Flexible Human-Robot Interaction in Domestic Environment Using Semantic Map

Yuga Yano
Graduate School of Life Science and Systems Engineering, Kyushu Institute of Technology,
2-4 Hibikino, Wakamatsu, Kitakyushu， Fukuoka 808-0196, Japan.

Yukiya Fukuda
Graduate School of Life Science and Systems Engineering, Kyushu Institute of Technology,
2-4 Hibikino, Wakamatsu, Kitakyushu， Fukuoka 808-0196, Japan.

Tomohiro Ono
Graduate School of Life Science and Systems Engineering, Kyushu Institute of Technology,
2-4 Hibikino, Wakamatsu, Kitakyushu， Fukuoka 808-0196, Japan.

Hakaru Tamukoh
Graduate School of Life Science and Systems Engineering, Kyushu Institute of Technology,
2-4 Hibikino, Wakamatsu, Kitakyushu， Fukuoka 808-0196, Japan.

E-mail: yano.yuuga158@mail.kyutech.jp
https://www.brain.kyutech.ac.jp/~tamukoh/

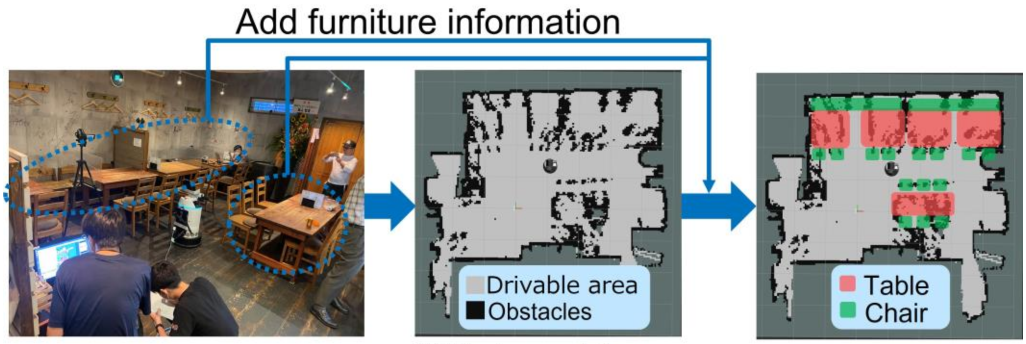
Abstract

We propose an efficient semantic map to realize flexible human-robot interaction (HRI) in domestic environments. Our proposed map is created from an output of Simultaneous Localization and Mapping and already-known environmental information such as furniture and room. In this study, we evaluated the effectiveness of our proposed method on two benchmark tests for HRI in RoboCup@Home held in Bangkok in 2022. In the RoboCup@Home, we employ 3D human recognition to apply our proposed map to HRI, such as "find and offer an empty seat." Our proposed method had the best score of all teams on both tests. The results of our experiments are available at https://youtube.com/playlist?list=PLfbN50Mwh2DG3OPDeCHo4TNuyrU4qYCrJ.

Keywords: Human-robot interaction, semantic mapping, Speech recognition, Human recognition

1. Introduction

In recent years, service robots have been widely used in domestic environments. Typical examples are nursing care robots, cleaning robots, and food delivery robots in restaurants [1], [2], [3], [4], [5] . However, these robots are still limited in their functions and cannot perform flexible actions. Therefore, home service robots that can realize flexible actions with Human-Robot Interaction (HRI) are needed. Estimating the surrounding environment and self-position is necessary to achieve autonomous mobility for home service robots. simultaneous localization and mapping (SLAM) is commonly used for creating an environmental map and self-localization simultaneously. Figure 1(b) shows the example of created environmental map by SLAM. However, SLAM only estimates the surrounding environment in terms of binary values, obstacle or not. Therefore, the robot cannot perform flexible actions such as guiding a person to an empty seat. To address this issue,



(a) Actual environment. (b) Environmental map created by SLAM. (c) Semantic map.

Fig. 1 An example of semantic map.

a semantic map, which assigns attributes to the surrounding environment, is effective. Figure 1(c) shows an example of a semantic map. In this figure, the robot can assign the attribute to an area previously judged to be an obstacle.

In this study, we propose an efficient and effortless semantic map for a more flexible HRI. We evaluate the effectiveness of our proposed semantic map in RoboCup@Home [6]. RoboCup@Home conducts multiple benchmark tests in a room that imitates an actual home environment. Therefore, we can evaluate the performance for social implementation.

2. Proposed Method

In this study, we propose an efficient semantic map. Figure 2 shows an overview of our method. First, our semantic map is created based on an environmental map using SLAM. Next, the semantic map is updated by manually mapping the furniture location to the coordinates in the environment map. Manual mapping is effortless and fast because it requires only two diagonal coordinates for each object. For example, a semantic map for a standard room can be created in less than an hour. In addition, we use the human and action recognition method proposed by Ono et al. [7] Figure 2 shows the semantic map. In this case, the robot estimates that the waving customer is sitting in the left chair.

3. Experiments

We participated in RoboCup@Home in Bangkok in July 2022 to evaluate the effectiveness of our proposed semantic map. We focus on two tasks which can evaluate

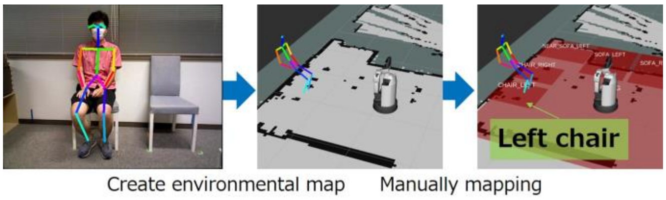


Fig. 2 Proposed semantic map

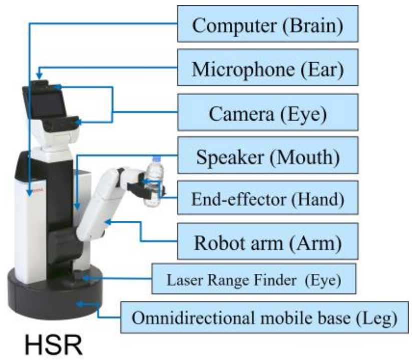
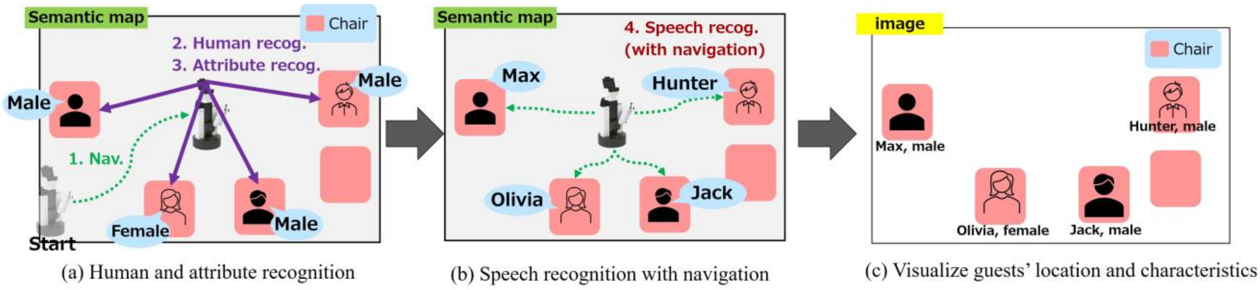


Fig. 3 overview of HSR.

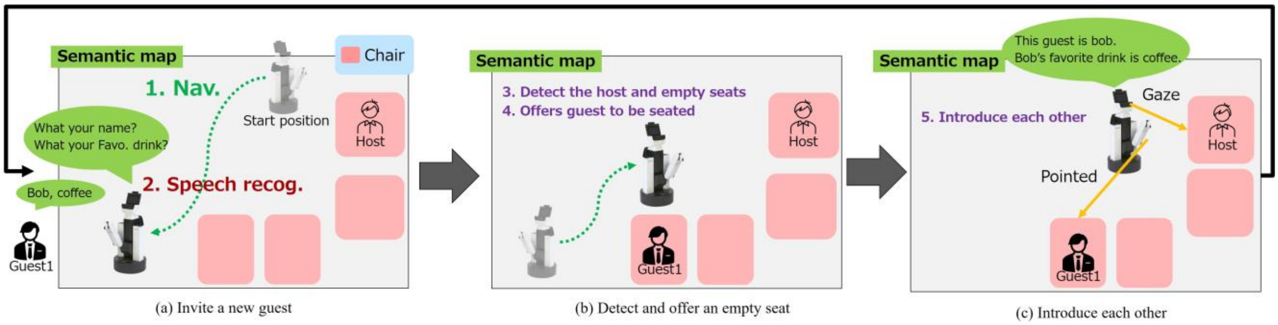
flexible HRI in the domestic environment. In RoboCup@Home, we use Human Support Robot (HSR) [8] as a domestic standard robot. Figure 3 shows the HSR appearance and mainly sensors. In addition, the PC specifications to control HSR are as follows: CPU: Intel core i7-7820HK, GPU: Geforce RTX 1080, Memory: 32 GB, OS: Ubuntu18.04.

In this experiment, we use Vosk [9] as a speech recognition method and class-specific residual attention [10] as a human attribute recognition system. In addition, we use real-time appearance-based mapping [11] as a SLAM.



(a) Human and attribute recognition (b) Speech recognition with navigation (c) Visualize guests' location and characteristics

Fig. 4 Solution for Find My Mates.



(a) Invite a new guest (b) Detect and offer an empty seat (c) Introduce each other

Fig. 5 Solution for Receptionist.

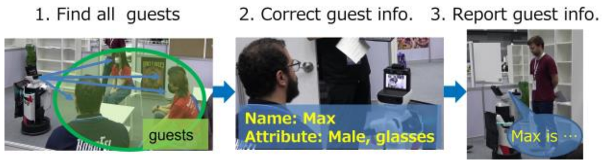


Fig. 6 Task flow of Find My Mates.

Fig. 7 Task flow of Receptionist.

3.1. *Find My Mates*

Find My Mates (FMM) is a benchmark test that collects information about guests visiting the house and reports it to the host. Figure 6 shows the task flow of FMM. Figure 4 shows our solution using our semantic map for FMM. First, HSR navigates to the center of the room and detects where guests are sitting using our proposed semantic map. Next, HSR navigates to the front of each guest based on the location of our semantic map and asks the guest's name using speech recognition. Furthermore, HSR detects the gender of the guest using attribute recognition. Finally, HSR visualizes collected information to inform the host using the head display. Figure 6 (c) shows the visualized image by our solution for FMM. Therefore, this task can evaluate a basic HRI that considers the locations of people and furniture.

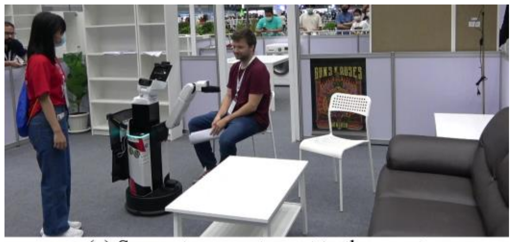
3.2. *Receptionist*

Receptionist (RP) is a benchmark test in which guests visiting a house are escorted to their rooms and introduced to their hosts. Figure 7 shows the task flow of RP. Figure 5 shows our solution using our semantic map for RP. First, HSR corrects a guest's name and favorite drink using speech recognition. Next, HSR detects which chair the host is sitting in using our proposed semantic map. In addition, HSR detects empty seats and offers guests to be seated. Finally, HSR realizes pointing to guests based on the seat locations registered in our semantic map and self-location. RP can evaluate more flexible HRI than FMM because RP needs to realize multiple approaches using furniture and human location.

Fig. 8 Created semantic map in RoboCup@Home.

(a) Actual environment (b) visualized image

Fig. 9 Visualized image to inform guests location and characteristics.



(a) Suggest an empty seat to the guest

(b) Introduce each other

Fig. 10 Behavior of HSR in Receptionist.

Table 1 Result of Find My Mates.

	First try	Second try
1. Navigation	-	✓
2. Human recognition	✓	✓
3. Attribute recognition	✓	✓
4. Speech recognition	-	-
5. Report to the host	Due to Navigation failed	✓

Table 2 Result of Receptionist.

	First guest	Second guest
1. Navigation	✓	✓
2. Speech recognition	-	-
3. Detect empty seats	✓	✓
4. Offer a guest to be seated	✓	✓
5. Introduce a new guest	✓	✓
5-A. Look at a guest	-	-
5-B. Point at a guest	-	-
6. Describe the first guest to the second guest		✓

4. Experimental Results

4.1. *Semantic map*

Figure 8 shows created semantic map in RoboCup@Home. In the experiment, we mapped only the chairs and sofas necessary to perform the task, omitting the desks and the like. We mapped the sofas onto the environmental map in three separate sections because the sofas were designed for three people. The creation of the semantic map, together with the acquisition of the environmental map, took about 30 minutes. It took about 30 minutes to create the semantic map.

4.2. *Find My Mates*

We tried FMM twice at RoboCup@Home. Table 1 shows our result of FMM. In the first trial, HSR could detect human location correctly using our semantic map. In the first trial, HSR failed in the initial navigation, but HSR could detect the person's location using a semantic map correctly. However, due to the low resolution of the guest images, HSR could not perform attribute recognition correctly. As a result, location reports were not accepted because the person images shown on the head display were unclear. In the second trial, we address the navigation issue of the first trial. HSR could get high-resolution guest images and recognize guest attributes correctly. In addition, HSR could navigate in front of the guest but could not get the guest's name by speech recognition. Figure 9(b) shows the image that visualizes guest locations and characteristics in the second trial. In this trial, guests were seated, as shown in Fig. 9(a), and the image correctly reports where guests were seated. Furthermore, the gender and name were also correct, resulting in a perfect score.

4.3. *Receptionist*

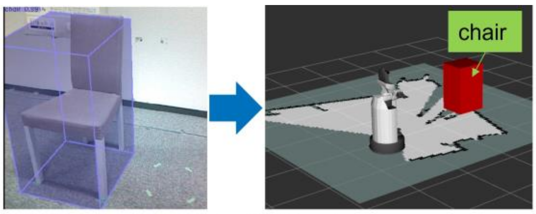
We tried RP only one time. Table 2 shows our results of RP. HSR could navigate in front of the door and ask the guests. However, HSR could not recognize their names and drinks by speech recognition. Furthermore, the chairs were placed in different locations from the predefined location beforehand. As a result, HSR misidentified one of the spectators as the host. Despite the misidentification, HSR could invite all guests and introduce each host and all guests. Figure 10 (a) shows that the HSR suggests an empty seat for the guest. HSR points toward the empty seat and looks at the guest to suggest guest be seated. Figure 10 (b) shows how HSR introduces each guest and host. In this case, the HSR points toward the host and introduces the second guest while looking at the second guest.

5. Discussion

In this study, we proposed a semantic map and conducted experiments in RoboCup@Home. Since HSR could report guest location using our proposed semantic map in FMM, our proposed semantic map can realize a basic HRI that considers the person's location and furniture information. However, HSR could not operate correctly in RP because of the chair position change. Because our proposed map is created manually and we cannot update the map in real-time. In future works, we will enhance the proposed semantic map to operate in dynamic environments using furniture detection. Figure 11 shows our new semantic map with 3D object detection. We use omni3d [12] as a furniture detection method capable of 3D detection and tracking detected furniture.

6. Conclusion

In this study, we proposed efficient semantic mapping and its application to realize flexible HRI using the domestic standard service robot. We also evaluated the effectiveness of our proposed method in RoboCup@Home. In RoboCup@Home, we achieved first place in each task. In future works, we will study semantic mapping using 3D object recognition.



(a) 3D furniture detection (b) Mapping to environmental map

Fig. 11 Semantic mapping using 3D object detection.

References

1. Y. Yano, I. Matsumoto, Y. Fukuda, T. Ono, H. Tamukoh, "Proposal for Solution of Human Interaction Task in RoboCup@Home," JSAI SIG on AI Challenge, SIG-Challenge-060-02, 2022.
2. Y. Tanaka, T. Morie, H. Tamukoh, "An amygdala-inspired classical conditioning model on FPGA for home service robots," IEEE Access, Vol. 8, pp. 212066-212078, 2020.
3. T. Matsushima, Y. Noguchi, J. Arima, T. Aoki, Y. Okita, Y. Ikeda, K. Ishimoto, S. Taniguchi, Y. Yamashita, S. Seto, S. Shane Gu, Y. Iwasawa, Y. Matsuo, "World robot challenge 2020 – partner robot: a data-driven approach for room tidying with mobile manipulator," Advanced Robotics, Vol. 36, No. 17-18, pp. 850-869, 2022.
4. KEENON website, catering robot T-8, https://www.keenonrobot.com/jp/index/Page/index/catid/32.html, (accessed 2022-12-22).
5. PUDO Robotics, Bella Bot, https://www.pudurobotics.com/products/bellabot, (accessed 2022-12-22).
6. " RoboCup@Home," https://athome.robocup.org/, (Accessed 2022-09-03)
7. T. Ono, D. Kanaoka, T. Shiba, S. Tokuno, Y. Yano, A. Mizutani, I. Matsumoto, H. Amano, H. Tamukoh, "Solution of World Robot Challenge 2020 Partner Robot Challenge (Real Space)," Advanced Robotics, Vol. 36 No. 17-18, pp. 870-889, 2022.
8. T. Yamamoto, K. Terada, A. Ochiai, F. Saito, Y. Asahara and K. Murase, "Development of Human Support Robot as the research platform of a domestic mobile manipulator," ROBOMECH Journal, Vol. 6, Art. no. 4, 2019.
9. alpha cephei Vosk Offline speech recognition. https://alphacephei.com/vosk/, (Accessed 2022-0904).
10. Ke Zhu, Jianxin Wu, "Residual Attention: A Simple But Effective Method for Multi-Label Recoginition," arXiv, 10.48550/ARXIV.2108.02456, 2021.
11. Mathieu Labbé and François Michaud, "RTAB-Map as an open-source lidar and visual simultaneous localization and mapping library for large-scale and long-term online operation," Journal of Field Robotics, Vol. 36, No. 2, pp. 416–446, 2019.
12. G. Brazil, J. Straub, N. Ravi, J. Johnson, G. Gkioxari, "Omni3D: A Large Benchmark and Model for 3D Object Detection in the Wild," arXiv, 2022.

Authors Introduction

Mr. Yuga Yano

He received the B.Eng. degree from Kyushu Institute of Technology, Japan, in 2022. He is currently in a master's course student in the graduate school of Life Science and Systems Engineering, Kyushu Institute of Technology. His research interest includes image processing, autonomous driving, and domestic service robots.

Mr. Yukiya Fukuda

He received the B.Eng. degree from National Institute of Technology, Ube College, Japan, in 2020. He received the M.Eng. from Kyushu Institute of Technology, Japan, in 2022. He is currently in a Ph.D. student in the graduate school of Life Science and Systems Engineering, Kyushu Institute of Technology. His research interest includes image processing.

Mr. Tomohiro Ono

He received the B.Eng. degree from National Institute of Technology, Ube College, Japan, in 2018. He received the M.Eng. from Kyushu Institute of Technology, Japan, in 2020. He is currently in a Ph.D. student in the graduate school of Life Science and Systems Engineering, Kyushu Institute of Technology. Since 2020, he has also been a research fellow of the Japan Society for the Promotion of Science (JSPS). His research interest includes image processing, motion planning and domestic service robots. He is a student member of RSJ.

Dr. Hakaru Tamukoh

He received the B.Eng. degree from Miyazaki University, Japan, in 2001. He received the M.Eng and the Ph.D. degree from Kyushu Institute of Technology, Japan, in 2003 and 2006, respectively. He was a postdoctoral research fellow of 21st century center of excellent program at Kyushu Institute of Technology, from April 2006 to September 2007. He was an assistant professor of Tokyo University of Agriculture and Technology, from October 2007 to January 2013. He is currently an associate professor in the graduate school of Life Science and System Engineering, Kyushu Institute of Technology, Japan. His research interest includes hardware／software complex system, digital hardware design, neural networks, soft-computing and home service robots. He is a member of IEICE, SOFT, JNNS, IEEE, JSAI and RSJ.