2018 IEEE International Conference on Systems, Man, and Cybernetics

3D Semantic Mapping in Greenhouses for Agricultural Mobile Robots with Robust Object Recognition using Robots' Trajectory

Shigemichi Matsuzaki, Hiroaki Masuzawa, Jun Miura and Shuji Oishi
Department of Computer Science and Engineering, Toyohashi University of Technology

***Abstract*—This paper describes a method of building a semantic map of a greenhouse for a robot path planning. Existing mapping methods only consider whether there are obstacles in a certain region. They are not sufficient for path planning in greenhouses where traversable regions are often covered by branches and leaves which are also recognized as obstacles. We propose a mapping method which generates a map with semantic information on the types of obstacles. By integrating RGB-D based visual SLAM (Simultaneous Localization And Mapping) and semantic segmentation by a deep neural network, we obtain a 3D map with semantic labels. In order to deal with the uncertainty of observations, we introduce a Bayesian label updating strategy which effectively utilizes the fact that the robot traverses a region. Through evaluations, we confirmed that the proposed method can perform a more accurate semantic labeling than the one only using SegNet.**

***Index Terms*—Agricultural robot, Mapping, Environment recognition**

I. Introduction

In recent years, agriculture has been facing various problems. Decreasing number of farmers and their aging are particularly serious problems. The number of people mainly engaged in farming in Japan was 1,571,100 in 2017, down by 79,000 (5.0%) from 2016 and by 246,700 (14.1%) from 2015. Moreover, 66.4% of the farmers are over 65 years old [1]. In the future, further manpower shortage will become more severe as this generation retires. Against this backdrop, labor reduction and improvement of production efficiency in agriculture are important issues.

As measures against such a situation, "smart agriculture" is being promoted in Japan, which applies RT (robot technology) and ICT to agriculture. Examples are autonomous driving of vehicles such as tractors and rice transplanters [2] and autonomous harvesting robots of fruits [3]. We are developing a mobile robot for supporting agricultural work in a greenhouse for flower cultivation [4]. This robot has an autonomous movement and a person tracking function, and supports workers in harvesting and transportation.

One of the functions necessary for an autonomous mobile robot is generation of a map for path planning. In greenhouses, paths are often covered by plants. While humans can recognize such paths to be traversable and go through by pushing the plants aside, robots with conventional mapping methods which recognize every object as an obstacle cannot recognize such traversable paths.

In this paper, we propose a method of generating a 3D map with object type information. A generated map has probabilistic semantic information on the type of objects calculated by image-based object recognition results and past robot trajectories. This realizes a path planning considering not only the presence of obstacles but also the traversability of each object.

The rest of the paper is organized as follows. Section II describes existing studies regarding with agricultral mobile robots. Section III describes the problems in mapping inside greenhouses and the elements necessary to solve them. Section IV describes the details of the proposed method. Section V shows the result of map generation by the proposed method and discusses its validity. Section VI summarizes this research and discusses future work.

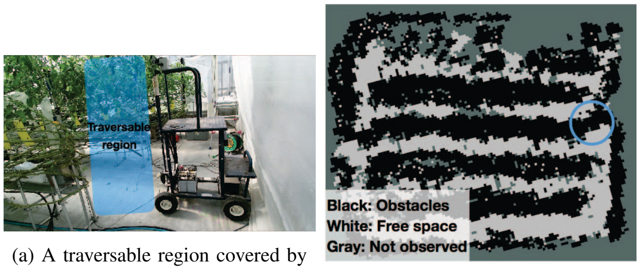
II. Related work

In this section, we describe relevant studies focusing on navigation and SLAM methods in agricultural robots and general semantic mapping, as well as applications of semantic mapping in agricultural robots.

There have been numerous studies in navigation, localization and mapping for autonomous agricultural robots as well-summarized in [2]. Nørremark et al. proposed a method of navigating a tractor using real-time kinematic GPS (RTK-GPS) and Kalman filter which enables it to perform real-time positioning and control of side-shift and cycloid hoe [5]. Kise et al. proposed a method which integrates RTK-GPS and a Fiber Optical Gyroscope (FOG) as navigation sensors [6]. When it comes to localization of robots inside greenhouses, on which we focus in this research, GPS is not suitable in terms of accuracy since there may be variety of physical barriers

Auat Cheein et al. proposed a SLAM method for precision agriculture mapping in olive groves [7]. This method detects olive stems utilizing both range data from a laser range finder and visual data from a monocular camera, and process a SLAM algorithm based on Extended Information Filter using the detected stems as landmarks. Shalal et al. proposed a mapping method for mobile robots targeting orchards [8]. Their method simply maps trees and non-tree objects detected by a camera and a laser scaner using the Extended Kalman Filter-based SLAM, also utilizing the trees as landmarks. While those methods are optimized for their specific operating environments, their assumptions of almost regularly arranged

2577-1655/18/$31.00 ©2018 IEEE
DOI 10.1109/SMC.2018.00070



(a) A traversable region covered by plant

(b) Two-dimensional map

Fig. 1. 2D map inside a greenhouse generated by conventional mapping method

trees and object-free paths are not suitable for the general autonomous agricultural robot we are seeking.

Bernuy et al. proposed a method of semantic mapping in outdoor environment aiming at the application in autonomous off-road driving [9]. In their method, 2D large-scale topological map is built using semantic image information. Weiss et al. proposed a method of semantic classification and mapping in agricultural fields [10]. In this work, they partition an agricultural field into several location classes such as *open field*, *row*, *row start* and *row end*. The method uses a low-resolution 3D laser sensor and classifies the sensor data into the classes. Like all of the work mentioned above, however, the map generated by their method does not consider any traversable objects.

III. Map for autonomous robots in greenhouses

In greenhouses for horticulture, paths are often covered by plants. In the case of rose cultivation, for example, branches emerging in the early stage of growth are artificially bent down as assimilation shoots for effective photosynthesis [11]. The bent assimilation shoots are spread out over the paths. The paths could also be covered by branches and leaves in the case of other crops. When moving on the paths, it is necessary for a robot to push those branches and leaves or step on them. However, conventional mapping methods consider only the presence of obstacles [12] [13] and regions without obstacles are treated as traversable regions. Crop harvesting is generally carried out with people entering between rows of plants. Therefore, a path planning function considering the traversable regions inside the greenhouses is important for autonomous mobile robots for supporting farmers.

Humans can recognize that the leaves and branches can be pushed aside or stepped on, and walk through the paths covered by plants by pushing them. In order to realize such inference on the robot, the following elements are necessary.

- Object recognition to distinguish possibly traversable objects such as plants.
- Knowledge on the likelihood that the region can be traversed according to a model of traversability for each object.

Probabilistic information on the type of the objects for each region is given to the map.

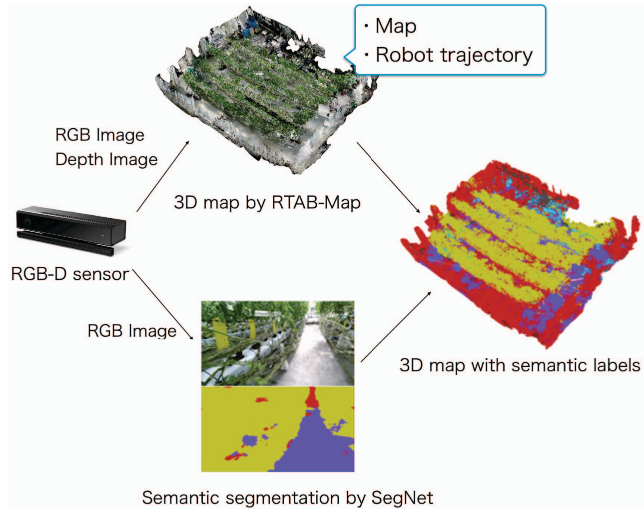


Fig. 2. Flow of labeling based on the kind of object

IV. Proposed method

A. Overview

We first generate 3D map by RTAB-Map [14] using RGB and depth images taken by Kinect v2 traveling through a greenhouse. SegNet [15] labels the type of object on each pixel of the RGB images. By associating the map given as point clouds generated by RTAB-Map with the object labels estimated by SegNet, we obtain a 3D map labeled by the types of objects (see Fig. 2).

The labeling by SegNet has uncertainty of result due to recoginition failures. In order to deal with the uncertainty, the object labels are treated probabilistically. At first, we segment the labeled 3D map into voxels. And then the probability of objects in each voxel is calculated from the frequency of the labels in the voxel.

In addition, we define the likelihood of traversability for each type of object in a heuristic manner. For voxels which the robot has traversed, the probabilities of objects are updated by Bayes' inference using the priors of the types of objects and the likelihood on trversability.

B. 3D Mapping

RTAB-Map [14] is used for generating a map. RTAB-Map is a visual SLAM (Simultaneous Localization And Mapping) methods which carries out camera localization and mapping simultaneously using data from RGB-D sensors. Using RTAB-Map, we obtain map which consists of an RGB image, a depth image and the camera pose at each location where the data have been obtained, and estimated trajectory of the camera. We can generate 3D point cloud map from those data.

C. Semantic labeling of 3D map

We first process semantic labeling of an RGB image from Kinect by SegNet [15]. SegNet is an encoder-decoder type

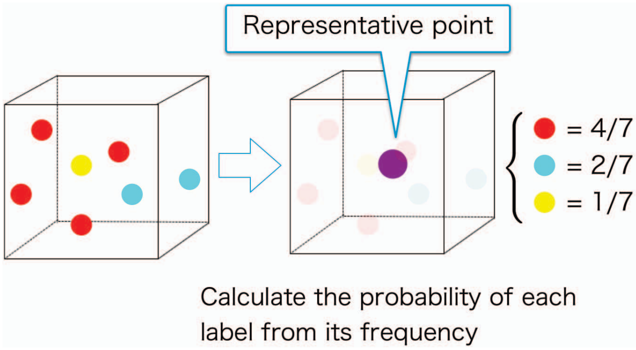


Fig. 3. Overview of voxelization of 3D map

deep neural network architecture for pixel-wise semantic labeling of an RGB image. We use a model trained by CamVid dataset [16] consisting of images of outdoor environment. Twelve object classes are defined in the model. Because some of the classes are not likely to exist in greenhouses (e.g., *Signs*, *Bikes* etc.), we limit the classes to the following three; *Building*, *Tree* and *Pavement* by normalizing the likelihood of those classes in the output and utilizing only them. The labels on the 2D image are then mapped onto the corresponding 3D points.

D. Voxelization and generating a histogram of the labels

The labeling by SegNet includes uncertainty due to recognition failures. With a naive mapping of 2D labels onto 3D point cloud, the result with such false recognition is simply reflected and may affect the estimation of traversability of the regions in a path planning based on the object types of regions. In order to realize robust environment recognition under such uncertainty, we divide the 3D space into voxels, and calculate the probabilities of objects of the sub-region from the frequency of each label.

A voxel is a cube with the edge length of 0.2 [m]. For each voxel, we calculate frequencies of labels of 3D points within it. By normalizing the frequencies, we obtain the probability $P(o)$ that the voxel is a certain object o as follows:

$$P(o) = \frac{N_o}{\sum_i N_i}, \quad (1)$$

where N_i represents the number of points with label i within the voxel. Fig. 3 illustrates the process of voxelization.

E. Label refinement by Bayesian inference

We can say that the objects in regions where a robot has traversed are more likely to be plants than to be the other two objects. For such an inference, we utilize the fact that the robot has traversed a region as evidence, and update the label probabilities by Bayesian inference. We assume the robot's shape to be a cuboid. Voxels within the cuboid at each node of the map can be considered as traversed voxels. Using that information, the probabilities of labels of the voxels are updated by following Bayes' rule:

TABLE I
LIKELIHOOD THAT OBJECT o IS TRAVERSABLE

	Building	Tree	Pavement
$P(\tau \mid o)$	0.05	0.6	0.05

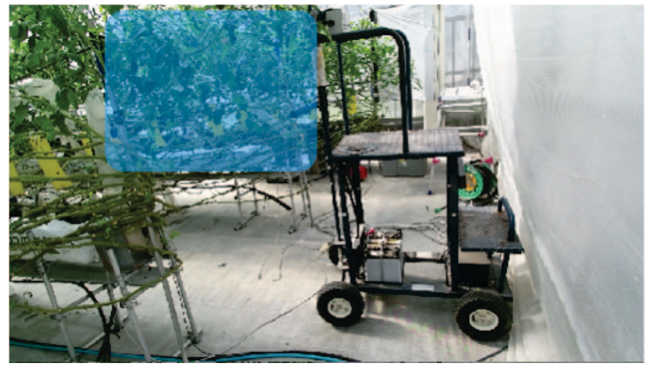


Fig. 4. A path covered by plant in a greenhouse at Toyohashi University of Technology. The region shown in a blue rounded rectangle is covered by the branches spreading out from the cultivation device on the left.

$$P(o \mid \tau) = \alpha P(\tau \mid o) P(o), \quad (2)$$

where o is a type of object, τ is an event that the voxel has been traversed, $P(o \mid \tau)$ is the posterior of object o after the event, $P(\tau \mid o)$ is the likelihood that object o is traversable, and α is a normalization constant. The prior $P(o)$ can be calculated by eq. (1). We define the likelihood $P(\tau \mid o)$ as shown in Table I. These values are heuristically set based on our knowledge that plants can sometimes be traversed by robots while buildings and pavements are usually not traversable.

V. EXPERIMENT

A. Experiment method

Our mapping method is implemented on ROS (Robot Operating System) [17]. ROS is a middleware which supports building robot applications by connecting units of functions called nodes.

In our experiment, we used data from the sensor recorded by rosbag, a functionality of ROS which records and replays ROS messages [18]. This allows us to reproduce the situation where the sensor data are recorded.

We recorded the data in a greenhouse for tomatoes at Toyohashi University of Technology. Some parts of paths were covered by leaves and branches (see Fig. 4).

B. Equipments

We used Kinect v2 as an RGB-D sensor. Kinect v2 is able to take depth data within the range of 0.5 to 8.0 [m] from the sensor, as well as 1920×1080 RGB images. The specification of the PC we used in our experiment is shown in Table II.

TABLE II
SPECIFICATION OF THE PC

OS	Ubuntu 16.04
Memory	32GB
CPU	Intel Core i7-6700HQ
GPU	Nvidia GeForce GTX 970M

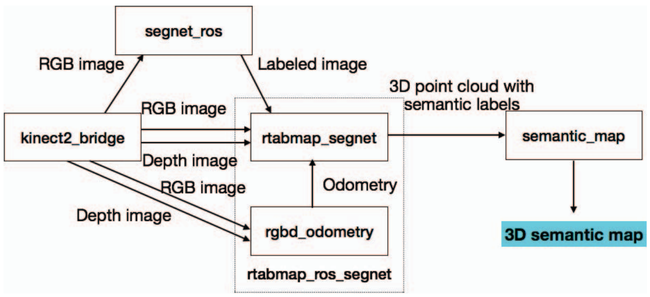


Fig. 5. Relation of the nodes

C. Software configuration

The nodes we developed are as follows.

- rtabmap_segnet
  A node to generate map data. This node is based on rtabmap node in rtabmap_ros package [19]. We modified the original rtabmap node so that it subscribes to label images from SegNet as well as RGB and depth images and odometry. It publishes a message *mapData* which includes RGB images, labeled images, depth images, camera poses etc. recorded at all the observation points.
- segnet_ros
  A node to use SegNet in ROS environment. It subscribes to RGB images and publishes semantically labeled images.
- semantic_map
  A node for constructing a 3D map with semantic labels. This subscribes to mapData published by rtabmap_ros_segnet and store them. After all mapData has been published, it processes the voxelization and the label the refinement using Bayesian inference, and publishes the resulting point cloud map.

Fig. 5 illustrates the entire structure of the software.

D. Result

Fig. 6 shows the result of the label update. The first row is a part of the 3D map generated with the original RGB images, the second row is the result of the labeling without the refinement, and the third row is the labels after the refinement by Bayesian inference. The colors of the voxels represent the label of the highest probability, red as Building, blue as Pavement and yellow as Tree. The regions indicated by the circles in the figures are areas which are covered by plants and the robot has traversed. While some of the voxels are misclassified as Building in the second row, the corresponding voxels in the third row have been changed to Plant. This means that more accurate classification is made using the observation of the robot traversing those voxels.

TABLE III
PROPORTION OF VOXELS WHOSE LABEL WITH THE HIGHEST PROBABILITY WAS CHANGED

Turned to Tree	Building and Pavement	Proportion
819	4244	0.193

Table III shows how many of the traversed voxels with the highest probability of Building or Pavement were changed to Plant, and their proportion against the voxles that are originally Building or Pavement. There were 819 voxels turned to Tree after the refinement out of 4244 voxels being Building or Pavement before the refinement.

TABLE IV
AVERAGE PROBABILITIES OF OBJECTS WITH AND WITHOUT THE REFINEMENT

	Tree	Building	Pavement
Without the label refinement	0.545	0.294	0.161
With the label refinement	0.625	0.235	0.140

Table IV shows the average probability of each object among the all voxels traversed by the robot. The avarage probability of Tree increased after the refinement while those of Building and Pavement decreased by approximately 20 [%] and 13 [%] respectively. Although the label refinement worked as expected by turning the traversed Building and Pavement to Tree, a significant number of traversed voxels remained Building or Pavement. This is because the probability of a certain object was significantly higher than of the other two, with some of those even having one on one object type and zero on the others, and thus the posterior of Tree could not be higher than the object.

The computational times with and without the label refinement are shown in Table V.

VI. CONCLUSIONS AND FUTURE WORK

We have proposed a method of generating a 3D map with semantic labels for autonomous robots used in greenhouses by integrating the 3D mapping and semantic labeling of 2D images. In order to realize robust environment recognition against observations with ambiguity, our method voxelizes the 3D map and calculate the probability of each object for each voxel from the frequency of the labels. Furthermore, the probabilities of voxels traversed by the robot are updated by Bayes' inference.

TABLE V
COMPUTATIONAL TIME (THE AVERAGE OF THREE TRIALS EACH)

Without the label refinement	With the label refinement
61,285[msec]	140,718[msec]

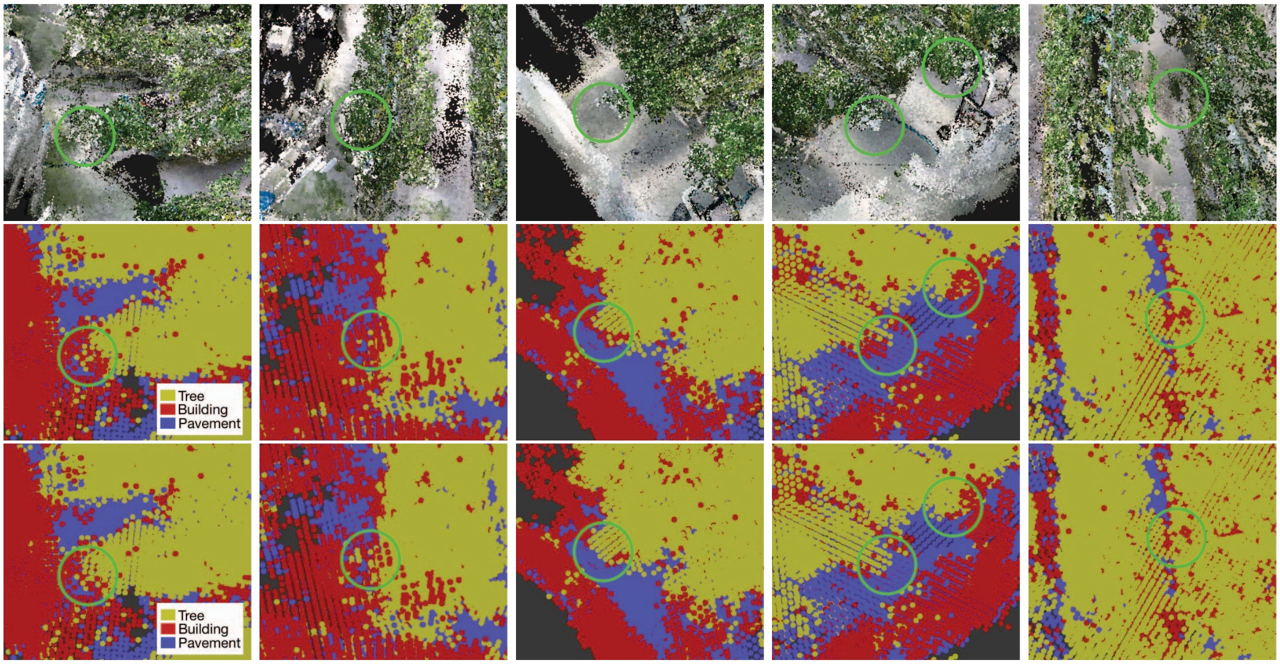


Fig. 6. The result of label refinement based on Bayesian inference

We conducted an experiment to generate a map from the data recorded in a greenhouse with paths partly covered by plants. We have shown that the proposed method can perform a more accurate semantic labeling than the one only using SegNet.

As future work, we are considering the following;

- Improving the method of generating histograms: In the proposed method, we generated the histograms of object labels in each voxel. It resulted in overly biased histograms with probabilities of zero on a object or two, which cannot be properly updated in our method. By utilizing the likelihood of the labels in SegNet in which a value other than zero is assigned for each label, this problem could be avoided.
- Designing a probabilistic model considering physical attributes of objects: Basically, it is difficult for robots to estimate the traversability of regions using only object type information. For example, branches and leaves of plants are flexible and likely to be traversable while their stem is usually not traversable. Such a fact indicates that the traversability of a region highly depends on physical attributes of each part of objects. We need to extend our method by taking them into consideration.
- Developing object recognition specialized for greenhouses: In the experiment, we used a pre-trained model for outdoor environment in the object recognition. Although the model showed quite good accuracy especially for the plants, there are a lot of recognition failure which cannot be updated by our method. By using a model trained for scenes inside greenhouses, the accuracy of the object recognition could be improved. We should also consider the image segmentation method other than pixel-wise ones, since our system does not require such detailed classification.

ACKNOWLEDGMENT

This work is supported by Knowledge Hub Aichi's Priority Research Project (2nd term). The authors would like to thank Mr. Mitsuo Tsume (Sinfonia Technology Co. Ltd.) for fruitful discussion. Aichi agricultural research center has cooperated with the experiment in this work.

REFERENCES

[1] Statistics Department of Ministry of Agriculture Forestry and Fisheries. Survey of movements in agricultural structure. Technical report, Ministry of Agriculture, Forestry and Fisheries, 2017.

[2] M. Li, K. Imou, K. Wakabayashi, and S. Yokoyama. Review of research on agricultural vehicle autonomous guidance. *Int. J. Agric. Biol. Eng.*, 2(3):1–16, 2008.

[3] K. Tanigaki, T. Fujiura, A. Akase, and J. Imagawa. Cherry-harvesting robot. *Computers and Electronics in Agriculture*, 63(1):65–72, 2008.

[4] H. Masuzawa, J. Miura, and S. Oishi. Development of a mobile robot for harvest support in greenhouse horticulture – person following and mapping. In *Proceedings of IEEE/SICE Int. Symp. on System Integration*, pages 541–546, 2017.

[5] M. Nørremark, H. Griepentrog, J. Nielsen, and H. Søgaard. The development and assessment of the accuracy of an autonomous gps-based system for intra-row mechanical weed control in row crops. *Biosystems Engineering*, 101(4):396 – 410, 2008.

[6] M. Kise, N. Noguchi, K. Ishii, and H. Terao. Development of the agricultural autonomous tractor with an rtk-gps and a fog. *IFAC Proceedings Volumes*, 34(19):99–104, 2001.

[7] F. Auat Cheein, G. Steiner, G. Perez Paina, and R. Carelli. Optimized EIF-SLAM algorithm for precision agriculture mapping based on stems detection. *Computers and Electronics in Agriculture*, 78(2):195–207, 2011.

[8] N. Shalal, T. Low, C. McCarthy, and N. Hancock. Orchard mapping and mobile robot localisation using on-board camera and laser scanner data fusion - Part B: Mapping and localisation. *Computers and Electronics in Agriculture*, 119:267–278, 2015.
[9] F. Bernuy and J. R. D. Solar. Semantic Mapping of Large-Scale Outdoor Scenes for Autonomous Off-Road Driving. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 124–130, 2015.
[10] U. Weiss and P. Biber. Semantic Place Classification and Mapping for Autonomous Agricultural Robots. *IEEE International Conference on Robotics and Automation, Workshop on Semantic Mapping and Autonomous Knowledge Acquisition*, 2010.
[11] A. Ushio. Studies on Development and Maintenance of Photosynthetic Capacity in Leaves of Assimilation Shoots in Rose. *Bull. Natl. Inst. Flor. Sci*, 40(8):15–40, 2008.
[12] S. Thrun, W. Burgard, and D. Fox. *Probabilistic Robotics*. The MIT Press. 2005.
[13] N. Correll. *Introduction to Autonomous Robots*. Magellan Scientific, 2nd edition, 2014.
[14] M. Labbe and F. Michaud. Appearance-based loop closure detection for online large-scale and long-term operation. *IEEE Transactions on Robotics*, 29(3):734–745, 2013.
[15] V. Badrinarayanan, A. Handa, and R. Cipolla. Segnet: A deep convolutional encoder-decoder architecture for image segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 39(12):2481–2495, 2017.
[16] G. J. Brostow, J. Fauqueur, and R. Cipolla. Semantic object classes in video: A high-definition ground truth database. *Pattern Recogn. Lett.*, 30(2):88–97, 2009.
[17] ROS.org. ROS.org — Powering the world's robots. http://www.ros.org.
[18] Y. Pyo, R. Kurazume, and Y. Watanabe. *ROS Robot Programming*. Kurazume Laboratory, 2015.
[19] M. Labbe. "rtabmap_ros - ROS wiki". http://wiki.ros.org/rtabmap_ros.