Come See This! Augmented Reality to Enable Human-Robot Cooperative Search

Christopher Reardon[1], Kevin Lee[2], and Jonathan Fink[1]

Abstract— **Robots operating alongside humans in field environments have the potential to greatly increase the situational awareness of their human teammates. A significant challenge, however, is the efficient conveyance of what the robot perceives to the human in order to achieve improved situational awareness. We believe augmented reality (AR), which allows a human to simultaneously perceive the real world and digital information situated virtually in the real world, has the potential to address this issue. Motivated by the emerging prevalence of practical human-wearable AR devices, we present a system that enables a robot to perform cooperative search with a human teammate, where the robot can both share search results and assist the human teammate in navigation to the search target. We demonstrate this ability in a search task in an uninstrumented environment where the robot identifies and localizes targets and provides navigation direction via AR to bring the human to the correct target.**

I. Introduction

One of the primary purposes envisioned for robots operating in field environments is to provide their human teammates increased situational awareness. The benefit of this increased information exchange can be seen in many potential applications. In search and rescue operations, the ability to quickly locate and navigate to victims while simultaneously avoiding dangerous areas would be a potentially life-saving advancement, and is indeed already being realized in teleoperation and limited autonomy scenarios.

A significant challenge that limits this application is the communication of the perception information from robots to human teammates. For example, teleoperation generally allows one human to access and control the information from one robot, at the expense of that person's time and focus. In this work, we wish to improve the efficiency of information exchange for situational awareness in field robotics scenarios. We accomplish this through the use of augmented reality (AR), in which a human's perceived reality is augmented by virtual information that is situated in the real world. We use an AR interface to visualize for the human only information perceived by the robot that is relevant for the task. By visualizing through AR information that is most useful for the human to perform his or her portion of the task, we can increase the human's situational awareness in such a way that the team's performance improves.

[1]U.S. Army Research Laboratory, Adelphi, MD 20783 USA {christopher.m.reardon3,jonathan.r.fink3}.civ@mail.mil

[2]U.S. Army Research Laboratory Research Associateship Program administered by Oak Ridge Associated Universities. klee23456@gmail.com

Recently, there is a growing interest in using AR and mixed-reality to convey robot perception and intent to humans in various human-robot interaction scenarios [1]. Much of the current work, however, is focused on cooperative manipulation of objects by stationary robots with manipulators, e.g., [2] and assembly line-type tasks [3]. Work that has addressed mobile robots, such as [4], has thus far been limited to constrained, instrumented environments.

The novelty of our work is in the application of AR for human-robot team performance in unconstrained field environments. Specifically we examine a human equipped with a wearable, AR head-mounted device (AR-HMD) that is able to convey information about the robot's perception of identified targets in a cooperative search task, and provide navigation planning to assist the human in efficient navigation to the correct search targets.

Fundamental to achieving this capability is the assumption that the robot and human teammates have autonomous localization capabilities, and that the robotic system is able to plan a global navigation path for the human from the current position to the target. In order to achieve this, a mutual alignment of coordinate frames between human and robot teammates must occur. We believe that the ability for robots and AR-wearing human teammates to dynamically achieve mutual alignment of coordinate frames online, without external input and maintain this alignment throughout the mission is a challenging requirement for real-world field robotics applications, which require decentralized position-based reasoning between multiple agents in non-engineered, uninstrumented environments.

Therefore, to generalize our approach to real field environments, we eschew engineering the environment with an external ground-truth calculation device such as motion capture, which has been used in previous works focused on indoor and manufacturing-type environment applications, or mutual observation, e.g., to localize the robot when seen in the AR device's FOV. We instead align our human and robot coordinate frames using the sensory information available from the robot and human-worn AR device.

The AR-HMD and the robot perform online mutual frame alignment of 3D point cloud data from the camera-based visual mapping of the AR-HMD and the robot's LiDAR scans of the environment. We show that by increasing the situational awareness of a human teammate with a robot's perception abilities in this manner, we can improve the human's ability to find and navigate to correct targets. We believe that this is the first application of AR to unconstrained field environments to enable cooperative search.

The remainder of this paper is organized as follows. Section II provides a review of the related research. Our problem is formally defined in Section III. We detail our approach in Section IV. Experimental results are presented in Section V, before finally concluding with Section VI.

II. Related Work

In this section, we provide a review of recent work on the intersection of augmented reality and human interaction with autonomous robots.

Spurred by the recent availability of commercial off-the-shelf human-wearable AR devices, there is a growing body of recent research using AR to provide insight and introspection into the decision space and intents of robotic systems. For example, recent work providing situational awareness to human teammates by visualizing robot video data in AR to see through walls in instrumented environments received attention [5]. The general use of virtual gestures from armless robots for interaction and an analysis of multiple dimensions and features for their application was proposed in [1].

Several recent works focus on human-robot cooperation in object manipulation and assembly scenarios. AR used to visualize Baxter robot motion to enable human perception of planned motion has been shown to enable faster and more accurate task performance in such a scenario [2]. A similar task was performed in [6] using mixed reality for object interaction with armed robot in shared workspace with the addition of analysis of different video perspectives. Robot intent was projected onto objects to visualize task and intention information for assembly line tasks in [3]. How a robot's ability to reveal intentions via AR affects plan cost, termed projection-aware planning, was explored in [7], and illustrated through object manipulation tasks.

Demonstration of visualizing movement and navigation intent in controlled environments has also been examined. Recently, [4] showed that visually signaled robot motion intent using AR improves task efficiency and human understanding of intent for UAVs performing assembly in indoor instrumented environments. On-board intention projection via an LED projector on robotic forklift on the shared floor space was shown to improve human response to the robot [8]. The Kilobot AR project [9], [10] utilizes both a virtualized environment and virtualized sensors for rapid prototyping and scaling to hundreds of Kilobots. For humans to understand robot soccer players' behavior, multiple robot behaviors were visualized on a screen in [11].

In terms of AR interface, light projected onto the environment offers one visualization modality that is particularly suited to well-structured environments, such as assembly lines [3] and factory floors [8]. Fixed [11] and portable [6] screens are another option for AR interface. For field robotics applications in unconstrained environments that are most relevant to safety, security, and rescue scenarios, the human-wearable head mounted display offers promise.

Our work examines the use of an AR-HMD for improving human-robot teaming in these types of scenarios. We believe we are the first to consider AR for human-robot teaming for search in an unconstrained, uninstrumented environment.

III. Problem Statement

We assume a single a ground robot that is able to autonomously perform simultaneous localization and mapping (SLAM) in an unstructured environment to produce consistent localization information for the robot and its sensor measurements that can be used to generate maps of the environment that include both point cloud and occupancy grid representations. These maps of the environment are used to autonomously navigate the environment where perception capabilities can be employed to detect and localize targets of interest, e.g., victims in a disaster scenario. We assume as human teammate equipped with AR-HMD that is capable of reliably tracking its own pose based on local sensor measurements and displaying augmented reality visualization information to the wearer. Furthermore, we assume the AR-HMD system makes its geometric representation of the environment available as a point cloud or triangle mesh. The fundamental problem we are addressing is to use the ground robot to explore an environment, detect targets, and guide the human to each target in turn.

The robot's SLAM system will define a map reference frame, typically with the origin at the starting location for the robot, such that we define its six degree-of-freedom pose as $\mathbf{x}_r \in SE(3)$. Sensor-based observations of targets are represented as $\mathbf{z}_r \in \Re^3$. Note that the subscript r denoting this measurement is represented in the map reference frame of the robot. At the same time, the AR-HMD system will separately define it's own reference frame in which the pose of the human can be represented $\mathbf{h}_a \in SE(3)$ where the a subscript denotes the AR-HMD system reference frame.

In this context, we define two concrete sub-problems that must be solved to address our search task. First, since the robot and AR-HMD system operate independently, we must solve for the rigid-body transformation that allows us to visualize robot-based observations with the AR-HMD. That is, we must compute the homogeneous transformation matrix $T_r^a \in SE(3)$ such that, e.g., we can represent an observation $\mathbf{z}_a = T_r^a \mathbf{z}_r$. Given proper frame alignment, the second sub-problem is to provide visualization through the AR-HMD device that helps direct the human teammate to the next target.

IV. Approach

Our approach to the problem leverages augmented reality to communicate though AR visualization both the position of each target and the path from the human teammate to the target, which is calculated by the robot teammate. By communicating both the relevant target as well as the path to the target, we are able to maximize the efficiency of the human teammate.

However, as noted above, this functionality requires a solution to the transformation T_r^a between the robot's AR-HMD and the system's map frames. We solve for this transformation in two steps. First, the human teammate

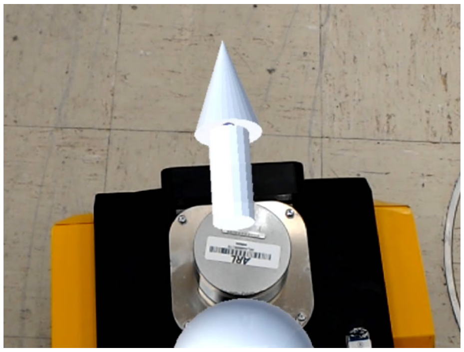

Fig. 1: Illustration of the initial pose estimate (white sphere for location and arrow for orientation) of the robot provided through the AR interface.

uses the AR-HMD system to interactively provide an initial estimate of T_r^a by "placing" an AR marker on the robot and indicating the direction the robot is facing as depicted in Fig. 1. We can use this placement as a "measurement" of the robot x_a in the AR-HMD frame. Then, we can initialize the reference-frame transformation as $T_r^a = x_a \cdot x_r^{-1}$. This provides a coarse estimate that can be further refined by performing the iterative closest point (ICP) algorithm [12] on point clouds produced by the robot and AR-HMD system in their respective reference frames. ICP has been shown to be a robust way to align point clouds even in the presence of noise and, indeed, is the basis for many LiDAR-based SLAM systems. This procedure can be recomputed as new information is observed by both the robot and the AR-HMD in order to improve the alignment as new parts of the environment are explored.

Given an accurate alignment of the robot's reference frame with the AR-HMD system, robot-based observations of targets in the environment can immediately be presented to the human teammate as AR visualizations. However, we hypothesize that in a sufficiently complex environment, e.g., a post-disaster scenario, additional guidance for how the human teammate can navigate to the target will be beneficial. To this end, we leverage the autonomous navigation capabilities that are already present on the robot in order to perform a motion plan that goes from the current human teammate's pose to the next target observed by the robot based on the robot's occupancy-grid representation of the environment. By displaying this path to the human in augmented reality, we can affect a "wayfaring" capability that steers the human along the shortest path to the target of interest.

V. Experimental Results

A. Implementation

The robotic teammate for this experiment is a Clearpath Robotics Jackal as depicted in Fig. 2a. It is equipped with a Velodyne VLP-16 LiDAR, Microstrain 3DM-GX4 inertial measurement unit (IMU) and an Orbbec Astra Pro camera. The robot operates with onboard algorithms to perform simultaneous localization and mapping as well as autonomous navigation as described in [13].

(a) (b)

Fig. 2: Mobile robot (a) AR head-mounted device (b) used in experiments.

Our human teammate was equipped with a Microsoft Hololens head-mounted AR device as depicted in Fig. 2b. Our system is implemented as a suite of C++, Python, and C# software modules. We leverage ROS [14] for messaging, interprocess communication, and common robotics libraries. We use the SIGverse [15] system for Hololens-ROS communication with custom extensions that allow for communication of the Hololens mapping information. The Hololens employs high-quality solutions to the visual SLAM problem, integrating images from five cameras and onboard inertial measurements to rapidly update it's 3D mesh-based map of the environment. This includes place-recognition technology to remember maps of previously visited locations. We found the Hololens to do a good job of accurately mapping complicated environments in bright sunlight, though we did apply a 95% tint to the device to enhance the visibility of the AR visualizations.

At the start of each experiment, the human teammate is asked to use the Hololens interface to "place" a spherical marker on the location of the robot as depicted in Fig. 1. After placing the sphere, the next interaction sets the current heading of the robot. Using this initial alignment, the system performs the ICP algorithm between the vertices of the Hololens mesh with a point cloud rendered by combining all of the point clouds observed by the robot with its Velodyne VLP-16 sensor. We found the ICP algorithm to be stable under a variety of conditions, likely due to the good initial guess provided by the user. This alignment procedure is recomputed as updated point clouds are generated by both the Hololens and the SLAM system on the robot so that the alignment actually improves as more structure of the environment is uncovered.

Aligning the Hololens frame with the robot's map not only allows us to accurately visualize augmented reality markers for the human teammate, but has the additional benefit that the robot knows the current pose of the human teammate. Using this information, when a target is detected, the robot is able to use its occupancy grid to plan a feasible path from the human teammate location to the target using the Search-Based Planning Library [16]. While we constrain the motion-planning search for our robot to account for it's non-holonomic constraints, we plan for the human using a simple grid connected lattice.

This implementation allows the autonomous robot to reason about the human and target location in the same coordinate system, thus enabling the robot to plan feasible paths for the human to localized targets and the visualization of paths and correct targets to the human via the Hololens.

B. Target Localization and Human-Target Path Planning

We demonstrate the increase in the human teammate's situational awareness in a scenario where the robot first searches for and localizes targets, then assists the human to each target.

A robot teammate navigates autonomously [13] to explore the environment, locate candidate targets within a specified area, and identify correct targets for human intervention.

In our experiment, we use AprilTags [17] for our targets. Upon identification of a target, the robot visualizes the location of the target to the human teammate and calculates a path from the human's position to the target using known human pose information in the shared aligned coordinate frame. That path is also visualized to the human via AR, and the human uses this information to navigate to and reach the target. At that point, the robot delivers the next target location and path, if any are available. When all targets are reached, the task is complete.

We demonstrate our system in two realistic environments. The first is an indoor building space with clutter and obstacles. The second is an outdoor environment consisting of multiple concrete buildings and a street arranged as a courtyard.

C. Results

The results of our experiment are depicted in the outdoor courtyard environment in Figs. 3 and 4, and in the indoor environment in Figs. 5 and 6.

For the outdoor environment, Fig. 3 shows the robot exploring the environment and performing autonomous SLAM, and the corresponding indoor environment is in Fig. 5. Figs. 3a-3d show the 2D occupancy map in ROS Rviz constructed by the robot as it searches the environment for targets in the outdoor environment, and Figs. 5a-5e for the indoor environment example. Correspondingly, the respective camera views from the robot's perspective as it performs target search are shown in Figs. 3e-3h for the outdoor environment and Figs. 5f-5j for the indoor environment.

Because the autonomous robot has searched the environment for valid targets, and because we dynamically compute the transformation to align robot and human coordinate frames, the robot is able to provide situational awareness through accurate navigation direction and target identification to the human. When the robot finds a target (Figs. 3d, 3h and 5e, 5j) the human is notified and provided navigation and target location via the AR-HMD.

Using the AR navigation assistance provided autonomously by the robot, the human was able to efficiently execute the path to each target. The human navigation is depicted in Figs. 4 (outdoor environment) and 6 (indoor environment). We show both the human navigation path visualized as a line in ROS Rviz (Figs. 4a-4d and 6a-6d) and from the human's reality-augmented perspective (Figs. 4e-4h and 6e-6h).

Over 30 trials run for these experiments, the human teammate was able to locate the correct object 100% of the time. While this is not a true human-studies trial, we believe that this result, while anecdotal, is representative of the efficiency of our approach. We can see that over the course of the distance traveled in Fig. 6, the close AR marker position relative to the correct physical position shown in Fig. 6h shows a good frame alignment maintained by our approach.

There are several observations that detail the challenges of providing navigation and target localization assistance to human teammates via AR, however. Planning for human navigation is not the same as planning for a robot, as standard settings for robot navigation such as costmap inflation are not appropriate. A human is both more agile than most robots, but also has traversability preferences. As can be seen in Fig. 4e, for example, a feasible path is not always a comfortable path for a human, as the path planned was too close to the walls and obstacles. Being so close to obstacle, this path could additionally introduce frame alignment issues if followed by the human. The AR-HMD device also provides its own challenges, such as limitations induced by its ability to sense and model the environment using the onboard cameras. The Hololens we used has a limited depth and field of view, compared to e.g. a LiDAR, which results in a smaller, directed point cloud. Further, to address warping caused by inconsistent alignment in the global map provided by our AR-HMD device, we performed our robot-to-human frame alignment using only the local Hololens point cloud. Despite these challenges, we see a promising path forward for using AR in human-robot teaming applications, particularly those involving improving human situational awareness in dangerous environments.

VI. Conclusion

We have shown that increased situational awareness can be provided by a AR device from a robot to a human teammate, for the purposes of enabling cooperative search. This was achieved by detecting and localizing targets, as well as providing navigation assistance to the human teammate to efficiently reach each desired target. This was demonstrated in a real-world field robotics context in two different environments, where the coordinate frames of the AR-HMD and robot were aligned automatically without the need for external instrumentation. We identified several challenges in enabling cooperative human-robot search via AR. We believe that this work represents an important first step in using AR to provide situational awareness in human-robot teams in field robotics settings.

References

[1] T. Williams, N. Tran, J. Rands, and N. T. Dantam, "Augmented, mixed, and virtual reality enabling of robot deixis," in *Proceedings of the 10th International Conference on Virtual, Augmented, and Mixed Reality*, 2018.

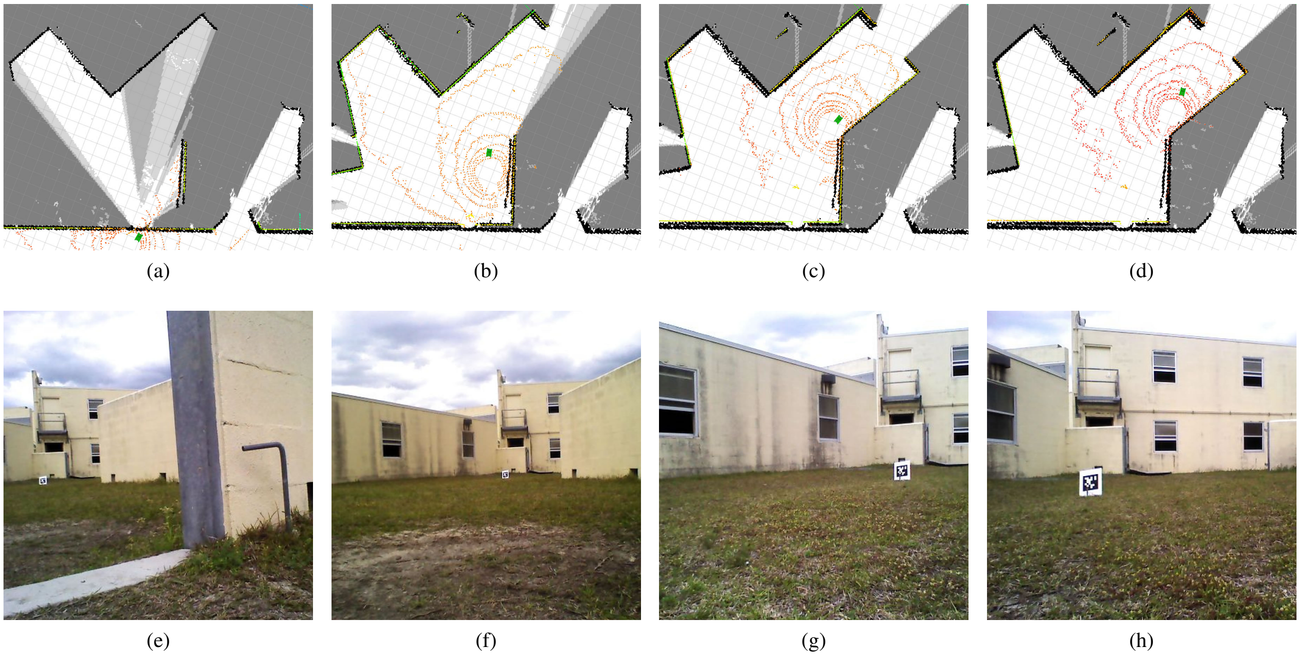

Fig. 3: Example exploration and target search of the outdoor environment by the robot teammate. (a)-(d) show the occupancy grid generated by the robot performing autonomous SLAM as it explores the environment. (e)-(h) show corresponding video captures from the viewpoint of the robot. In (e) and (a), the robot enters a courtyard. In (f)-(g) and (b)-(b), the robot closes distance to a possible target. In (h) and (d), the robot identifies the target (AprilTag).

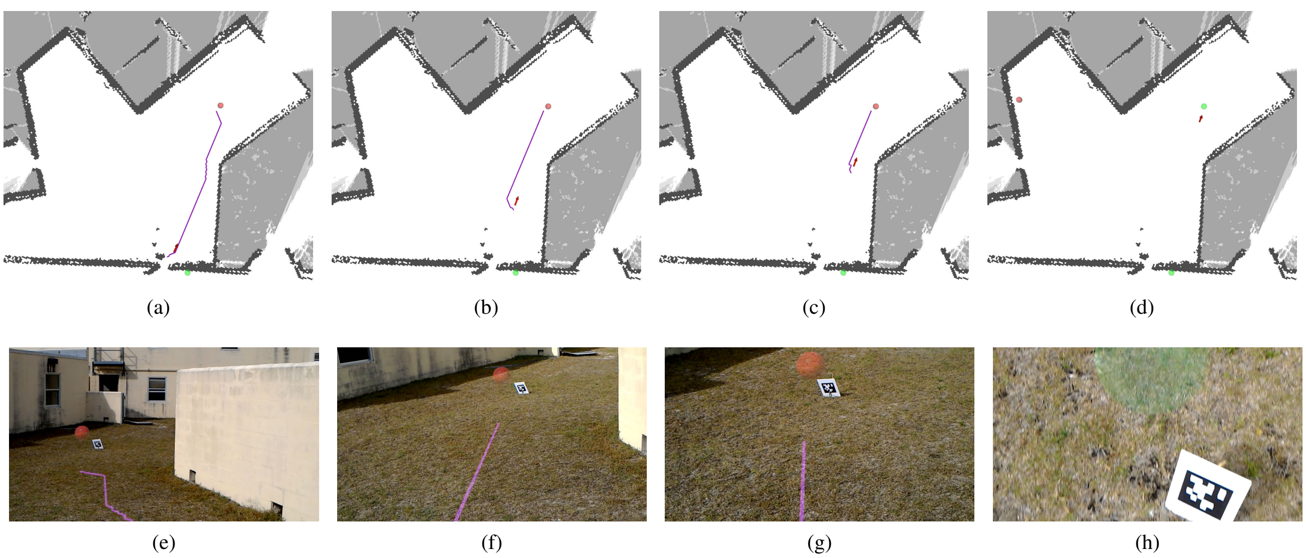

Fig. 4: Example human navigation to targets for the outdoor environment. (a)-(d) show a progression as the human navigates through an outdoor courtyard, using a navigation path and target location generated by the robot, to a target detected by the robot. The map is generated from a 2D laser scan of the environment. The human pose is displayed as a red arrow. Targets are shown as spheres, green for targets the human has reached, red for unreached targets. The autonomously generated path from the human teammate's location to the next target is shown in cyan. (e)-(h) show the AR viewpoint of the human at time instances corresponding to (a)-(d), respectively. Path guidance shown in cyan and the target is shown as a red sphere.

[2] E. Rosen, D. Whitney, E. Phillips, G. Chien, J. Tompkin, G. Konidaris, and S. Tellex, "Communicating robot arm motion intent through mixed reality head-mounted displays," *arXiv preprint arXiv:1708.03655*, 2017.

[3] R. S. Andersen, O. Madsen, T. B. Moeslund, and H. B. Amor, "Projecting robot intentions into human environments," in *Robot and Human Interactive Communication (RO-MAN), 2016 25th IEEE International Symposium on*. IEEE, 2016, pp. 294–301.

[4] M. Walker, H. Hedayati, J. Lee, and D. Szafir, "Communicating robot motion intent with augmented reality," in *Proceedings of the 2018*

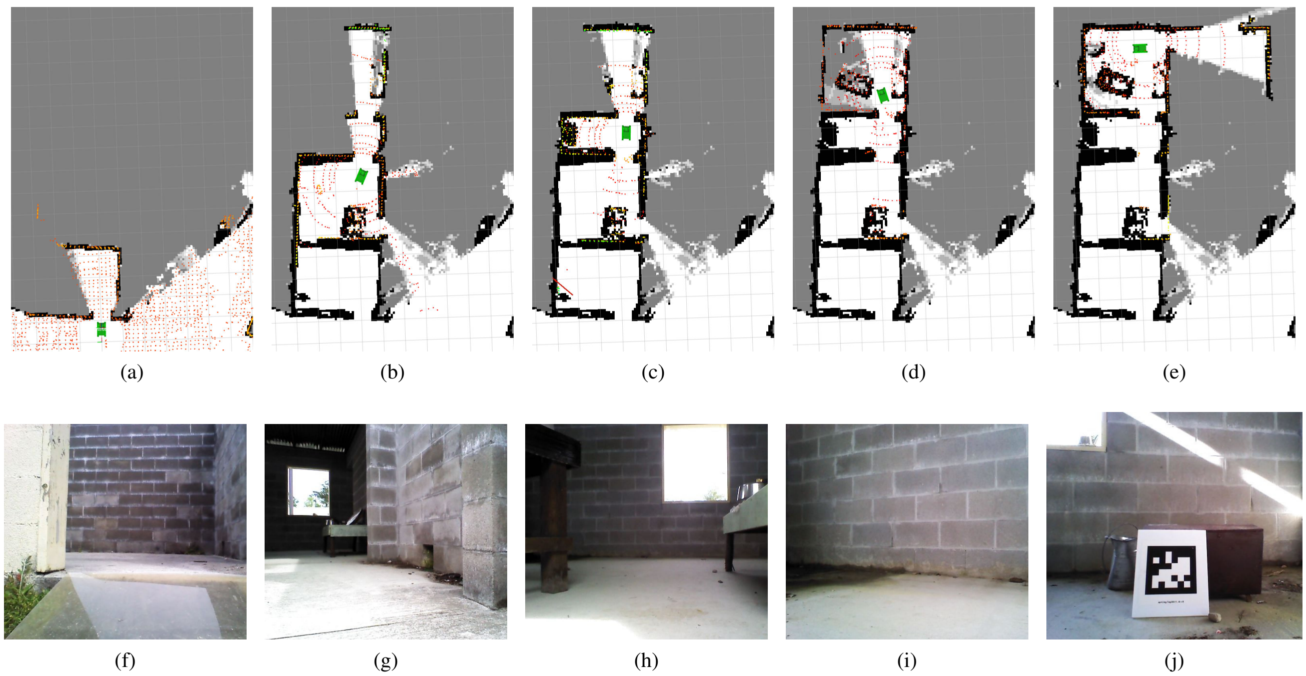

Fig. 5: Example exploration and target search of the indoor environment by the robot teammate. (a)-(e) show the occupancy grid generated by the robot performing autonomous SLAM as it explores the environment. (f)-(j) show corresponding video captures from the viewpoint of the robot. In (f) and (a), the robot enters the building. In (b)-(d) and (g)-(i), the robot explores the environment searching for targets. In (e) and (j), the robot finds a target (AprilTag).

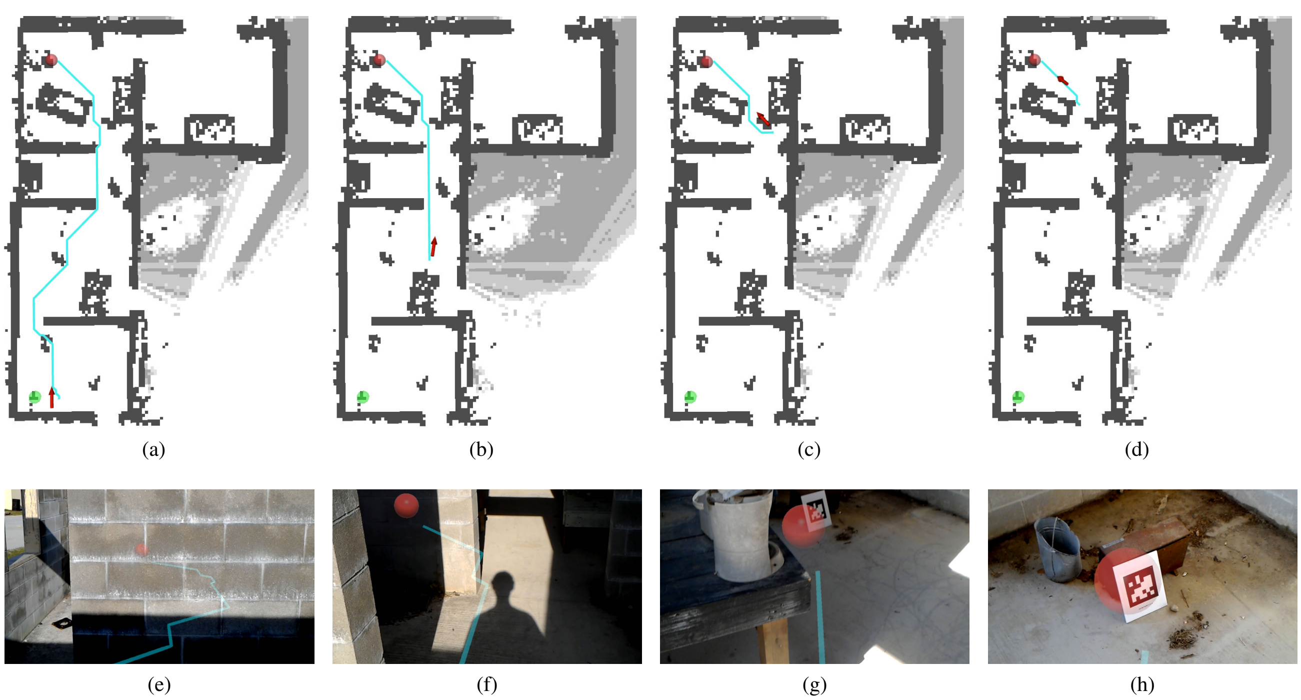

Fig. 6: Example human navigation to targets for the indoor environment. (a)-(d) show a progression as the human navigates through the interior of a building, using a navigation path and target location generated by the robot, to a target detected by the robot. The map is generated from a 2D laser scan of the environment. The human pose is displayed as a red arrow. Targets are shown as spheres, green for targets the human has reached, red for unreached targets. The autonomously generated path from the human teammate's location to the next target is shown in cyan. (e)-(h) show the AR viewpoint of the human at time instances corresponding to (a)-(d), respectively. Path guidance is shown in cyan and the target is shown as a red sphere.

ACM/IEEE International Conference on Human-Robot Interaction. ACM, 2018, pp. 316–324.

[5] O. Erat, W. A. Isop, D. Kalkofen, and D. Schmalstieg, "Drone-augmented human vision: Exocentric control for drones exploring hidden areas," *IEEE transactions on visualization and computer graphics*, vol. 24, no. 4, pp. 1437–1446, 2018.

[6] J. A. Frank, M. Moorhead, and V. Kapila, "Mobile mixed-reality interfaces that enhance human–robot interaction in shared spaces," *Frontiers in Robotics and AI*, vol. 4, p. 20, 2017.

[7] T. Chakraborti, S. Sreedharan, A. Kulkarni, and S. Kambhampati, "Alternative modes of interaction in proximal human-in-the-loop operation of robots," *arXiv preprint arXiv:1703.08930*, 2017.

[8] R. T. Chadalavada, H. Andreasson, R. Krug, and A. J. Lilienthal, "That's on my mind! robot to human intention communication through on-board projection on shared floor space," in *Mobile Robots (ECMR), 2015 European Conference on*. IEEE, 2015, pp. 1–6.

[9] A. Reina, A. J. Cope, E. Nikolaidis, J. A. Marshall, and C. Sabo, "Ark: Augmented reality for kilobots," *IEEE Robotics and Automation Letters*, vol. 2, no. 3, pp. 1755–1761, 2017.

[10] A. Antoun, G. Valentini, E. Hocquard, B. Wiandt, V. Trianni, and M. Dorigo, "Kilogrid: A modular virtualization environment for the kilobot robot," in *Intelligent Robots and Systems (IROS), 2016 IEEE/RSJ International Conference on*. IEEE, 2016, pp. 3809–3814.

[11] D. Zhu and M. Veloso, "Visualizing robot behaviors as automated video annotations: A case study in robot soccer," in *Intelligent Robots and Systems (IROS), 2017 IEEE/RSJ International Conference on*. IEEE, 2017, pp. 6408–6413.

[12] A. Segal, D. Haehnel, and S. Thrun, "Generalized-icp." in *Robotics: science and systems*, vol. 2, no. 4, 2009, p. 435.

[13] J. Gregory, J. Fink, E. Stump, J. Twigg, J. Rogers, D. Baran, N. Fung, and S. Young, "Application of multi-robot systems to disaster-relief scenarios with limited communication," in *Field and Service Robotics*. Springer, 2016, pp. 639–653.

[14] M. Quigley, K. Conley, B. P. Gerkey, J. Faust, T. Foote, J. Leibs, R. Wheeler, and A. Y. Ng, "ROS: an open-source robot operating system," in *Workshop of International Conference on Robotics and Automation*, 2009.

[15] T. Inamura, T. Shibata, H. Sena, T. Hashimoto, N. Kawai, T. Miyashita, Y. Sakurai, M. Shimizu, M. Otake, K. Hosoda *et al.*, "Simulator platform that enables social interaction simulationsigverse: Sociointelligenesis simulator," in *IEEE/SICE International Symposium on System Integration (SII)*. IEEE, 2010, pp. 212–217.

[16] M. Likhachev, "Search-Based Planning Library," https://github.com/sbpl/sbpl. [Online]. Available: https://github.com/sbpl/sbpl

[17] E. Olson, "AprilTag: A robust and flexible visual fiducial system," in *Proceedings of the IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, May 2011, pp. 3400–3407.