Development of a Mobile Robot for Harvest Support in Greenhouse Horticulture – Person Following and Mapping –

Hiroaki Masuzawa, Jun Miura, and Shuji Oishi

Abstract— **Agriculture has been suffering from many problems nowadays, and labor shortage is one of the most serious ones. Automation of agricultural works is a possible solution and many efforts have been made for developing harvest support robots. This paper discusses an application of mobile robot technologies to harvest support in the greenhouse horticulture domain. Although there are several harvest support robots for fruits and vegetables, they usually require modification of a greenhouse for autonomous navigation. We therefore aim to develop a mobile robot that can operate and support harvesting works in normal greenhouses, especially for flowers. Based on an investigation of necessary technologies for harvest support robots, we develop a prototype mobile robot with person following and 3D mapping capabilities. We performed preliminary experiments in a real greenhouse and evaluated the developed robot system.**

I. Introduction

Agriculture has been suffering from many problems nowadays, and labor shortage is one of the most serious ones. The portion of the elderly in agriculture is significantly increasing in Japan; as of 2015, the ratio of farmers over 65 years old reached 63.5% (1.32 million) [1]. Improvement of the quality of agricultural crops and the reduction of production cost are also issues to be resolved for making Japanese agriculture be more competitive in the world market. In addition, as the total population in the world is expected to reach 9.2 billion in 2050, increasing the productivity of agriculture is a common problem in the world.

In order to solve these problems, efforts towards the automation of agricultural works are being carried out under the initiative of the Government of Japan. One popular example is the unmanned operation of agricultural machinery in outdoor such as tractors and rice planters [2]. Such a machine can move with a positional error within $5cm$ using a highly accurate GPS sensor and an inclinometer. In the greenhouse horticulture domain, automatic harvesting robots for, for example, strawberries [3] and tomatoes [4] are being developed; computer vision techniques are used for recognizing fruits [5]. In many cases, however, in order to use such automated systems, it is necessary to modify a greenhouse for robots; for example, they sometimes need to install rails for guiding a robot. It is still difficult to use them in the current greenhouse. Therefore, in this research, we aim to develop a mobile robot that can support harvesting work in greenhouses, especially those for flowers.

The authors are with the Department of Computer Science and Engineering, Toyohashi University of Technology, Toyohashi, Japan.

In this paper, we first discuss necessary technologies for mobile robots for harvest support. We then describe the implementation of some of key technologies, that is, person following and map generation in greenhouse domain. We also describe experimental results to confirm the feasibility of the developed system.

The rest of the paper is organized as follows. Sec. II discusses necessary technologies for harvest support mobile robots. Sec. III explains the design and the specifications of a prototype robot. Sec. IV and Sec. V describe the person following and the mapping function of the robot. Sec. VI show experimental results in a real greenhouse. Sec. VII concludes the paper and discusses future work.

II. Necessary Technologies for Harvest Support Mobile Robots

Aichi prefecture where our university exists has highest share of many flowers including roses and chrysanthemums. We visited several farms cultivating flowers in greenhouses in order to investigate necessary functions that mobile robots for flower harvest support need to possess.

One of the key issues in keeping the quality of cut flowers is to shorten the time between reaping flowers and putting them into water. For this purpose, we are aiming at developing a mobile robot which carries a water tank and automatically follows a worker. The robot also moves back to a station once an enough amount of flowers are reaped. Fig. 1 illustrates how the robot works and what its necessary functions is. In this paper, we focus of two of them. One is to follow people for carrying items necessary for harvesting. The other is to make a map and localize the robot in the map for autonomously carrying harvest to a specific station.

Although these two functions are quite popular in the mobile robotics domain, realizing them for flower harvest support requires us to consider a special characteristic of greenhouse. In a greenhouse, we often see heavily-grown plants as shown in Fig. 2. Such a situation can arise as plants grow naturally, but also is artificially made by farmers. Since photosynthesis processes are very important for the growth of plants, branches of plants are expanded and/or bent from crop rows to aisles, and this expansion/bending is different from farm to farm. As a result, plants often occlude aisles in various ways and ordinary crop row detection approaches (e.g, [6], [7], [8]) cannot be applied. This also makes difficult to detect people as well as perform mapping and localization.

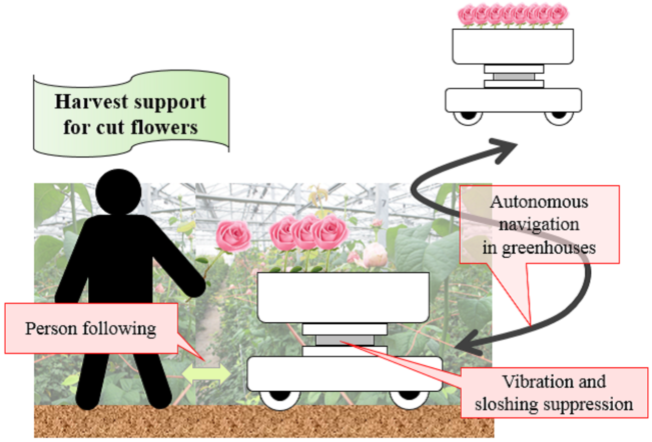


Fig. 1. Tasks and functions of flower harvest support robots.

Concerning people detection and tracking, we need to develop a method which can cope with heavy occlusions by and/or contact with plants. Leg and/or body detection by a 2D LRF (e.g., [9]) or blob detection using 3D point clustering (e.g., [10])do not work well.

Concerning mapping and localization, since traversable regions (i.e., aisles) are sometimes not visible (see Fig. 2), we need to discriminate fake obstacles, which are usually plants that can be pushed away, from real obstacles to avoid. Navigation in such a scene is very challenging. We also consider plants' flexibility in 3D scan matching for ego-motion estimation and localization. A long-term change of plants could also be factors to be considered.

(a) greenhouse 1.

(b) greenhouse 2.

Fig. 2. Greenhouse scene examples.

III. Prototype Robot

A prototype robot is shown in Fig. 3. The size of the robot is $450mm$ in width, $700mm$ in depth, and $1350mm$ in height. The width and the depth were determined based on those of a cart actually used for carrying plants in greenhouse. The height is determined so that the view of the camera on the top is not obstructed by plants. In addition to a computer controlled usage, the robot has a power assist mechanism which makes the manual control easy.

We use an RGB-D camera (Kinect v2 by Microsoft) for measuring distance to objects. We could use a 3D LIDAR instead but it is too costly for a usual operation. A 2D LIDAR is also a possible option but it is not suitable for plant-covered environments. Although the robot is equipped with a 3D LIDAR (VLP-16 by Velodyne) and a 2D LIDAR (UST-20LX by Hokuyo), they are used only for evaluation purposes. The PC used for control and sensor data processing is the one with Core i7-6700HQ CPU and NVIDIA GeForce GTX 960M GPU.

Fig. 3. Prototype robot

IV. Person Following

The person following capability is realized by a combination of an image-based person detection and tracking and a trajectory generation for following movements.

A. Person detection and tracking

We compared three image-based person detection methods. The first one is YOLO, which is an deep neural network-based person detection by Redmon and Farhdi [11]. The

TABLE I

COMPARISON RESULT OF PERSON DETECTION METHOD.

Method	YOLO	Depth image	Viola & Jones
Precision rate	0.938	0.823	0.190
Recall rate	0.875	0.438	0.492

second one is a depth camera-based person detection by Munaro and Menegatti [12]. The third one is to apply a usual Haar-like feature-based method [13] to human upper body images. Using image sequences of person following in a greenhouse as inputs, we compared these methods in terms of recall and precision of person detection. Examples of detection results are shown in Fig. 4 and a performance comparison is in Table I.

Based on the comparison results, we decided to use YOLO for a high performance; especially in recall, it exhibits about 1.7 times higher recall rate than the others. We also analyzed false positive cases by YOLO and found what detected are person images reflected by glass walls of a greenhouse; such false positives can easily deleted using depth information.

Person tracking uses the Kalman filter to estimate the position of a person using the system equation of constant linear motion based on the coordinate position of the detected person. The position and velocity of a person are represented by a robot coordinate system, and coordinates are converted appropriately by using the amount of movement of the robot obtained from odometry.

For a more reliable person tracking, we use Kalman filter for person position estimation. We use a constant velocity model as a state transition model. The position and the velocity of a tracked person are represented in a robot coordinate system, and a wheel odometry is used for a coordinate transformation due to the robot movement.

B. Motion planning for person following

In a usual person tracking scenario (e.g., [14]), a robot can make a collision-free path to the detected person position. Detecting aisles is, however, difficult in a greenhouse, as mentioned above, because aisles are often occluded by grown plants. We therefore use the movement of a person as a guide to a safe path planning; that is, the robot tries to trace the person's trajectory. As a person moves by $0.5m$ or more from the last position, the robot adds it to the sequence of person positions, which are then used as sub-goals of the robot movement. The robot repeatedly moves to the nearest sub-goal to approximately trace the person trajectory.

The translational and the angular velocity at time t, denoted as v_t^r and ω_t^r respectively, are determined by:

$$v_t^r = \begin{cases} v_{max}^r & (d_t^p \geq d_{th1}) \\ \frac{d_t^p - d_{th2}}{d_{th1} - d_{th2}} v_{max}^r & (d_{th1} > d_t^p > d_{th2}) \\ 0 & (d_{th1} \geq d_t^p) \end{cases}, \quad (1)$$

$$\omega_t^r = K_\theta \left(\theta_t^{sg} - \theta_t^r\right) + K_d d_t^{diff}, \quad (2)$$

where d_t^p is the distance between the target person and the robot, θ_t^{sg} is the direction of the line connecting the last and the current sub-goal, θ_t^r is the robot orientation with respect to that line, d_t^{diff} is the distance to that line (see Fig. 5).

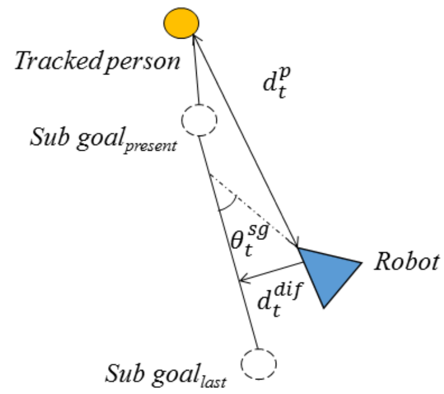


Fig. 5. Calculation of the translational and the angular velocity.

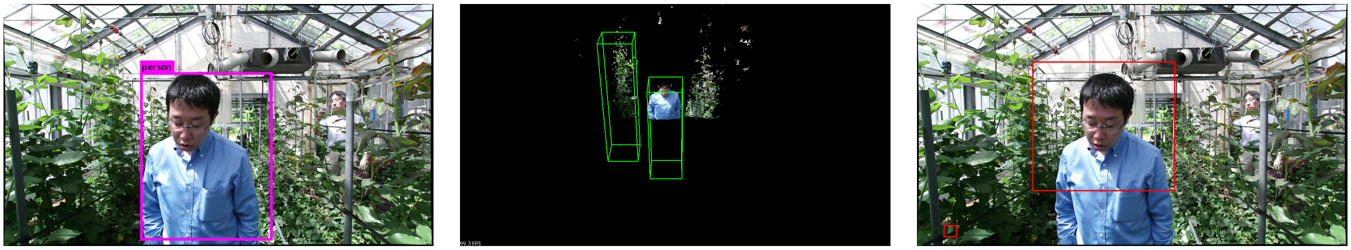
Equation (1) means that the robot moves at the maximum speed if the distance to the target person is larger than d_{th1}, stops if the distance is less than d_{th2}, and moves at an interpolated speed if the distance is in between the thresholds. The parameters are currently set as follows: $v_{max}^r = 0.5m/s$, $d_{th1} = 1.5m$, and $d_{th2} = 0.5m$.

Equation (2) means that the robot is controlled so that the robot is aligned to the line connecting the last and the current sub-goal and comes closer to that line. The gain parameters are currently set as follows: $K_\theta = 0.5$ and $K_d = 0.25$. If the distance to the target person is small enough (less than $1.5m$), however, the robot is controlled to directly move towards the latest target person position.

V. 3D MAPPING OF INTERIOR OF GREENHOUSE

Mapping and localization (or SLAM) [15] are necessary for a robot to autonomously move in the environment. The system uses RTAB-Map [16] for 3D mapping of an interior of a greenhouse. RTAB-Map extracts feature points in the color images, calculates their 3D position using the corresponding depth images, and estimates the camera motion from a set of corresponding 3D points in consecutive frames. Since the wheel odometry is not very accurate in the environments like greenhouses which could cause frequent slippage, we use this method which does not rely on any odometry information. Note that the above feature matching is also used for loop closure detection in RTAB-Map.

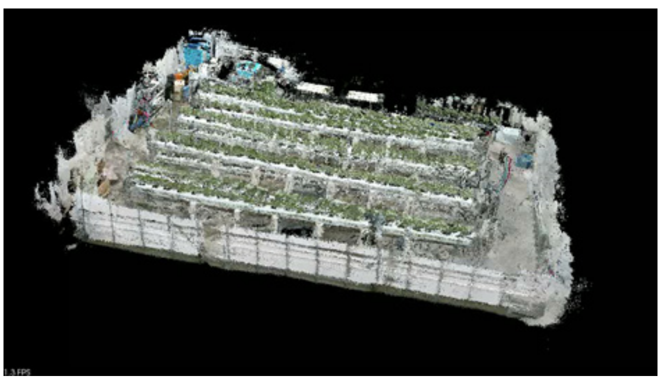
In outdoor environments, the appearance of a scene changes with those of sunlight, and this may degrade image feature extraction and matching, and a final mapping result. We therefore conducted a preliminary mapping experiment in a greenhouse in the TUT campus. Fig. 6(a) shows the experimental scene. The size of the greenhouse is $7m \times 10m$. We manually moved Kinect v2 and collected data over seven minutes of movement. Fig. 6(b) shows the mapping result, which is reasonably consistent over the entire region. This result confirms the use of the combination of Kinect v2 and RTAB-Map.



(a) YOLO [11]. (b) Depth camera-based [12]. (c) Viola& Jones [13].

Fig. 4. Comparison of person detection methods.

(a) Experimental scene.



(b) Mapping result.

Fig. 6. A preliminary mapping experiment in a TUT greenhouse.

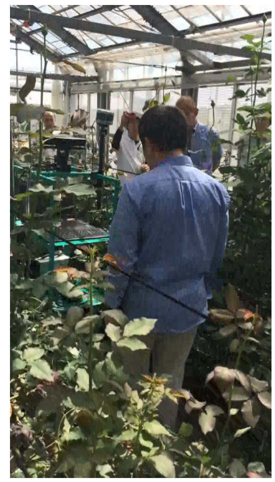


Fig. 7. Experimental scene in a greenhouse.

Fig. 8. Mapping result using RTAB-Map.

Exploration and mapping is one of the approaches to autonomously generating a map of an unknown environment (e.g., [17]). However, this may not work well if detection of traversable regions is difficult in a greenhouse due to heavily-grown plants. One practical approach is to guide a robot throughout the greenhouse aisles using the person following capability. In that case, we need to eliminate depth data of the person so that those data are not included in the final map.

VI. Experiment

We carried out person following and mapping experiments at a rose cultivating site in Aichi Agricultural Research Center. Fig. 7 shows a snapshot of the person following experiment in an aisle of the greenhouse. We got 1,800 frames of data in total and generated a map by using RTAB-Map, as shown in Fig. 8. We also took another approach to making a map using 3D LIDAR data for a comparison purpose. This method uses an NDT-based ego-motion estimation [18] for a 3D LIDAR-based mapping, and Fig. 9 shows the mapping result. The colors in the map indicate the height of the 3D points. The areas to which the maps are generated by both methods are rather different due to different fields of view of the sensors. For the commonly mapped area, they are sufficiently consistent with each other. Therefore, an RGB-D camera is a feasible sensor for mapping a greenhouse with

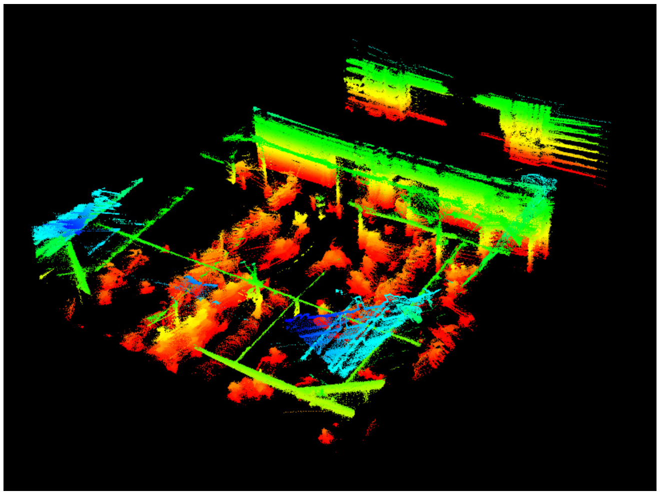


Fig. 9. Mapping result using 3D LIDAR.

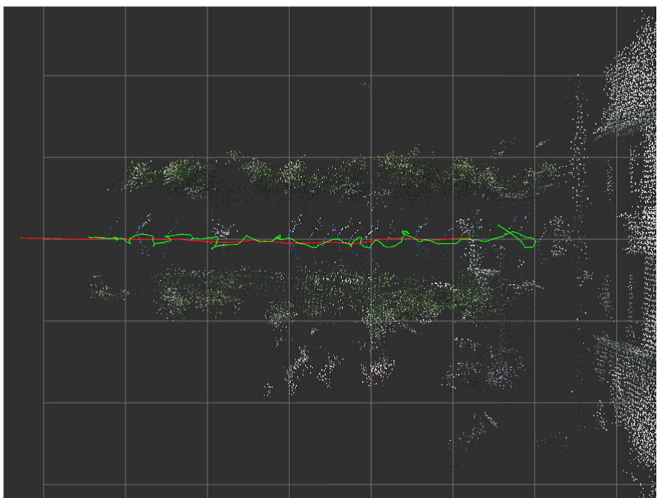
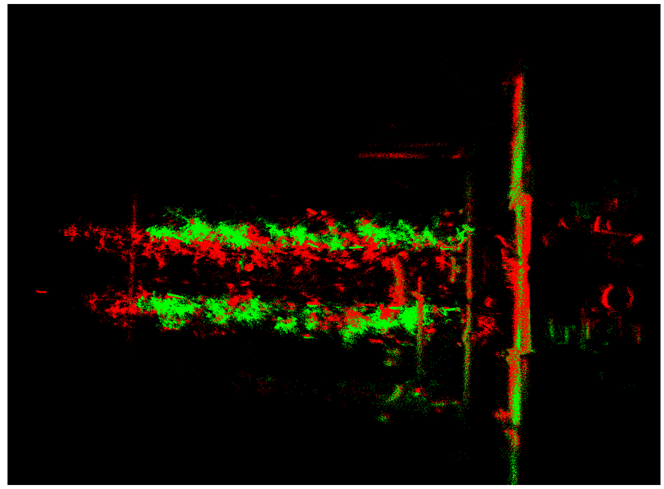


Fig. 10. The person and the robot trajectory during a person following experiment inside an aisle.

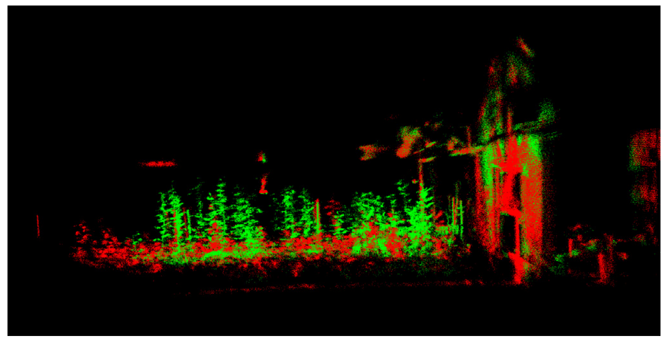
grown plants.

Fig. 10 shows the results of person following in an aisle where the robot took a set of data used for making the map in Figs. 8 and 9. The green and the red line indicate the person and the robot trajectory, respectively, during the person following (and mapping) experiment. Most of time the distance between the robot and the person is small enough, the robot almost always moved toward the current person position (see Sec. IV-B). The robot successfully tracked the person.

We then compared the map with the one which is generated from a set of data taken about two months before. One of challenging characteristics of such a natural scene is that plans grow and their shapes change sometimes drastically. Fig. 11 shows the comparison results. The two independently generated maps are aligned using NDT with a manually-given initial alignment. Thanks to large structural objects such as the walls on the right, two maps are well aligned. We can observe that the greenhouse structure is considered



(a) top view.



(b) side view.

Fig. 11. Comparison of the maps generated using new (in green) and old (in red) data.

sufficient to localize the robot even if the date of mapping and that of localization are largely different. At the same time, we can see the growth of plants clearly due to a nice alignment. This would be beneficial to a precise plant matching over times for watching a growing process for timely harvesting.

VII. Conclusions and Future Work

This paper investigated necessary technologies for mobile robots that support harvesting in greenhouse horticulture. Based on the investigation results, we developed a prototype mobile robot with person following and mapping capabilities. We compared three person detection methods in an actual greenhouse scene, and selected the YOLO detector as most suitable one. We also developed a path planning method that controls the robot so that it follows the trajectory of the detected target person, and that can therefore be applied to aisles with heavily-grown plants. Concerning the mapping, we evaluated RTAB-Map, a SLAM method which uses an RGB-D camera and has been shown to be effective in various environments, also in the same greenhouse scene and confirmed that it can generate a map which is comparable to the one generated by a more reliable 3D LIDAR. We also compared two maps generated with a large interval to show structural objects are effective for localization under a large change of growing plants.

For increasing the efficiency of harvesting work, it is desirable that a robot can autonomously move around in a greenhouse for carrying harvested plants or other items. For this purpose, we are now developing a system which combines the local navigation capability with a global route planning and realize from-to operations. It is also necessary to develop a reliable localization and traversable region detection using RGB-D data for a robust local navigation.

ACKNOWLEDGMENT

The authors would like to thank Mr. Kiyoaki Kakihara (KER Co. Ltd.) for constructing a mobile robot base and Mr. Mitsuo Tsume (Sinfonia Technology Co. Ltd.) for fruitful discussions on flower harvest support. They would also like to thank members of Aichi Agricultural Research Center for supporting them for conducting experiments.

This work is supported by Knowledge Hub Aichi's Priority Research Project (2nd term) "Next Generation Robot Social Development Technology Development Project".

REFERENCES

[1] Yano Research Institute Ltd., "Current status and future prospects of smart agriculture in 2016", 2016 (in Japanese).

[2] N. Noguchi , "Agricultural ICT and Vehicle Robot based on Spatial Information", *J. of the Robotics Society of Japan*, Vol. 35, No. 5, pp. 372–375, 2017 (in Japanese).

[3] Y. Takahashi, Y. Kakigi, H. Kumagai, Y. Abe, H. Seki, T. Fukuda, M. Kashiwazaki, and K. Ozaki, "Prototype of contactless quality inspection device of strawberry using near infrared region and CIE 1931 XYZ color space camera", The 34th Annu. Conf. of the Robotics Society of Japan, 2016 (in Japanese).

[4] T. Matsumoto, G. Nakamura, K. Tomoda, K. Nakao, H. Miki, H. Hamamatsu, and T. Matsuo, "Verification of end effector and image processing system of tomato harvesting robot", 2017 JSME Conference on Robotics and Mechatronics, 2017 (in Japanese).

[5] K. Kapach, E. Barnea, R. Mairon, Y. Edan, and O. Ben-Shahar. Computer vision for fruit harvesting robots – state of the art and challenges ahead. *Int. J. of Computational Vision and Robotics*, Vol. 3, No. 1/2, pp. 4–34, 2012.

[6] S. Jay, G. Rabatel, and N. Gorretta. In-field crop row stereo-reconstruction for plant phenotyping. In *Proceedings of 2nd Int. Conf. on Robotics and Associated High-Technologies and Equipment for Agriculture and Forestry (RHEA-2014)*, pp. 339–348, 2014.

[7] J. Romeo, A. Benítez, J.M. Guerrero, M. Montalvo, and G. Pajares. Tractor guidance in maize fields using computer vision and geometry. In *Proceedings of 2nd Int. Conf. on Robotics and Associated High-Technologies and Equipment for Agriculture and Forestry (RHEA-2014)*, pp. 257–265, 2014.

[8] T.A. Troyer, S. Pilta, and E. Nutter. Inter-row robot navigation using 1d ranging sensors. *IFAC-PapersOnLine*, Vol. 49, No. 16, pp. 463–468, 2016.

[9] Z. Zainudin, S. Kodagoda, and G. Dissanayake. Torso Detection and Tracking using a 2D Laser Range Finder. In *Proceedings of Australasian Conf. on Robotics and Automation 2010*, 2010.

[10] K. Kidono, T. Miyasaka, A. Watanabe, T. Naito, and J. Miura. Pedestrian Recognition Using High-Definition LIDAR. In *Proceedings of 2011 IEEE Intelligent Vehicles Symp.*, pp. 405–410, 2011.

[11] J. Redmon and A. Farhadi. "YOLO9000: Better, Faster, Stronger", arXiv preprint arXiv:1612.08242, 2016.

[12] M. Munaro and E. Menegatti, "Fast RGB-D people tracking for service robots", *Autonomous Robots*, Vol. 37 No. 3, pp. 227–242, 2014.

[13] P. Viola and M. Jones, "Rapid object detection using a boosted cascade of simple features", *Proceedings of 2001 IEEE Conf. on Computer Vision and Pattern Recognition*, Vol. 1, pp. 511–518, 2001.

[14] J. Miura, J. Satake, M. Chiba, Y. Ishikawa, K. Kitajima, and H. Masuzawa, "Development of a Person Following Robot and Its Experimental Evaluation", *Proceedings of the 11th Int. Conf. on Intelligent Autonomous Systems*, pp. 89–98, 2010.

[15] S. Thrun, W. Burgard, and D. Fox. *Probabilistic Robotics*. The MIT Press, 2005.

[16] M. Labbe and F. Michaud, "Online Global Loop Closure Detection for Large-Scale Multi-Session Graph-Based SLAM", *Proceedings of 2014 IEEE/RSJ Int. Conf. on Intelligent Robots and Systems*, pp. 2661–2666, 2014.

[17] Y. Okada and J. Miura, "Exploration and Observation Planning for 3D Indoor Mapping", *Proceedings of 2015 IEEE/SICE Int. Symp. on System Integration*, pp. 599–604, 2015.

[18] M. Magnusson, A. Lilienthal, and T. Duckett, "Scan registration for autonomous mining vehicles using 3D-NDT", *J. of Field Robotics*, Vol. 24, No. 10, pp. 803–827, 2007.