static objects without a sensor fusion. The key idea of this algorithm is the vertical grid conversion

that abstracts binarized labeled data into a horizontal vector. The vertical grid encoding enables searching, grouping, and comparing in one dimension. Furthermore, it does not require line fitting algorithms such as the Hough transform, Radon transform, or RANSAC to detect a parking space or primitive feature extraction such as edge or corner.

4.2.1 Binarization

The pixel-wise inference image of the semantic segmentation has four channels, which include probabilities of each class per pixel. For binarization, optimal thresholding values are applied to each channel except for the free-space channel as in Fig. 1. The optimal thresholding value of each channel, which results from the model validation after training, is determined to maximize MaxF1 of the semantic segmentation.

4.2.2 Vertical grid encoding

Initially, the ROI is individually specified in the upper and lower areas of an AVM image because the vehicle should scans left and right areas to identify vacant parking spaces. A vertical grid is generated with a fixed resolution, which considers the width of the slot marking in each ROI. Next, the vertical grid encoder counts all valid pixels in a grid, and the corresponding value of the grid is either set to one or not, depending on a threshold for the cumulative value. This operation is performed along the horizontal direction. As a result, we obtain three 1-D horizontal vectors to represent the labeled data for each ROI. Three vectors result in the final 1-D horizontal vector by using OR operation. In Fig. 1, two blue vertical grids indicate the final vectors for upper and lower ROIs. A colored grid means that the grid is set as 1, whereas an empty grid shows the grid is set as 0.

4.2.3 Grouping

The adjacent empty grids are grouped as parking space candidates. The parking space candidates might be unoccupied slots formed by parking slot markings, empty spaces surrounded by adjacent static obstacles or free spaces. To classify the parking space candidates into perpendicular and parallel parking spaces, we define range specifications considering parking slot regulations. Each group has start and end positions in the 1-D horizontal vector.

4.2.4 Refinement

To estimate an accurate position of the destination, we propose a refinement method because unoccupied slots and empty spaces do not provide accurate position due to the usage of the vertical grid coordinate. Figure 5 shows examples of parking space refinement for unoccupied slot (a) and

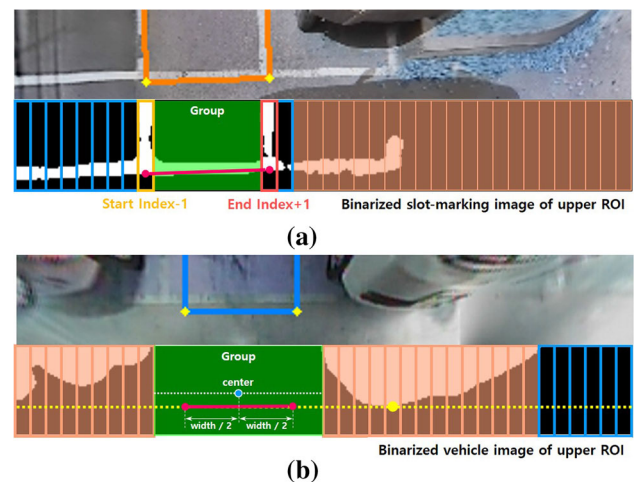


Fig. 5 Refinement of the unoccupied slots (a) and the empty spaces (b)

empty space (b). For the unoccupied slots, we scan two grids in the vertical grid of the binarized slot marking image, indicated by start index-1 and end index+1 of a group as in Fig. 5a. Next, we find the end of the pixels in each grid. Two end positions are connected and defined as the entrance to the unoccupied slot. For the empty spaces, the parking space can be identified by the location information of the surrounding static objects. The entrance is calculated based on the center point of the grouping grids and the lower limit of the static obstacles, assuming that nearby objects are aligned in parallel as in Fig. 5b. The width is defined considering the width of ego-vehicle. Subsequently, accurate position and heading of the parking destination are estimated based on the regular specification of the parking slot as orange and blue boxes in Fig. 5a, b, respectively.

5 Experimental results

5.1 Introduction to public datasets for parking space detection

There had not been any public dataset for parking space detection by using AVM system. To encourage related research, we firstly release two types of datasets of AVM images. The semantic segmentation (SS) dataset aims to the pixel-wise inference, whereas the parking space (PS) dataset is dedicated for parking space detection. Both datasets can be downloaded in [1]. SS dataset is composed of AVM and corresponding annotation images collected from various parking conditions in outdoors and indoors. With SS dataset, we can train and validate the model of deep-learning-based semantic segmentation. PS dataset contains AVM images and vacant parking spaces of 75 parking sequences in outdoors

Table 2 SS dataset composition (PS: parking space, PAR: parallel, PER: perpendicular)

Condition	Slot type	PS type	Frames
Outdoor	Closed	PER	2005
	Opened	PER	674
	No marking	PER	19
	Closed	PAR	686
	No marking	PAR	230
Indoor	Closed	PER	2642
	No marking	PER	340
	Closed	PAR	67
	No marking	PAR	100
Total number of frames			6763

Table 3 PS dataset composition (PS: parking space, PAR: parallel, PER: perpendicular)

Condition	Slot type	PS type	Num. of PS
Outdoor	Closed	PER	7765
	Opened	PER	2704
	Closed	PAR	1862
	Opened	PAR	429
	No marking	PAR	51
Indoor	Closed	PER	20,548
	No marking	PER	920
Total number of PS			34,279

and indoors. Total number of AVM images is 21581 frames. Tables 2 and 3 show details of dataset compositions including slot types, parking space types, and parking lot conditions.

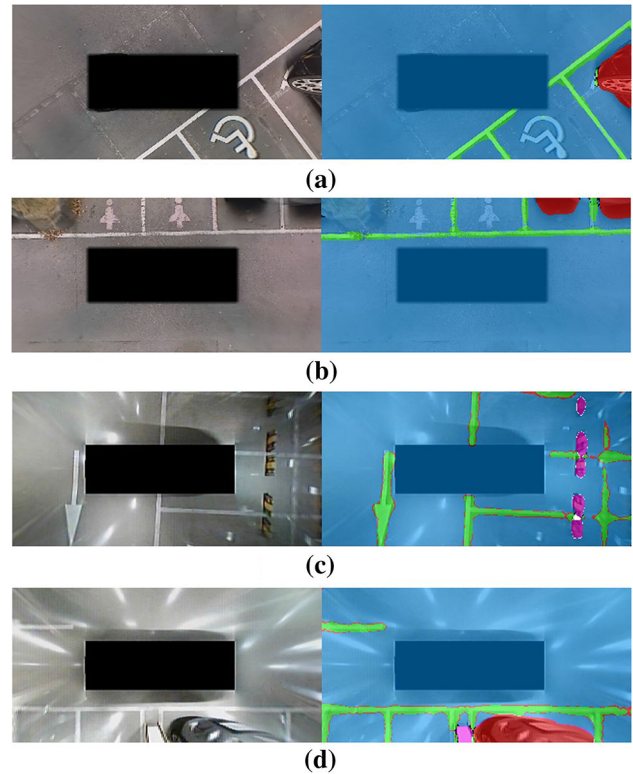
5.2 Evaluation of semantic segmentation for AVM image

5.2.1 Training with SS dataset

The training environment consists of Intel-Xeon E5-2630@2.30 GHz and 32 GB memory. A graphic processing unit (GPU) can accelerate the computation; however, a minimum of 12 GB of GPU memory is required to train the model of semantic segmentation using a VGG encoder [11]. The training sets consist of 3849 images. To perform supervised learning, each image is annotated according to class. Figure 3 shows the original and annotated images. Because of the lack of samples in the original data alone, we perform boosting using the data-augmentation method. A method of randomly setting the position of the upper left and lower right corners and cropping the image was applied. In addition, no dependence exists on the top, bottom, left, and right of the sample data. Therefore, four times more data are generated using

Table 4 Evaluation of the semantic segmentation model for test datasets

Class	Max $F1$ (%)	Average precision (AP) (%)
Free space	97.95	98.81
Slot marking	74.31	76.23
Vehicle	90.16	89.41
Other objects	78.23	81.39

**Fig. 6** Results of the semantic segmentation for the disabled parking lot, women-only parking lot, stoppers, and pillars

vertical and horizontal flipping. We divide SS dataset into 4057 frames for training and 2706 frames for test. Among the training set, 10% frames are used for validation. Before the training is started, all weights of the encoder are initialized using a pre-trained VGG model on ImageNet.

For the segmentation evaluation, we employ the F -measure derived from the precision and recall values. F -measure has been used in road and lane estimation of the KITTI Vision Benchmark suite [4]. The traditional F -measure or balanced F -score denotes the harmonic mean of precision and recall, i.e., is equal to one. In this case, it is called as $F1$ -measure.

5.2.2 Inference results

Maximum $F1$ -measure is used to evaluate the proposed semantic segmentation method. Table 4 lists the segmen-

tation results of the data that include various lighting and parking conditions. Through the proposed semantic segmentation method, the location and shape information of the free space, slot markings, vehicles, and other objects under outdoor, indoor, and rainy conditions can be effectively extracted from the image. The second column of Figs. 7 and 8 shows examples of the semantic segmentation results under different conditions. The proposed method has several advantages. First, it does not receive much shadow effect. A parking line is effectively recognized even when a shaded area due to surrounding objects is present outdoors. In addition, because it recognizes not only the shape of the vehicle but also the shadow portion as a vehicle, it is very useful for determining whether or not the parking lot is occupied. Finally, it is not affected by reflected light, which often occurs in an indoor environment. The reflected light is hard to filter because it has a similar color and shape to the parking line; however, the proposed method can solve the problem. Finally, because free space can be obtained, strict constraints can be provided to the parking-path planning and vehicle control for auto-parking. The result of the slot marking segmentation is not affected by other road markings. The disabled parking mark and women-only parking mark are not recognized as slot marking as in Fig. 6a, b. Because each mark is very similar to the slot marking in terms of color and thickness, it cannot be classified by the method using a Sobel or a Canny edge filter. The erroneously detected slot marking may degrade the parking slot detection. The height of the stopper in an indoor parking lot is low, and it is difficult to detect by an ultrasonic sensor. However, we successfully recognized the stopper as a fourth class as in Fig. 6c. The stopper position can be used as a stopping guide for longitudinal control during parking. Unlike outdoor parking lots, pillars are present in indoor parking lots. In most cases, a parking lot is located next to a pillar. However, it is very unlikely to be recognized as a slot marking because the slot marking is very close to the pillar. If parking is required between a parked vehicle and the pillar, we need to recognize the pillar to accurately

prevent a collision. The proposed method shows the correct recognition of a pillar indoor, as shown in Fig. 6d.

5.3 Evaluation of vertical-grid-based parking space detection

The proposed method was tested with PS dataset including indoor and outdoor conditions. We counted the ground truths of the parking spaces presented in each frame and calculated the recall and precision, as listed in Table 5. We accomplished 96.81% precision and 97.80% recall rates. The performance under indoor environment (precision 96.03%, recall 97.88%) notably did not deteriorate compared with the performance (precision 98.14%, recall 97.67%) under outdoor environment owing to the robustness of the semantic segmentation.

Figures 7 and 8 show the detection process in various parking conditions. The input image was grouped into semantic classes through semantic segmentation. As shown in the overlay image in the second column of Figs. 7 and 8, the free space is represented in blue, the occupied area is represented in red, the slot marking is shown in green, and the unoccupied slot is identified by a light green color. After limiting the ROI to the top and bottom regions of the image, a vertical grid was created, and we considered the slot marking, vehicle, and other objects as encoded data. Finally, the area detected as an empty space provided a precise parking position through refinement.

Both opened and closed types were correctly detected because the vertical shape of the slot marking was effectively interpreted in the vertical grid with half resolution of the slot marking width. The refinement method for finding an empty space is excellent in recognizing parking space even in the absence of a slot marking. Figure 7e shows that a parking space was detected based on the left pillar and right parked vehicle when the slot marking was entirely damaged. Figure 8c shows an empty space between the parked front and rear vehicles as a parallel parking space. Parallel and perpendicular parking spaces are classified according to the horizontal

Table 5 Validation performance of the vertical grid-based parking space detection (GT: ground truth, TP: true positive, FP: false positive, FN: false negative)

Type	Condition	Slot type	GT	TP	FP	FN	Precision (%)	Recall (%)
Perpendicular	Outdoor	Closed	7765	7650	111	115	98.57	98.52
Perpendicular	Outdoor	Open	2704	2704	25	0	99.08	100.00
Perpendicular	Indoor	Closed	20548	20113	833	435	96.02	97.88
Perpendicular	Indoor	No marking	920	900	35	20	96.26	97.83
Parallel	Outdoor	Closed	1862	1686	101	176	94.35	90.55
Parallel	Outdoor	Open	429	426	0	3	100.00	99.30
Parallel	Indoor	No marking	51	47	0	4	100.00	92.16
Total			34279	33526	1105	753	96.81	97.80

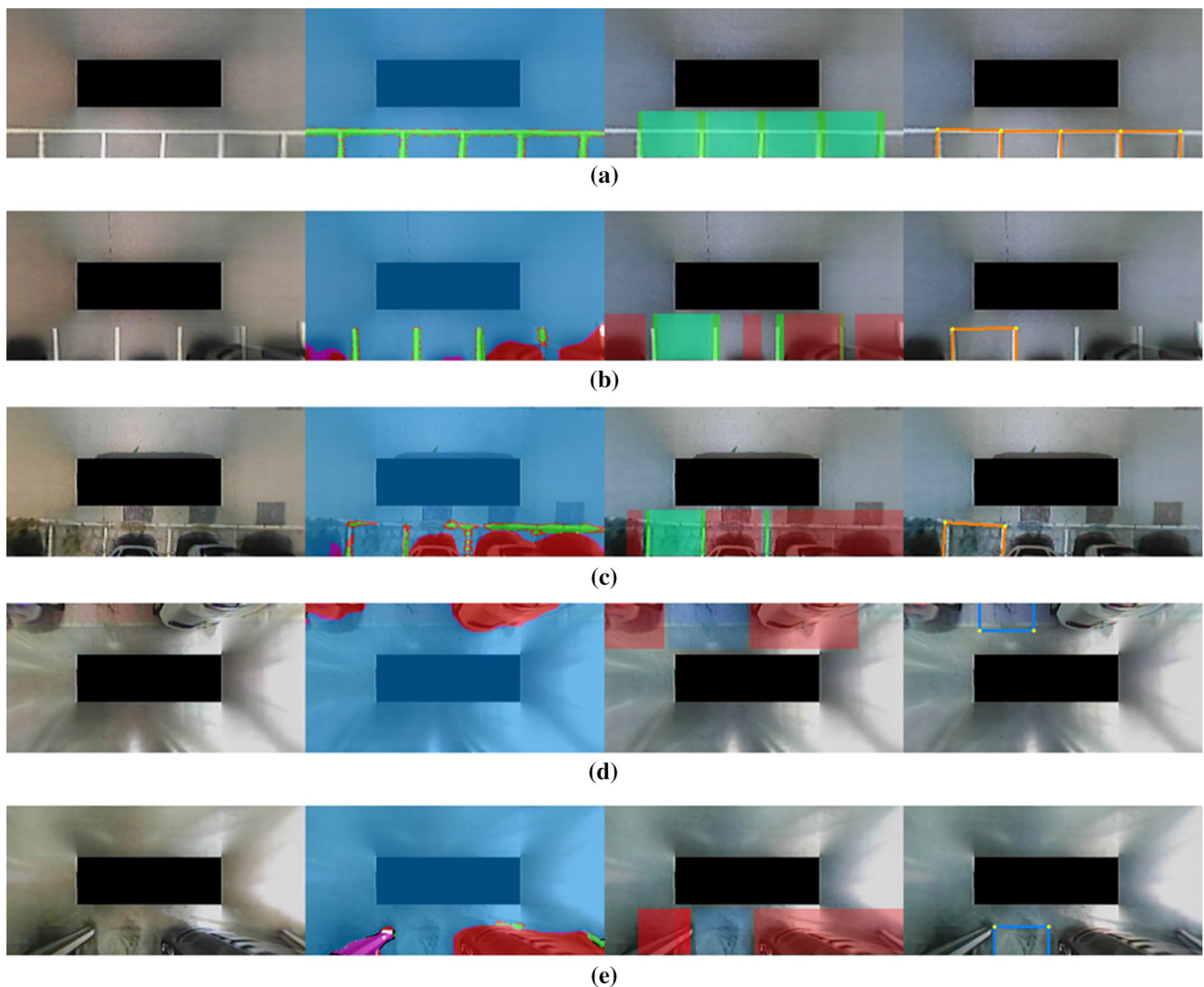


Fig. 7 Detection results of the perpendicular parking in various conditions: **a** outdoor, closed type, **b** outdoor, open type, **c** outdoor, closed type, gravel ground, **d** indoor, closed type, and **e** indoor, closed type,

severely damaged slot marking. As column-wise, input, semantic segmentation, vertical grid, and refinement images are displayed in order

length of the detected parking spaces based on Korea regulations on parking slot markings.

6 Analysis

6.1 Accuracy of semantic segmentation

The semantic segmentation performs upsampling in the decoder. Although bilinear interpolation provides high precision for free space (AP = 98.81%) and vehicle (AP = 89.41%), the AP of the slot marking was 76.23%. Therefore, we performed a detailed analysis. Figure 9a shows one of the validation images with ground truth annotation, and Fig. 9b shows the corresponding result of the semantic seg-

mentation for the slot marking. We find that the pixels of the slot marking semantics have significant location errors by comparing the magnified images, as shown in the top-left corner in Fig. 9. The reason is that the width of the slot marking is only six pixels, and the upsampling rate is at least $8\times$; thus, accurate inference is not possible. To minimize the location error, details of the location information in the semantic segmentation should be improved.

6.2 Failure cases of semantic segmentation

The failure cases of the semantic segmentation are described as follows. First, the recognition rate of the slot marking is degraded if contrast with the ground is not remarkable. Slot markings that are severely damaged are also classified as

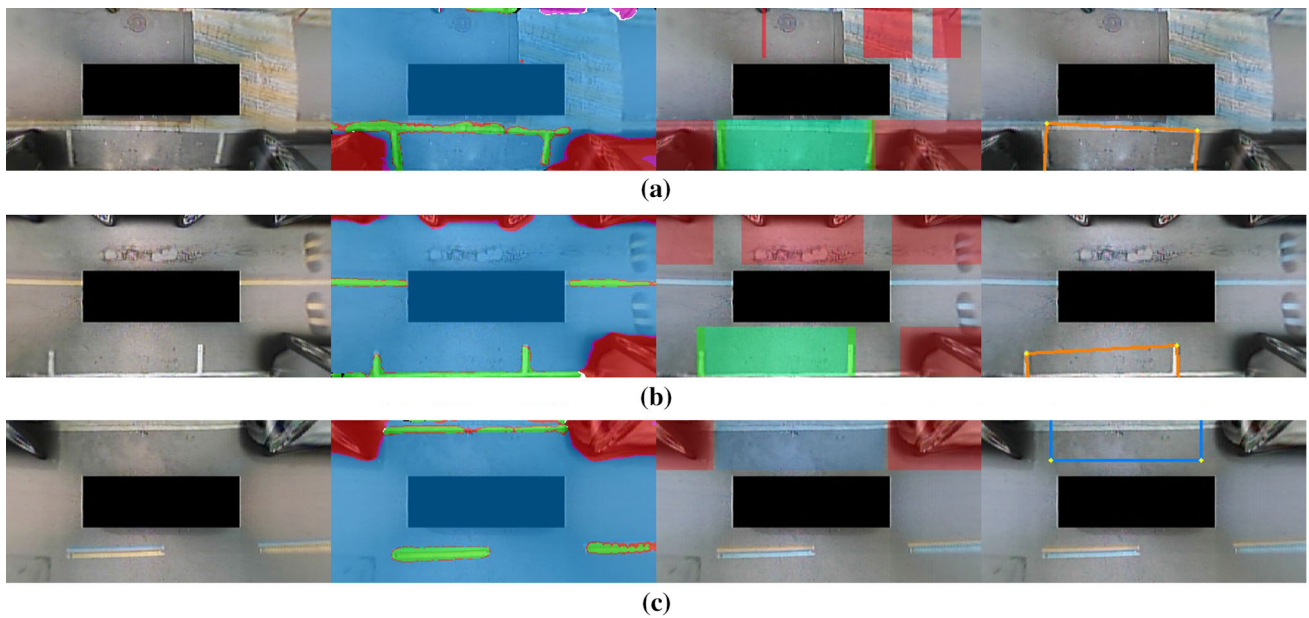


Fig. 8 Detection results of parallel parking in various conditions: **a** outdoor, closed type, **b** outdoor, open type, and **c** outdoor, no marking. As column-wise, input, semantic segmentation, vertical grid, and refinement images are displayed in order

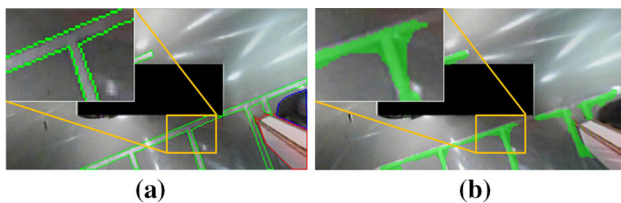


Fig. 9 Precision analysis of semantic segmentation

free space, as shown in Fig. 10a, b. In addition, a vehicle is mistakenly recognized due to the white light reflection. We can see that an unoccupied slot is classified as being occupied by a vehicle, as shown in Fig. 10c.

6.3 Execution times

The proposed algorithm is implemented in Python language. In the NVIDIA GTX 680 environment, the semantic segmentation takes 100 ms. The vertical-grid-based parking space detection takes 1.157 ms.

6.4 Limitations

The vertical-grid-based parking space detection assumes parallel scanning during parking. Therefore, we analyze the acceptability of the proposed algorithm according to the angle variation. In the presence of slot markings, we find that the refinement process can detect unoccupied slots within $\pm 9^\circ$.

Furthermore, since the AVM system that we used for this experiment is a type of commercialized products, we were

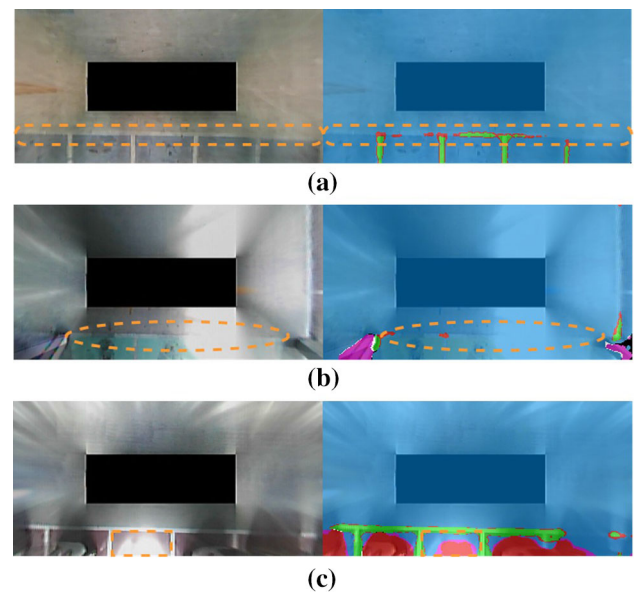


Fig. 10 Failure cases of the semantic segmentation

not able to adjust the field of view of the AVM image. To clearly identify whether any obstacles exist or not in parking spaces before starting automated parking, the lateral field of view should be extended from 6 to at least 12 m.

6.5 Comparison with previous research

In the research of parking space detection, there is no public dataset which would be helpful to accelerate research activities. Due to the lack of public dataset, it was difficult to compare the results qualitatively and quantitatively because

each proposition was evaluated using different datasets and different configurations. Therefore, we summarized previous research including performance and difference with the our proposed method. In [12], Wang et al. accomplished 97.1% precision and 81.5% recall in 2626 frames that have normal parking conditions. The method was not evaluated in various parking conditions and has the low recall rate. Furthermore, edge-based occupancy classification might be easily affected by falsely detected edges. In [5], Han et al. achieved 99.08% precision and 99.95% in 4024 parking spaces. The authors proposed specialized filter for slot marking and used dense motion stereo to detect vehicles. However, they did not perform the intensive evaluation for indoor parking conditions using only small dataset. The matching performance of dense motion stereo might be degraded in low illumination conditions. In [10], Suhr et al. obtained 97.4% precision and 99.2% recall in 265 parking spaces in indoor. The authors fused ultrasonic-based, pillar-based, and slot-marking-based approaches to improve recognition rate. They aimed to recognize the closed type of parking space having the guide line and handled three types of detection results for the fusion separately. The Sobel edge detection could generate many candidates in reflected light conditions of indoor; thus, the guide line estimation using RANSAC might be carefully tuned and the line tracking might be required. In [7], Lee et al. mentioned that precision is 97.9% and recall is 95.1% in 7790 slots. This method focused on slot marking detection without static obstacle detection. In [9], Suhr et al. extended their previous work for various types of parking space. Overall performance was 97.64% precision and 95.24% recall in 609 parking spaces. In [14], Zong et al. proposed the fusion method with AVM system and ultrasonic sensors; however, they had low recall rates of 80.18–90.97%. Comparing the previous research, our proposed method uses only the AVM system and can simultaneously detect slot marking and static obstacles in single image. Therefore, any fusion process is not required. In various parking conditions, we achieved 96.81% precision and 97.80% recall rates without tracking.

7 Conclusion

We employed deep-learning-based semantic segmentation on an AVM image. This method can immediately recognize free spaces, slot markings, vehicles, and other objects, including curbs, pillars, walls, and even stoppers. Therefore, in contrast to previous studies, we do not need a range sensor or 3D reconstruction algorithm that uses stereo matching. In particular, we showed that the semantic segmentation performed excellently even in indoor parking lots, which are among the challenging conditions. On the basis of the semantic segmentation, we developed a vertical-grid-based parking space detection. To detect a parking space, many previous

studies conducted primitive feature extractions of the parking slot markings. However, our proposed method does not need such operations. A 1-D search in the vertical grid can identify whether parking spaces are empty or not. Various parking conditions in 75 scenarios were experimentally validated using the proposed method. Precision and recall rates of 96.81% and 97.80%, respectively, were achieved.

From the experiments, we realized that several colors are used for parking slot markings, such as white, yellow, pink, and blue, depending on the parking lots. Pink and blue colors are difficult to distinguish from the floor color due to low intensity differences. Therefore, false negatives may occur in such environments. Although the AVM image that we used in the experiments could cover 360°, the covered area was slightly narrow compared with that identified by ultrasonic sensors. In particular, the height of the AVM image should be extended from 6 to at least 12 m. The current system provides a 37.5 mm/pixel resolution. If the image resolution is increased, the precision can be improved; however, more computational time and resources for processing are required.

One of the advantages of the proposed system is the unified structure for line filtering and obstacle detection. In contrast to previous studies, parking slot markings and static obstacles can be simultaneously recognized using the identical framework. Thus, integration or fusion process is not required. The framework can be obtained by end-to-end training. Further, the vertical grid encoding enables straightforward and fast occupancy determination. On the other hand, the proposed system requires large labeled datasets for the training to achieve robustness. Moreover, the vertical grid-based parking space detection can detect only two parking types: perpendicular and parallel parking.

In our future work, we plan to carry out parking space tracking based on semantic segmentation. Parking space tracking should be performed during parking after destination selection. However, because the designated parking space rotates and can be occluded by the ego-vehicle, a different approach from the parking space detection might be required. Furthermore, we plan to build maps of parking lots using sequential inference images of the semantic segmentation and on-board sensors in vehicles. The semantics in a inference image could be used as features for map matching in precise localization process as proposed in [6].

Acknowledgements This work was financially supported by the BK21 plus program (22A20130000045) under the Ministry of Education, Republic of Korea, the Industrial Strategy Technology Development Program (No. 10039673, 10060068, 10079961, 10080284), the International Collaborative Research and Development Program (N0001992) under the Ministry of Trade, Industry and Energy (MOTIE Korea), and National Research Foundation of Korea (NRF) grant funded by the Korean government (MEST) (No. 2011-0017495).

References

1. AVM (Around View Monitoring) System Datasets for Auto Parking. https://github.com/ChulhoonJang/avm_dataset (2017). Accessed Oct 2018
2. A Sudden Bang When Parking—Allianz. <https://goo.gl/DeE6UR> (2017). Accessed Oct 2018
3. Bertozzi, M., Broggi, A.: GOLD: a parallel real-time stereo vision system for generic obstacle and lane detection. *IEEE Trans. Image Process.* **7**(1), 62–81 (1998)
4. Fritsch, J., Kuhn, T., Geiger, A.: A new performance measure and evaluation benchmark for road detection algorithms. In: 16th International IEEE Conference on Intelligent Transportation Systems (ITSC), 2013, pp. 1693–1700. IEEE (2013)
5. Han, S.J., Choi, J.: Parking space recognition for autonomous valet parking using height and salient-line probability maps. *ETRI J.* **37**, 1220–1230 (2015). <https://doi.org/10.4218/etrij.15.0114.0112>
6. Jo, K., Jo, Y., Suhr, J.K., Jung, H.G., Sunwoo, M.: Precise localization of an autonomous car based on probabilistic noise models of road surface marker features using multiple cameras. *IEEE Trans. Intell. Transp. Syst.* **16**(6), 3377–3392 (2015). <https://doi.org/10.1109/TITS.2015.2450738>
7. Lee, S., Hyeon, D., Park, G., Baek, I.J., Kim, S.W., Seo, S.W.: Directional-DBSCAN: parking-slot detection using a clustering method in around-view monitoring system. In: Proceedings of the IEEE Intelligent Vehicles Symposium, vol. 2016, pp. 349–354. <https://doi.org/10.1109/IVS.2016.7535409> (2016)
8. Long, J., Shelhamer, E., Darrell, T.: Fully convolutional networks for semantic segmentation. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 3431–3440 (2015)
9. Suhr, J., Jung, H.: A universal vacant parking slot recognition system using sensors mounted on off-the-shelf vehicles. *Sensors (Switzerland)* **18**(4), 1213 (2018). <https://doi.org/10.3390/s18041213>
10. Suhr, J.K., Jung, H.G.: Automatic parking space detection and tracking for underground and indoor environments. *IEEE Trans. Ind. Electron.* **63**(9), 5687–5698 (2016). <https://doi.org/10.1109/TIE.2016.2558480>
11. Teichmann, M., Weber, M., Zoellner, J.M., Cipolla, R., Urtasun, R.: Multinet: real-time joint semantic reasoning for autonomous driving. *CoRR*. [arXiv:1612.07695](https://arxiv.org/abs/1612.07695) (2016)
12. Wang, C., Zhang, H., Yang, M., Wang, X., Ye, L., Guo, C.: Automatic parking based on a bird's eye view vision system. *Adv. Mech. Eng.* **6**, 847,406 (2014)
13. Zeiler, M.D., Fergus, R.: Visualizing and understanding convolutional networks. In: European Conference on Computer Vision, pp. 818–833. Springer (2014)
14. Zong, W., Chen, Q.: A robust method for detecting parking areas in both indoor and outdoor environments. *Sensors (Switzerland)* **18**(6), 1903 (2018). <https://doi.org/10.3390/s18061903>

Chulhoon Jang received B.S. degree in Mechanical Engineering and Ph.D. in automotive engineering from Hanyang University, Seoul, Korea, in 2011 and 2017, respectively. His main fields of interest are autonomous driving system, image processing, and deep learning. His current research activities include object detection and classification using multiple sensor fusion for urban autonomous driving.

Myoungcho Sunwoo received B.S. in Electrical Engineering from Hanyang University in 1979, M.S. in Electrical Engineering from the University of Texas at Austin in 1983, and Ph.D. in System Engineering from Oakland University in 1990. He joined General Motors Research (GMR) Laboratories, Warren, MI, in 1985 and has worked in the area of automotive electronics and control for 30 years. During his nine-year tenure at GMR, he worked on the design and development of various electronic control systems for powertrains and chassis. Since 1993, he has led research activities as a Professor with the Department of Automotive Engineering at Hanyang University. His work has focused on automotive electronics and controls, such as modeling and control of internal combustion engines, design of automotive distributed real-time control systems, intelligent autonomous vehicles, and automotive education programs.