Feasibility Study of Intersection Detection and Recognition Using a Single Shot Image for Robot Navigation

Takuto Watanabe, Kouchi Matsutani, Miho Adachi, and Takuro Oki
Dept. of Computer Science, Graduate School of Science and Technology, Meiji University, Japan
Email: {wttk, coach, miho, o_tkr}@cs.meiji.ac.jp

Ryusuke Miyamoto
Dept. of Computer Science, School of Science and Technology, Meiji University, Japan
Email: miya@cs.meiji.ac.jp

***Abstract*—This study is an attempt to actualize the autonomous movement of a robot using a navigation system with a camera, instead of expensive external sensors such as light detection and ranging. The present implementation of our approach basically consists of road-following, intersection detection, and intersection recognition, using the results of semantic segmentation. In this study, we focus on the accuracy improvement of the intersection detection and recognition. Classifiers are constructed for these tasks using deep neural networks. We evaluated the proposed classifier using three-dimensional computer graphics generated from the CARLA simulator and the Ikuta dataset composed of actual images that we took. The Experimental results demonstrated that the proposed system could detect and recognize intersections accurately; the F measure exceeded 0.96 for detection, and the actual images were recognized and classified with perfect accuracy.**

***Index Terms*—intersection detection, intersection recognition, semantic segmentation, robot navigation**

I. Introduction

Recently, it has become important to develop autonomous mobile robots and vehicles to expand the working area of robots, and reduce the rate of traffic accidents, respectively. Most of the current schemes for autonomous mobility [1]-[3] use environmental maps around the robot's movement area. To construct the environmental map, accurate three-dimensional (3D) sensors, such as LiDAR, are used. Then, localization is performed based on the correlation between the current sensing data obtained by sensors mounted on the robot and the prior-created 3D environmental map. Although this approach was effective in some scenes, it has some significant shortcomings, one of which is the expensiveness of accurate 3D sensors.

To reduce the price of the sensors required for autonomous mobility, the authors attempt to construct a novel scheme using only vision sensors [4]. The final goal of our project is to actualize the autonomous movement of robots using only a camera as an external sensor. In the current implementation of our scheme, a robot moves by performing the following two main procedures iteratively: road-following between intersections and route-changing at intersections. For these operations, a topological map [5] that has information on intersections and their connections is adopted in the scheme, instead of the widely used 3D metric map; localization is performed only at intersections.

The performance of the road-following operation based on the runnable region extracted through the semantic segmentation [6]-[8], with the additional modifications proposed in [9], was sufficient for practical applications. The accuracy of semantic segmentation is quite important; however, ICNet [8] can yield effective results in our scheme, if a suitable dataset is used for training [10]. To improve the robustness of the proposed approach it is necessary to enhance the accuracy of the intersection detection and recognition, because a robot changes its course at only intersections. Thus, if both are inaccurate, the robot inevitably loses its way.

Currently, our robot uses side cameras to detect intersections using the results of semantic segmentation. A robot can notice that the current location is an intersection where the center of a segmented image obtained from a side camera is runnable. In addition to the intersection detection, a correlation-based scheme using the results of the semantic segmentation proposed in [11] demonstrated moderate accuracy for intersection recognition. However, there still remains room for the accuracy improvement of intersection recognition.

The feasibility of deep learning is evaluated in this study as it shows remarkable results in the field of image recognition and intersection detection and recognition. To evaluate the detection accuracy of the proposed scheme, datasets are created using the CARLA [12] simulator to generate many kinds of images near intersections that are particularly useful for this purpose. The detection and classification accuracies are evaluated using actual

Manuscript received September 4, 2020; revised March 5, 2021.

doi: 10.18178/joig.9.2.39-44

images taken at our university campus and the synthetic images.

II. Related Work

This section explains our vision-based navigation scheme, as proposed in [4].

A. A Topological Map

Before explaining our vision-based navigation scheme, this subsection introduces a topological map that is not popular in autonomous mobility. The topological map used in our scheme has information on intersections and their connections. Therefore, it can be represented using an undirected graph, where intersections and connections can be represented by nodes and edges, respectively. Fig. 1 shows an example of a topological map corresponding to an actual map obtained from Google map.

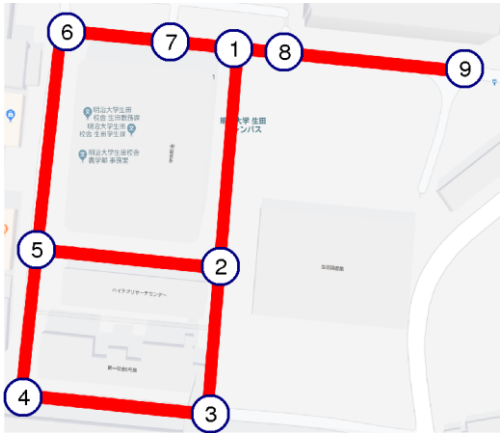


Figure 1. An example of a topological map corresponding to an actual map.

B. Visual Navigation Based on a Topological Map

The visual navigation scheme based on a topological map using the results of the semantic segmentation enables robots to move autonomously by performing the following operations iteratively:

1. Intersection detection,
2. Intersection recognition,
3. Change of course based on a topological map and the intersection recognition result, and
4. Road-Following to the next intersection.

These operations are discussed in details in the rest of this section.

C. Road-Following Using Results of Semantic Segmentation

The road-following scheme based on the results of the semantic segmentation controls a robot using a target point defined in an input image, as determined from the runnable area extracted using semantic segmentation. In our present implementation, we adopted the ICNet after considering the processing speed and the segmentation accuracy. Fig. 2 shows an example of the target point for robot control. The road-following scheme attempts to keep the control point at the center of input image based on the runnable area extracted by the ICNet. This procedure includes obstacle avoidance; therefore, further object detection is not necessary for autonomous movement.

Figure 2. A target point for robot control computed using the result of semantic segmentation.

D. Intersection Detection

In our present implementation, intersections are detected using the segmented images obtained from the side cameras. The detection is based on simple thresholding; the current location is classified as an intersection when the number of pixels classified as runnable regions in a rectangular region that is defined prior exceeds the pre-defined threshold. Fig. 3 and Fig. 4 show sample images corresponding to intersection detection.

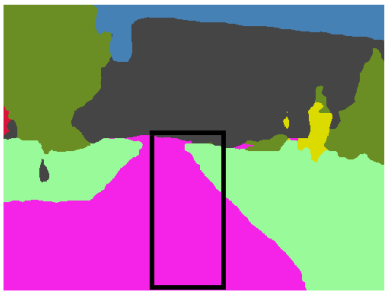


Figure 3. A segmented image obtained from a side camera just before a robot entering to intersection.

Figure 4. A segmented image obtained from a side camera at the inside of intersection.

This simple scheme is effective, because there must be runnable areas corresponding to the intersecting roads that can be observed by the side cameras. However, this scheme has a significant shortcoming, the computational complexity owing to the semantic segmentation of images obtained from the side cameras. To reduce the computational complexity required for intersection detection, this study attempts to construct a novel scheme that uses only images obtained from the frontal camera.

E. Intersection Recognition

Once an intersection is detected, the robot must recognize the current location based on a topological map given to the robot prior to autonomous running. To perform this process, the robot must classify the intersection at the current location using the input images. Our current implementation adopts a correlation-based scheme using the results of semantic segmentation [11].

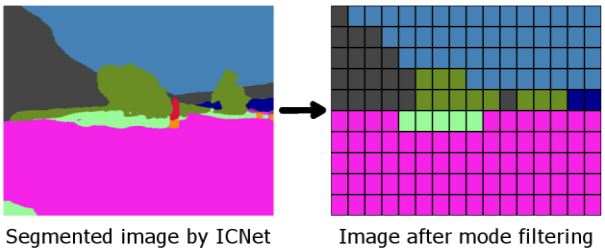


Figure 5. Feature extraction for intersection recognition proposed in [11].

In this scheme, a segmented image is divided into several rectangular regions, and the mode filter is applied to these regions. After these operations, we can obtain an image, as shown in Fig. 5. Correlation for intersection classification is simply computed using feature vectors obtained directly from the image generated by the previous operations. Here, the weighting coefficients are multiplied by the feature vectors according to class labels before correlation computation, to reduce the influence of the runnable area.

The classification accuracy of this scheme is acceptable for short courses with few intersections. However, this scheme may not be sufficiently accurate when the navigation area of the robot becomes larger. Therefore, the classification accuracy should be improved, and for this reason, we attempt to construct a scheme that is more accurate than the present implementation in this study.

III. Intersection Detection by Deep Neural Networks

This section explains how to detect intersections using deep learning, and evaluates the detection accuracy of the proposed schemes using two kinds of datasets: the CARLA and Ikuta datasets.

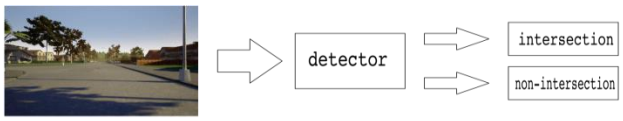


Figure 6. Operation flow of intersection detection.

A. How to Detect Intersections

In this study, the authors attempt to construct an intersection detector using a two-class classifier; the trained classifier determines whether there is an intersection or not. Fig. 6 shows the operation flow of the proposed classifier. To construct an effective classifier appropriately, positive and negative samples corresponding to intersections and non-intersections are created, as detailed in the following subsections. In this study, the effectiveness of the architecture of VGG16, Resnet-50, Resnet-101, Resnet-152, and Inception v3 as two-class classifiers were verified.

B. CARLA Dataset

CARLA [12] is a simulator developed for the autonomous mobility of vehicles; it is also applicable to robots. This simulator can reproduce things ranging from roads, signals, buildings, cars, to humans on a virtual space rendered by the three-dimensional Computer Graphics (CG). We reiterate at this juncture that it is possible to obtain arbitrary views using a camera mounted on a robot.

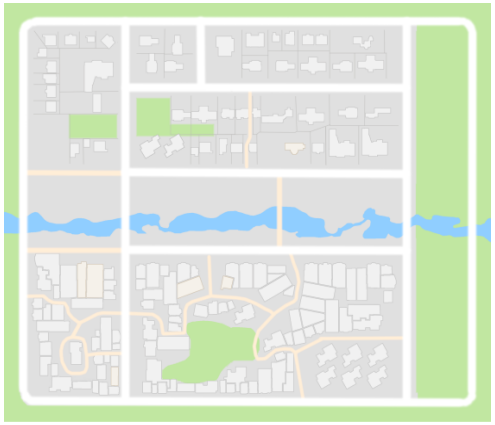


Figure 7. A map for dataset creation using the CARLA simulator.

To create positive and negative samples, the border between the intersections and non-intersections were set according to the distance: regions within 5.5 m from the intersections were defined as positive. As may be seen on the map in Fig. 7, twelve intersections were to be detected.

The positive samples were generated from several images based on the simulation of the view of a robot, and randomly varying the following parameters: location, view angle, and the weather. Here, the nearest distance from the intersections was set to 1 m for the positive samples, because an intersection must be detected before a robot enters it. The yaw angle of the camera was set to the range of -20 to 20 degrees, simulating an actual situation. The weather was varied according to three conditions: sunny, cloudy, and rainy. All the parameters, except the location, were also varied to generate the negative samples.

Subsequently, the CARLA dataset composed of 4320 images, with equal numbers of positive and negative samples, was created. The training samples included in the dataset are shown in Fig. 8.

Figure 8. Example of the images in the CARLA dataset created by the authors.

C. Ikuta Dataset

The Ikuta dataset using actual images taken at the Ikuta campus of Meiji University was created to evaluate the detection accuracy of the proposed scheme. For this dataset, several images of a course with eight intersections was captured, as shown in Fig. 1. The course terminated at the ninth intersection. There were 418 and 417 positive and negative samples, respectively. Both were classified as determined by humans. Fig. 9 shows the example of the images in the Ikuta dataset.

Figure 9. Example of the images in the Ikuta dataset.

D. Training of Deep Neural Networks

For evaluating the classification accuracy using the created dataset, five kinds of deep neural networks (DNNs) were trained as two-class classifiers: VGG16 [13], Resnet-50, Resnet-101, Resnet-152 [14], and Inception v3 [15]. They were all implemented using Keras [16]. The size of the input images was 224×224. For the optimizer, we adopted a Momentum-SGD [17] with the following parameters: 0.001 for the learning rate, 0.9 for the momentum, 32 for the batch size, and cross entropy for the loss function. For the training process, we adopted fine-tuning from the pre-trained models using ImageNet [18].

E. Evaluation

Table I and Table II show the evaluation results of intersection detection using the CARLA and Ikuta dataset, respectively. As can be seen, all the networks achieved quite good results, with the Resnet-50 or Resnet-101 and Resnet-152 being the most effective on the CARLA and Ikuta dataset, respectively. However, the results by the VGG16 was also good, although its structure is much simpler than Resnet's. Considering that precision is the most important factor in our scheme, the VGG16 appears to be the best network on the basis of the results of the Ikuta dataset; false positives may be removed during the intersection classification performed after detection.

TABLE I. EVALUATION RESULTS OF INTERSECTION DETECTION USING THE CARLA DATASET

	Precision	Recall	F measure
VGG16	0.987	0.996	0.991
ResNet-50	0.985	0.996	0.991
ResNet-101	0.991	0.993	0.992
ResNet-152	0.989	0.994	0.992
InceptionV3	0.985	0.994	0.990

TABLE II. EVALUATION RESULTS OF INTERSECTION DETECTION USING THE IKUTA DATASET

	Precision	Recall	F measure
VGG16	0.995	0.939	0.966
ResNet-50	0.910	0.989	0.948
ResNet-101	0.985	0.938	0.961
ResNet-152	0.965	0.970	0.967
InceptionV3	0.970	0.956	0.963

IV. INTERSECTION RECOGNITION BY DEEP NEURAL NETWORKS

This section explains how to classify intersections using DNNs, and evaluates the classification accuracy of the proposed classifier using two datasets, CARLA and Ikuta.

A. How to Classify Intersections

Similar to the intersection detection, intersection classification was performed using only a single-shot image taken with a camera mounted on a robot, the only difference being that a multi-class classifier was constructed for intersection classification. Fig. 10 depicts the process by which the intersection classifier recognizes intersections.

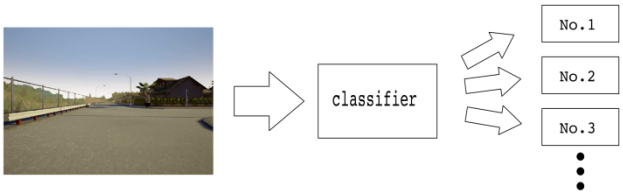


Figure 10. Intersection classification by a multi-class classifier.

B. CARLA Dataset for Intersection Recognition

The training samples for intersection recognition were created in a similar way to the detection dataset, the only difference being the distance of the viewpoints from the intersections. In this case, only viewpoints nearer than 5.5 m were used for dataset generation, because intersection classification is performed only when an intersection is found through the detection process. The total number of samples included in the dataset was 4320, 360 for each intersection.

C. Ikuta Dataset for Intersection Recognition

For the intersection recognition experiment, the Ikuta dataset was used. The Ikuta dataset was created in a similar way to the CARLA dataset, the only difference being how the distance from the intersections was measured: only images classified as near samples to the intersections by a human were used in the classification dataset. Consequently, the total number of images included in the dataset was 807, with approximately 100 being selected for each intersection.

D. Evaluation

The multi-class classifiers based on the VGG16 [13], Resnet-50, Resnet-101, Resnet-152 [14], and Inception v3 [15] were created in the same way as the intersection detection. Only the training samples were different.

Table III shows the classification accuracy of these models. As can be seen, perfect classification was achieved by all the models. However, only a single test sample was classified incorrectly in the Ikuta dataset by Resnet-152, as shown in Table IV. Fig. 11 shows the incorrectly classified image. This may be attributed to the number of training samples being too few for the Resnet-152 that had more layers than the other models.

TABLE III. CLASSIFICATION ACCURACY OF INTERSECTIONS USING THE CARLA DATASET

	Accuracy
VGG16	1.00
ResNet-50	1.00
ResNet-101	1.00
ResNet-152	1.00
InceptionV3	1.00

TABLE IV. CLASSIFICATION ACCURACY OF INTERSECTIONS USING THE IKUTA DATASET

	Accuracy
VGG16	1.00
ResNet-50	1.00
ResNet-101	1.00
ResNet-152	0.997
InceptionV3	1.00

Figure 11. The image incorrectly classified by Resnet-152.

V. CONCLUSION

The authors attempted to actualize a visual navigation scheme using a single camera as the external sensor. For this purpose, it is essential to detect and recognize intersections using a single-shot image while a robot moves autonomously. To confirm the applicability of DNNs to intersection detection and recognition, the intersection detection and recognition accuracy using the CARLA and Ikuta datasets were evaluated in this study. Experimental results showed that VGG16, Resnet-50, and Resnet-152 demonstrated good accuracy for intersection detection, and nearly perfect results were achieved by all the models used in the evaluation. These results also verified the effectiveness of DNNs for intersection detection and recognition.

In the future, to implement a practical system for intersection detection and recognition, we will focus on the following tasks: creating a smaller architecture for this task, investigating the required number of intersections to be classified with the support of a topological map, creating a single multiclass classifier to merge detection and recognition into a single process, and evaluating the proposed classifier based on real-time experiments.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

AUTHOR CONTRIBUTIONS

All authors conducted the research; T. Watanabe, T. Oki, and R. Miyamoto wrote the paper; all authors had approved the final version.

ACKNOWLEDGEMENT

This research was partly supported by Research Project Grant (B) by Institute of Science and Technology, Meiji University.

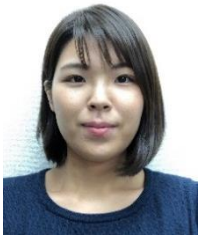
REFERENCES

[1] D. Hahnel, W. Burgard, D. Fox, and S. Thrun, "An efficient fastSLAM algorithm for generating maps of large-scale cyclic environments from raw laser range measurements," in *Proc. IEEE/RSJ International Conference on Intelligent Robots and Systems*, 2003, vol. 1, pp. 206-211.
[2] H. Durrant-Whyte and T. Bailey, "Simultaneous localization and mapping: Part I," *IEEE Robotics Automation Magazine*, vol. 13, no. 2, pp. 99-110, June 2006.
[3] S. Thrun, M. Montemerlo, H. Dahlkamp, D. Stavens, A. Aron, J. Diebel, *et al.*, "Stanley: The robot that won the DARPA grand challenge," *Journal of Field Robotics*, vol. 23, no. 9 pp. 661-692, 2006.
[4] R. Miyamoto, Y. Nakamura, M. Adachi, T. Nakajima, H. Ishida, K. Kojima, *et al.*, "Vision-Based road-following using results of semantic segmentation for autonomous navigation," in *Proc. International Conference on Consumer Electronics*, Berlin, 2019, pp. 194-199.
[5] B. J. Kuipers, "Modeling spatial knowledge," *Cognitive Science*, vol. 2, no. 2 pp. 129-153, 1978.
[6] H. Zhao, J. Shi, X. Qi, X. Wang, and J. Jia, "Pyramid scene parsing network," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2017, pp. 2881-2890.
[7] H. Zhao, Y. Zhang, S. Liu, J. Shi, C. Change Loy, D. Lin, *et al.*, "PSANet: Point-Wise spatial attention network for scene parsing," in *Proc. European Conference on Computer Vision*, 2018, pp. 267-283.
[8] H. Zhao, X. Qi, X. Shen, J. Shi, and J. Jia, "ICNet for real-time semantic segmentation on high-resolution images," in *Proc. European Conference on Computer Vision*, 2018, pp. 418-434.
[9] M. Adachi, S. Shatari, and R. Miyamoto, "Visual navigation using a webcam based on semantic segmentation for indoor robots," in *Proc. IEEE Conf. Signal Image Technology and Internet Based Systems*, 2019, pp. 15-21.
[10] R. Miyamoto, M. Adachi, Y. Nakamura, T. Nakajima, H. Ishida, and S. Kobayashi, "Accuracy improvement of semantic segmentation using appropriate datasets for robot navigation," in *Proc. International Conference on Control, Decision and Information Technologies*, 2019, pp. 1610-1615.
[11] H. Ishida, K. Matsutani, M. Adachi, S. Kobayashi, and R. Miyamoto, "Intersection recognition using results of semantic segmentation for visual navigation," in *Proc. International Conference on Computer Vision Systems*, 2019, pp. 153-163.
[12] A. Dosovitskiy, G. Ros, F. Codevilla, A. Lopez, and V. Koltun, "CARLA: An open urban driving simulator," in *Proc. 1st Annual Conference on Robot Learning*, 2017, pp. 1-16.
[13] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," in *Proc. International Conference on Learning Representations*, 2015.
[14] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2016, pp. 770-778.
[15] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, "Rethinking the inception architecture for computer vision," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2016, pp. 2818-2826.
[16] F. Chollet, *et al*. (2015). Keras. [Online]. Available: https://keras.io
[17] D. E. Rumelhart, G. E. Hinton, and R. J. Wilson, "Learning representations by back-propagating errors," *Nature*, vol. 323, no. 6088, pp. 533-536, 1986.

[18] J. Deng, W. Dong, R. Socher, L. Li, K. Li, and F. Li, "Imagenet: A large-scale hierarchical image database," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2009, pp. 248-255.

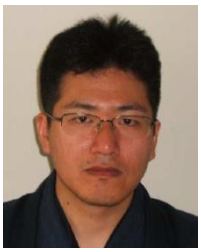
Takuto Watanabe received the B.S. degree in computer science from Meiji University, Japan, in 2019. His research interest includes reinforce learning for robot navigation using only visual information, image-based sensing using machine learning, etc.
He is currently a master course student.

Kouchi Matsutani received the B.S. degree in computer science from Meiji University, Japan, in 2019. His research interest includes intersection detection and classification for robot navigation in urban scenarios, deep learning for image classification, etc.
He is currently a master course student.



Miho Adachi received the B.S. degree in computer science from Meiji University, Japan, in 2019. Her research interest includes visual navigation of autonomous robot, localization of a robot using only visual information, machine learning for robot applications.
She is currently a master course student. She is a member of IEEE and SICE.

Takuro Oki received the B.S. degree, the M.S degree in, and the D.S. degree in computer science from Meiji University, Japan, in 2015, 2017, and 2020, respectively. His research interest includes visual object detection based on machine learning, parallel and real-time implementation of image processing applications, machine learning for sports applications, etc. He is currently an engineer at DWANGO corporation. He is a member of IEEE and IEICE.



Ryusuke Miyamoto received the B.E. degree in industrial chemistry, the M.Sc. degree in communications and computer engineering, and the Ph.D. degree in communications and computer engineering from Kyoto University, Kyoto, Japan, in 1998, 2001, and 2007, respectively. He is currently a senior assistant professor in the school of science and technology, Meiji University. He is a member of IEEE, IEICE, IIEEJ, IPSJ, and SICE.