Presented at Workshop on Social and Cognitive Interactions for Assistive Robotics held in conjunction with IROS 2022, October 27, 2022.

Observing Human Behaviors for Robotic Assistance in Human-Robot Collaborative Object Search

Takahiro Ishii, Jun Miura, and Kotaro Hayashi
Department of Computer Science and Engineering
Toyohashi University of Technology

***Abstract*— Human-robot collaborative object search is a problem where an appropriate sharing of search areas realizes an efficient search. This paper describes a method of determining the robot's search strategy based on observing human search behavior. The behavior is estimated by matching the estimated travel time for possible behaviors with the actual elapsed time. We confirmed that the proposed method could reduce the overlapped searched areas and the time to find the target objects in the simulation experiments. Future directions toward more dense interactions are also discussed.**

I. Introduction

Mobile service robots have recently been gaining popularity for supporting people in various scenarios such as delivery and attending. In the context of lifestyle support, finding and bringing a use-specified object is a common task for mobile service robots. It has been used as a typical task in many robotic competitions [1], [2]. Another aspect of service robots is human-robot interaction. A human and a robot can collaborate physically or virtually to achieve a task. This paper deals with the problem where a human and a robot are in the same environment and collaborate to find a specific object.

One approach to collaboration is that the human takes the lead by giving commands or suggestions to the robot. There are systems with which humans can assist the robot's recognition or decision [3], [4], [5]. In this approach, a human needs to know how the robot works to give appropriate commands and suggestions. Another approach is robotic support to human task execution [6], [7], where robots refer to the model of the task or the human state and decide the type and the timing of supportive actions. A robot and a human share the task more equivalently in a collaborative object search task.

Collaborative object search can also be considered a distributed search problem. In the case of robot-robot collaboration, robots can easily share respective data through the network (e.g., [8]). In the collaborative search context, for example, one robot knows where the other robots have searched and are going to search and can choose its search area. However, in the case of human-robot collaboration, humans cannot give their knowledge digitally to the robot, and the robot must obtain it in other ways such as observation and dialog, just as in the case of human-human collaboration. This paper focuses on obtaining such information by observation and proposes a new method for estimating and predicting human behaviors. The use of dialog-based interaction is also discussed.

We use the following problem settings. A robot and a human jointly search for a specific object in a common environment with several tables. The map of the environment is given in advance. The target object is on one of the tables, but its location is initially unknown. The robot can detect and locate the human when he/she is visible.

II. Observing and estimating human behavior

An efficient collaborative search can be realized by appropriately sharing the robot and human search areas. To this end, the robot observes human behavior and estimates where the human has searched (estimation of searched area) and will search (prediction of search area). The following subsections explain how to carry them out.

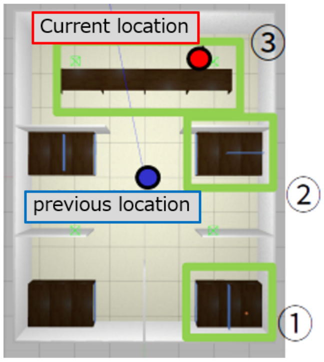
A. Estimation of searched area

In the collaborative search, the robot and the human examine different areas for an efficient search. Therefore, the human is not always visible to the robot, and the robot intermittently sees the human in various places. The valuable information for the robot to determine its action is the area where the human has searched during invisible periods. Therefore, the robot estimates the area from a consecutive pair of human locations and times.

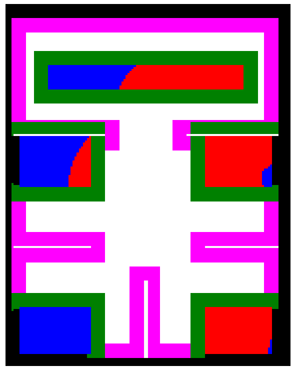
As the object is assumed to be on a table, we set a representative location for each table, and suppose the human goes there for observing the tables. This makes it possible to calculate the time for a sequence of human-visited tables and the previous and the current human location and we choose the best sequence whose total time is the closest to the actual elapsed time (i.e., the time difference between the previous and the current observation).

The time for traversing between two locations is calculated using the Fast Marching Method (FMM) [9] on a grid map. By applying FMM to a grid map with a single starting point (source point), we have a distance map, the pixel value of which indicates the distance from the starting point. As we have a given map of the environment, we precalculate the distance maps from all table locations.

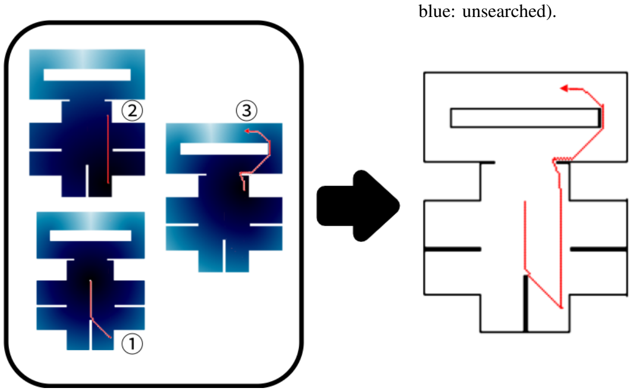
We do a breadth-first search with the previous location as the root. The time of a route from the root to an open node is the summation of the time of movements between nodes, that of movements from the open node to the current location,



(a) Sequence of visited desks.



(b) Searched area estimation (red: searched, blue: unsearched).



(c) Trajectory estimation.

Fig. 1: Example of searched area estimation.

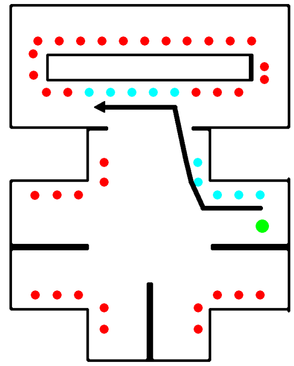
and that for observation at each node (set to 10[s]). We set the allowable time range by the elapsed time with a margin (set to 3[s]) and use the range for pruning branches. The one with the smallest time difference is selected among the sequences within the range. Finally, the estimated trajectory and the searched area are calculated.

Fig. 1 shows an example of trajectory and searched area estimation. Fig. 1(a) shows a scene where the human is considered to visit three tables. Fig. 1(c) illustrates that using three FMM distance maps from the three tables, we can calculate the trajectory and the time of the whole sequence. We then have a map of searched areas shown in Fig. 1(b).

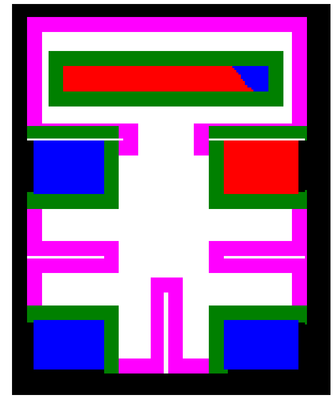
B. Prediction of search area

Predicting where the human will search is crucial for avoiding conflict in search areas between the human and the robot. We thus predict the most probable motion of the human until a fixed future time point for calculating the human's future search area.

In prediction, different from the case of estimating the human's past behavior explained above, we only have one constraint on the human location, which is the location obtained by the latest observation. Considering all possible destinations may cause a combinatorial explosion, we take a best-first search strategy in this paper. We put a set of waypoint candidates and follow them from the latest observed location by repeatedly choosing the nearest unvisited neighboring waypoint. Fig. 2 shows an example case. In Fig. 2(a), the green, the red, and the blue points indicate the human's latest observed location, the waypoint candidates, and the chosen waypoints, respectively. Based on the predicted trajectory, the areas to be observed are calculated (red areas in Fig. 2(b)).



(a) Predicted trajectory (in blue).



(b) Predicted search area (in red).

waypoint candidates chosen waypoints human's current location

Fig. 2: Example of future search area prediction.

III. Robot Motion Planning

The robot takes a next best view (NBV) approach to determine the robot motion. As the location of the target object is initially unknown, we place a set of viewpoint candidates around the tables, as in the case of human motion prediction (see Fig. 2(a)), and choose the best one which maximizes the following score function:

$$score = \frac{S_c}{T_{mot} + T_{obs}},$$

where S_c is the size of the currently-unknown area to be observed by a viewpoint, T_{mot} is the time cost to reach the viewpoint from the current location, and T_{obs} is the time cost for observation there. This score evaluates the reduction of unknown areas per time.

This viewpoint selection method was developed for a single robot object search [10]. We apply this method to the current collaborative search problem by excluding the estimated and the predicted human search areas from the unknown area of the robot's map. As a result, the robot tends to choose the motion toward the areas far from the human.

Fig. 3 shows the robot's observation and motion planning algorithm. We use move_base [11] for path planning and YOLOv3 [12] for object and human detection.

IV. Experiments

A. Simulation environment

We implemented the experimental environment using SIGVerse [13] environment. In SIGVerse, we can enter a virtual environment as an avatar and interact with robots and other avatars. We use a model of Toyota's HSR (Human Support Robot) [14] as the simulated robot and control it from the

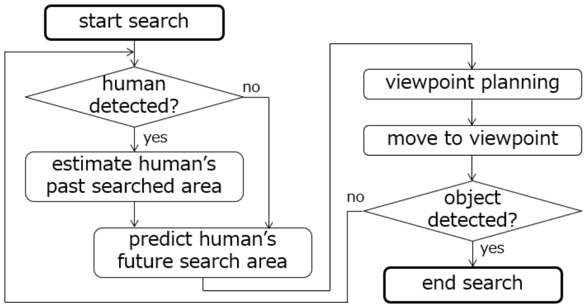


Fig. 3: Diagram for robot motion planning.

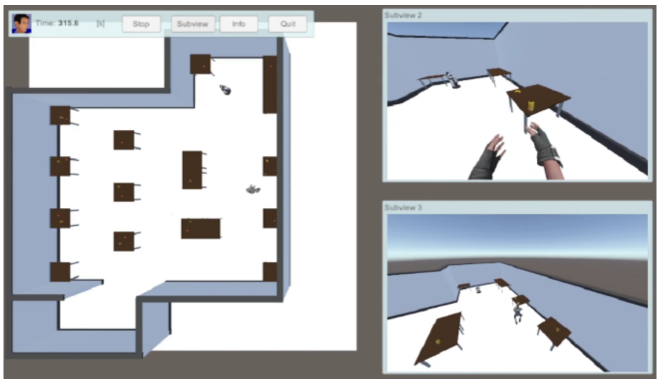


Fig. 4: SIGVerse simulation environment. (left) Evironment setup. (top-right) avatar view. (bottomo-right) human-shooting camera.

ROS environment. Fig. 4 shows the simulated environment. In the experiment, the human subject searches the environment for the specified object freely. As we did not give additional instructions and suggestions, the subject naturally considers the robot's motion in deciding on his/her future motion.

B. Experimental procedure

We compare the following three methods:

(1) The method which does not consider human motions (baseline).

(2) The method which estimates the human past behavior but does not predict human future behavior.

(3) The method which carries out both the human behavior estimation and prediction (proposed).

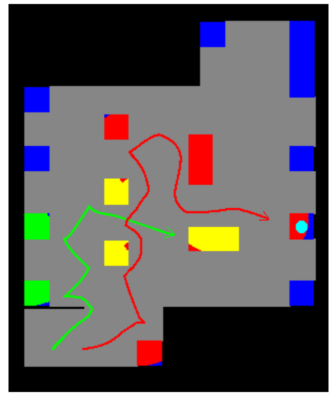
We tested the methods for six locations of the target object. We calculate the time to find the object and the searched areas by the human avatar and the robot. Since it is difficult to know the area precisely a human subject searched, we assume that the human avatar searched the areas within a certain distance from its trajectory.

C. Results

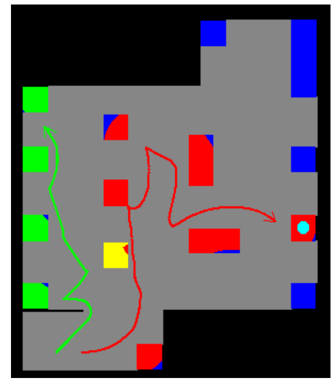
Table I summarizes the experimental results. The proposed method outperforms the others in both the search time and the size of the overlapped search area. Fig. 5 compares the robot and the human trajectories taken by each method for one situation. In the figures, blue, red, and yellow areas indicate the human-searched area, the robot researched area, and the area searched by both the robot and the human, respectively. The green and the red lines indicate the robot and the human trajectories, respectively. From the figures, we can see that by the proposed method, the robot and the human shared the search areas most efficiently by taking trajectories far from each other.

TABLE I: Experimental results.

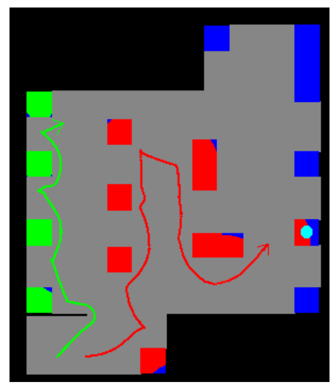
	search time [s]	overlapped search area [%/min.]
(1) No estimation nor prediction	181	6.18
(2) Only estimation	139	3.34
(3) Estimation and prediction	123	1.46



(1) No estimation / prediction.



(2) Only estimation.



(3) Estimation and prediction (proposed).

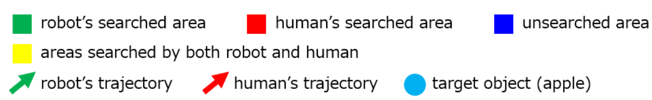


Fig. 5: Comparison of the human and the robot behaviors.

V. Discussion

The experimental results show that the robot and the human can efficiently share the search area by observing each other but without explicit communication. This collaboration was possible because they can frequently see each other in the environment used. However, when the space is more complex (many rooms, for example), or when the map is initially unknown, the uncertainty of estimation and prediction increases because the interval between observations tends to be long and therefore, the decision based on such uncertain information may lead to inefficient robot motion. This problem also applies to the human; it is hard to decide their action without knowing the robot behavior.

Since the robot and the human cannot directly share their information, a promising approach is to use dialog to exchange information among them. It may be sufficient to prepare a set of simple questions for asking the respective search areas. However, in object search in a vast space, as they are sometimes too far from each other to talk, it would also be necessary to choose the right dialog timing. This further raises an interesting "To ask or not to Ask" [15] problem, where uncertainty-driven dialog control [16] becomes one of the necessary functions.

VI. Summary

This paper deals with collaborative object search as a typical problem for human-robot interaction research. We have implemented a method to estimate and predict human past and future behavior. With the method, the robot can decide on its action such that the robot and the human efficiently share the search area, thereby reducing the time to find the target object. We also discussed introducing dialog for further improving the collaboration.

Acknowledgment

This work is in part supported by JSPS KAKENHI Grant Numbers 22H04864.

References

[1] R. Memmesheimer, V. Seib, and D. Paulus, "Winning team of the robocup@home open platform league 2017," in *RoboCup 2017: Robot World Cup XXI*. Springer International Publishing, 2018, pp. 509–520.

[2] "Wrs2020 partner robot challenge (real space)," https://wrs.nedo.go.jp/en/wrs2020/challenge/service/partner.html.

[3] M. Mast, M. Burumester, B. Graf, F. Wiesshardt, G. Arbeiter, M. Španěl, Z. Materna, P. Smrž, and G. Kronreif, "Design of the human-robot interaction for a semi-autonomous service robot to assist elderly people," in *Ambient Assisted Living*, 2015, pp. 15–29.

[4] K. Nagahama, K. Takeshita, H. Higuchi, K. Yamazaki, T. Yamamoto, and M. Inaba, "Estimating door shape and manipulation model for daily assistive robots based on the integration of visual and touch information," in *Proceedings of 2018 IEEE/RSJ Int. Conf. on Intelligent Robots and Systems*, 2018, pp. 7660–7666.

[5] M. Matsushita, K. Hayashi, and J. Miura, "An operator and obserer style teleoperation user interface for remote object search," in *Proceedings of 2021 IEEE/SICE Int. Symp. on System Integration*, 2021.

[6] S. Nikolaidis and J. Shah, "Human-robot teaming using shared mental models," in *Proceedings of HRI2012 Workshop on Human-Agent-Robot Teamwork*, 2012.

[7] H. Goto, J. Miura, and J. Sugiyama, "Human-robot collaborative assembly by on-line human action recognition based on an fsm task model," in *Proceedings of HRI2013 Workshop on Collaborative Manipulation: New Challenges for Robotics and HRI*, 2013.

[8] D. Fox, J. Ko, K. Konolige, B. Limketkai, D. Schulz, and B. Stewart, "Distributed multi-robot exploration and mapping," *Proceedings of IEEE*, vol. 94, no. 7, pp. 1325–1339, 2006.

[9] A. Valero-Gomez, J. Gomez, S. Garrido, and L. Moreno, "Fast marching methods in path planning," *IEEE Robotics and Automation Magazine*, vol. 20, no. 4, pp. 111–120, 2013.

[10] Y. Miake and J. Miura, "Viewpoint planning for object search on the desk using a mobile robot," in *2022 JSME Robotics and Mechatronics Conf.*, 2020, (in Japanese).

[11] "move_base," http://wiki.ros.org/move_base/.

[12] J. Redmon and A. Farhadi, "YOLOv3: an incremental improvement," arXiv:1804.02767, 2018.

[13] T. Inamura and Y. Mizuchi, "Sigverse: A cloud-based vr platform for research on multimodal human-robot interaction," *Frontiers in Robotics and AI: Human-Robot Interaction Section*, 2021.

[14] 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*, 2019.

[15] H. Cai and Y. Mostofi, "To ask or not to ask: A foundation for the optimization of human-robot collaborations," in *Proceedings of 2015 American Control Conf.*, 2015, pp. 440–446.

[16] M. Serras, N. Perez, M. Torres, and A. del Pozo, *Entropy-Driven Dialog for Topic Classification: Detecting and Tackling Uncertainty*. Springer Singapore, 2017, pp. 171–182.