Socially Adaptive Manner for Motion Planning from Human Server in Cafe

Ryo Hachiuma[1], Yuko Ozasa[1], and Ayanori Yorozu[1]

***Abstract*— We aim to generate a heuristic knowledge from an action of a human server to make the robot provide a socially adaptive service. We label an action trajectory and action types of the server on the 3D environment map. Experimental results show that the motion planning could be conducted by referring to the labeled map.**

I. Introduction

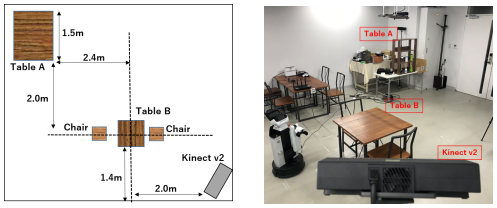
The robots which provide a service to humans work closely with humans in the same environment. As humans behave in a socially adaptive manner which is affected by personal spaces or social rules, we aim to make the robot provide socially adaptive service to humans. For example, the socially adaptive manner for the serving task at the cafe is to consider the following: a path while the server walks to the table, a distance between the server and the table when the server begins to put a cup on the table, and a trajectory while the server puts the cup on the table. In this paper, we extract these manners as a heuristic knowledge from the action of the human server and give the knowledge to the robot as the reference for motion planning.

Unlike imitation learning [1], it is not desirable for the service robots to accomplish tasks with the exact same motion as that of humans because of the physical differences between humans and robots. To achieve socially adaptive navigation, O'Callaghan et al. [2] labeled the walking trajectory on the 2D environment map. For the serving task, the action trajectory of the server while walking and putting the cup on the table should be labeled on the 3D map because taking the shortest trajectory at the navigation and manipulation planning is not appropriate. Moreover, the positions of switching the action types should also be labeled on the map to consider the distance between the server and the table when the server begins to put the cup on the table.

In this paper, we label the action trajectory and action types of the human server on the 3D space of the environment. The labeled space enables the robot to move within the space and switch the action types. We make the robot provide the socially acceptable service to humans by referring to the labeled space. To label the trajectory on the space, the 3D positions of joints are estimated in a single frame. As the positions of joints are distant between frames, we interpolate joints between frames.

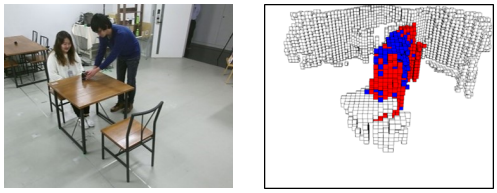
In the experiment, the dataset that four people serve a drink to the table was constructed using an RGBD sensor.

[1]Ryo Hachiuma, Yuko Ozasa and Ayanori Yorozu are with the Graduate School of Science and Technology, Keio University, Yokohama, Japan. {ryo-hachiuma, yuko.ozasa, ayanoriyorozu}@keio.jp



(a) (b)

Fig. 1: The layout and the image of the experimental environment.



(a) (b)

Fig. 2: Visualization of the action trajectory and the action types.

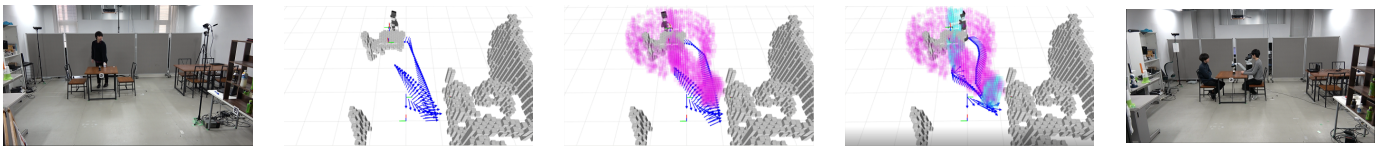
We construct the labeled space for each server and verified that the robot could conduct the motion planning and switch the planning algorithms by referring to the constructed space.

II. Labeling the action of the server

The 3D environment map is constructed from a background point cloud. First, we remove the points of the floor plane from obtained point cloud by RANSAC [3]. Second, isolated points in the point cloud are filtered out by Euclidean clustering [4]. Third, the 3D environment map is constructed by voxelization of preprocessed point cloud.

The 3D positions of the server's joints are estimated to label as the trajectory in the map. The positions of the joints are estimated at each frame [5]. The 3D positions of joints are also voxelized, and corresponding voxels in the map are labeled as the trajectory. These joints are interpolated between frames by Bresenham algorithm.

The labeled voxels are segmented into voxels of the manipulation and navigation to label the action types. The distances between joints in the current frame and joints in the previous frames are used to segment the joints. In this paper, manipulation is an action that moves the arm in a direction vertical to the floor.The joints in the frame are segmented based on the following assumption: if the arm moves vertically to a greater extent than the foot does, the voxels of joints in the frame are labeled as a manipulation, and vice versa.



(a) (b) (c) (d) (e)

Fig. 3: An example of the motion planning results.

III. Experiments

A. Dataset

The layout and an image of the experimental environment are shown in Figure 1 (a) and (b). We constructed a dataset of four people providing the service of transporting a bottle from a table (Table A) placed against a wall to another table (Table B), which was placed in the center of the room as shown in Figure 1. Between zero and two people are seated on a chair, and each server transports a bottle/bottles. Scenes were captured with Kinect v2 which was placed at the corner of the room at a height of $1.8m$ above the floor.

B. Evaluation

We used the robot, HSR of Toyota Motor Corporation for the evaluation. If the robot enters into the voxel of manipulation, the robot performs motion planning using bidirectional constrained rapidly-exploring random trees (RRT) for manipulation and navigation [6], and otherwise, the robot performs the motion planning of navigation using a fuzzy algorithm [7]. To make the robot move along the voxels of navigation a priority, we integrated the membership function that generates the motion.

We compared the motion planning result with two baselines. Baseline A is a method that conducts the motion planning of the robot's arm and an omnidirectional driving system simultaneously by constrained bidirectional RRT [6] and inverse kinematics. Baseline B uses the same motion planning algorithm as Baseline A while passing through the server's trajectory without the segmentation. We compare Baseline A with B to verify the effectiveness of motion planning based on the trajectory of the server. We also compare Baseline B with our approach to verify the effectiveness of switching the motion planning.

IV. Discussion and Conclusion

Figure 2 shows an example image of the labeled space. (a) is the RGB image captured with Kinect, (b) is the segmented trajectory. The segmentation result shows the voxels in red (manipulation) that were located close to the table when the server moved their arms to put the bottle down on the table.

Figure 3 shows an example of the motion planning results by two baselines and the proposed. (a) is the RGB image while the human server put the bottle down on the table. (b) shows the path of Baseline A which takes the shortest path. (c) shows the path of Baseline B which makes the robot pass through the trajectory without the segmentation of action types. (d) represents the path by referring to the proposed space, which makes the robot pass through the trajectory and switch the planning algorithm. (e) shows the result of motion planning with the labeled space in a real environment. The arrows represent the position and posture of the robot at every 0.5 second interval. The root of the arrow is the position, while the direction of the arrow is the posture (yaw angle). The trajectory is colored pink, and the voxels of manipulation are colored light blue.

As Baseline A is to take the shortest path, we can confirm that the robot took the path which is close to the table. In contrast, the path of Baseline B is relatively distant from the table. It shows that motion planning with the trajectory made the robot to maintain a safer distance from obstacles. Comparing between the results of Baseline B and the proposed, our approach makes it possible for the robot to switch motion plannings. Our labeled space enables to generate the knowledge from the action of the server and give it to the robot as the reference for motion planning.

Acknowledgment

This work was supported by JST CREST "Intelligent Information Processing Systems Creating Co-Experience Knowledge and Wisdom with Human-Machine Harmonious Collaboration."

References

[1] Y. Duan, M. Andrychowicz, B. Stadie, O. Jonathan Ho, J. Schneider, I. Sutskever, P. Abbeel, and W. Zaremba, "One-shot imitation learning," in *Advances in Neural Information Processing Systems 30*, I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett, Eds. Curran Associates, Inc., 2017, pp. 1087–1098. [Online]. Available: http://papers.nips.cc/paper/6709-one-shot-imitation-learning.pdf

[2] S. T. O'Callaghan, S. P. N. Singh, A. Alempijevic, and F. T. Ramos, "Learning navigational maps by observing human motion patterns," in *International Conference on Robotics and Automation (ICRA)*, May 2011, pp. 4333–4340.

[3] M. A. Fischler and R. C. Bolles, "Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography," *Commun. ACM*, vol. 24, no. 6, pp. 381–395, Jun. 1981.

[4] R. B. Rusu, "Semantic 3d object maps for everyday manipulation in human living environments," Ph.D. dissertation, Computer Science department, Technische Universitaet Muenchen, Germany, Oct. 2009.

[5] J. Shotton, T. Sharp, A. Kipman, A. Fitzgibbon, M. Finocchio, A. Blake, M. Cook, and R. Moore, "Real-time human pose recognition in parts from single depth images," *Commun. ACM*, vol. 56, no. 1, pp. 116–124, Jan. 2013.

[6] K. Terada, K. Murase, and T. Yamamoto, "Utilization of ros in human support robot (hsr)," *Journal of the Robotics Society of Japan*, vol. 35, no. 4, pp. 280–283, 2017.

[7] A. Yorozu and M. Takahashi, "Obstacle avoidance with translational and efficient rotational motion control considering movable gaps and footprint for autonomous mobile robot," *International Journal of Control, Automation and Systems*, vol. 14, no. 5, pp. 1352–1364, Oct 2016.