Augmented Reality for Human-Robot Teaming in Field Environments

Christopher Reardon[1]([✉]), Kevin Lee[2], John G. Rogers III[1], and Jonathan Fink[1]

[1] U.S. Army Research Laboratory, Adelphi, MD 20783, USA
{christopher.m.reardon3,john.g.rogers59,jonathan.r.fink3}.civ@mail.mil
[2] U.S. Army Research Laboratory Research Associateship Program administered by Oak Ridge Associated Universities, Oak Ridge, USA
klee23456@gmail.com

Abstract. For teams of humans and mobile robots to work together, several challenges must be overcome, including understanding each others' position, merging map information, sharing recognition of salient features of the environment, and establishing contextually-grounded communication. These challenges are further compounded for teams operating in field environments, which are unconstrained, uninstrumented, and unknown. While most modern studies that use augmented reality (AR) in human-robot teaming side-step these challenges by focusing on problems addressable in instrumented environments, we argue that current AR technology combined with novel approaches can enable successful teaming in such challenging, real-world settings. To support this, we present a set of prototypes that combine AR with an intelligent, autonomous robot to enable better human-robot teaming in field environments.

Keywords: Human-robot teaming · Augmented reality · Field robotics

1 Introduction

Intelligent robots working cooperatively alongside human teammates in unconstrained, uninstrumented, and unknown field environments represents a formidable vision with the potential to impact application domains such as disaster relief, search and rescue, environmental monitoring, and military operations. Particularly in disaster and search and rescue work, the need for robots need to perform better alongside humans is well-recognized [8,10].

A specific technical challenge to achieving this vision is the need to cooperate in a natural, non-invasive manner. We believe that this challenge is important for co-located human-robot teams, and can be addressed by using augmented reality (AR). Indeed, a number of recent works have used AR to demonstrate improved human-robot teaming – albeit in structured [5] or instrumented [16] environments.

We believe that it is not only possible but important to transition from this reliance on highly-structured environments to begin using AR for building and studying teams of humans and intelligent mobile robots capable of operating outside the laboratory. Conclusions drawn under the imposition of the constraints of instrumentation could be strengthened and clarified in the transition to creating a human and robot team operating in the real world. To support this argument, we present four approaches that enable essential abilities for humans and robots to work together in these situations through the sharing of metric and symbolic information:

1. Understanding teammate position (i.e., where the teammates are),
2. Merging map information (i.e., where they have been),
3. Recognizing and sharing salient information about objects in their environment (i.e., what they have seen),
4. Communicating understanding and receiving feedback (i.e., deciding what to do next).

To explore these capabilities, we equip a human teammate with an augmented reality head-mounted device (AR-HMD), which we use to collect metric information from the human's task performance, share through visualization metric and symbolic information from the robot's reasoning system, and communicate using augmented reality visualizations and simple dialogue to achieve shared semantic understanding. We present our approaches to enabling each capability, share exemplar experimental results for each, and discuss outcomes.

2   Background and Related Work

With the recent advent of inexpensive, commercial-off-the-shelf AR-HMDs there is a growing interest in using such devices for human-robot teaming tasks [17]. There is growing consensus that AR can be used to provide humans with insight into the perception and reasoning of their robot teammates.

The potential for augmented reality is recognized in several related fields. Versions of augmented reality for supporting maintenance has been a research topic for 50 years. A review of that research is presented in [9]. Several recent papers have examined AR for human-robot teaming in scenarios inspired by maintenance and assembly. For example, using AR to visualize the planned motion of a Baxter robot for faster, more accurate performance of a manipulation task [13]. Mixed reality was used in [5] with armed robot performing object manipulation in a shared workspace. Recent works have used projected light in structured environments such as assembly lines [1] and factory floors [2]. How a robot's ability to reveal intentions via AR affects plan cost, termed projection-aware planning, was explored in [3], and illustrated through object manipulation tasks. AR for medical robotics, for example overlaying stiffness information in surgical applications [18], is also believed to have significant potential.

This paper focuses on using AR with mobile robots to improve human-robot teaming in field environments. Recent work has examined using AR with mobile

robots and shown great potential for enabling human-robot teaming. For example, AR was used to visually signal robot motion intent for UAVs performing an assembly task in [16] and found to improve task efficiency and human understanding of intent. Robot video data was projected to allow humans to "see through walls" in [4] and thereby improve human situational awareness. Methods of communicating robot field of view via AR to improve teleoperation were examined in [7]. [19] used AR via a screen to convey an understanding of a team of robot soccer players' behavior.

However, these previous works were limited to instrumented environments, e.g., using motion capture to localize and perform the coordinate transformations necessary to share any information between the human and robot or robots. Preliminary work by the authors [11], which presented a method using AR to enable human-robot cooperative search, eschewed an instrumented environment and is reviewed as part of this work.

3   AR-Based Approaches to Enabling Human-Robot Teaming

In this section, we present technical approaches to enable each of the four capabilities outlined in Sect. 1. For each approach, a corresponding experimental validation in a motivational application scenario that is enabled by each capability and is relevant to the human-robot teaming in field environments domain is presented in Sect. 4.

3.1   Dynamic Frame Alignment to Understand Teammate Position

In scenarios multiple physical agents working together, an important first step is for each agent to understand where the other agents are. Practically, this requires aligning the coordinate frames of each agent. Environmental instrumentation is often used as a shortcut to bypass this problem, e.g., using motion capture or fiducial markers to trivially locate and directly compute the transforms between agents. While this approach is perfectly valid in constrained environments such as laboratory and factory settings, we believe there is a great potential for impact in unconstrained field environments, for example in cooperative search and rescue in resource-denied locations, such as disaster scenarios.

In these scenarios, teammates perform SLAM independently and the transformation between agents' frames must be computed online. To accomplish this, we take advantage of modern AR-HMD devices' (i.e., Microsoft HoloLens - see Sect. 4.1) ability to localize itself and its wearer through the performance vision-based SLAM, and team a human wearing such a device with a mobile robot performing LiDAR-based SLAM.

In particular, we use the approach presented in previous work [11] to align the human and robot teammates' coordinate frames. We assume that both the robot and the AR-HMD generate a geometric representation of the environment in point cloud format, and compute the homogeneous transformation matrix

between the robot and human point clouds using the Iterative Closest Point (ICP) algorithm [14]. This is initially performed on a coarse estimate provided by the human. It is recomputed online thereafter as the robot and human move throughout the environment.

Knowing this transform allows the robot teammate to understand and reason about the human's position. We previously demonstrated that this capability enables a robot to perform cooperative search with a human teammate [11]. We review the outcomes of enabling this capability in Sect. 4.2.

3.2   Merging Robot and AR-HMD Map Information

In addition to understanding where other agents are, a second critical capability in cooperative teaming tasks is to understand where one's teammates have been. In field applications where teammates must maneuver in the same environment, fusing of map information allows each agent to reason over the other's map information. Our human and robot teammates generate maps using different sensor modalities and at different scales. The challenge is to fuse these heterogeneous maps - one generated by a human wearing an AR-HMD performing vision-based SLAM and another by a mobile ground robot performing SLAM with a LiDAR sensor.

The ground robot uses an OmniMapper-based [15] mapping system, which uses pose graph vertices to represent the robot's location and sensor measurements. Sensor measurements associated with pose graph vertices are then used to generate local occupancy grid maps by iterating through the point cloud and setting an obstacle for points which fall within a height filter. Local occupancy grid maps are generated from the vertices of a pose graph, as points in the point cloud which fall within a height filter are treated as occupied.

The AR-HMD uses an onboard, proprietary SLAM system to generate a model of the environment and localize the wearer within that model. We translate the internal mesh-based representation into a point cloud, from which we are able to similarly generate an occupancy grid for compositing with the robot's map. We use the relative transform method from Sect. 3.1 to effect this transform, and composite the AR-HMD occupancy grid information only into the unmapped area of the robot's occupancy grid.

With this composite map, each teammate is effectively able to make use of the other teammate's exploration efforts in addition to its own when making decisions. We demonstrate this ability in experiments where the robot and the human perform cooperative exploration of an unknown, uninstrumented environment in Sect. 4.3.

3.3   Shared Object Recognition

A third significant capability for enhancing human-robot teaming that can be enabled with AR is the ability to reason about other objects in the team's environment. Building upon the abilities to understand and reason about teammate position and fused map information, we explore how recognizing objects and

localizing them in the shared reference frame can be used to facilitate shared semantic understanding.

An initial implementation of a modern online classifier [12] allows the robot to classify objects from its video feed. Then, as a first step, we achieve semantic understanding for the human through visualizing via the AR-HMD objects recognized by the robot teammate. Visualized objects are highlighted and annotated with information about the robot's knowledge of those objects. For example, at the most basic level this could be object class and a unique index. Sharing this information to the human teammate gives the human insight into the robot's semantic understanding of the environment.

We show in Sect. 4.4 that with this understanding, the human is able to provide clear instruction with regard to the objects in the environment.

3.4   AR-Enhanced Dialogue

A fourth and final powerful capability enabled by AR for human-robot teaming is through the incorporation of basic dialogue for human-robot team cooperative decision making. This capability is particularly complimentary to the shared object recognition capability (Sect. 3.3), as sharing recognized objects with the human allows the human insight into robot's perception and provides a corpus over which dialogue can occur.

We construct this preliminary basic dialogue system by combining commands for basic mobile robot capabilities (e.g., "go to," "explore," "examine") with the set of known object classes to form a dialogue corpus. Then, we use basic speech recognition through the AR-HMD to allow the robot to follow the commands of the human teammate. This allows the human to give instruction in the context of objects, for example, "go to the door." It also makes dialogue particularly powerful in resolving situations of ambiguous semantic grounding, as the robot is able to request feedback in scenarios where instructions are ambiguous.

In Sect. 4.5, we present a scenario where such dialogue is used to clarify ambiguous instructions and correctly perform the command, showing that incorporating dialogue with AR in this way allows a human-robot team to rapidly achieve mutual understanding in decision-making situations.
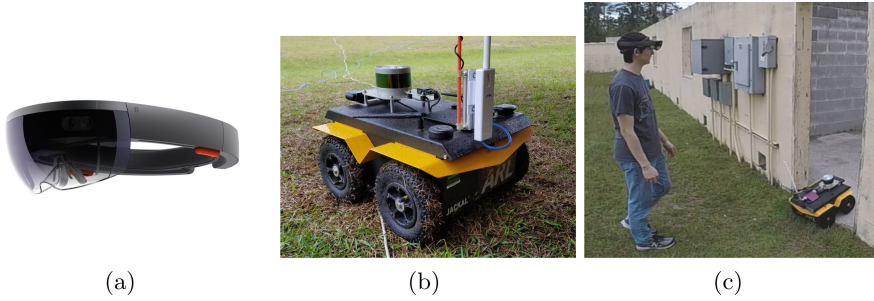
4   Experiments

4.1   AR and Robot Hardware Implementation

The human teammate's AR-HMD used for these experiments is the Microsoft HoloLens[1] shown in Fig. 1a. The HoloLens performs vision-based SLAM onboard using a forward-facing camera array and internal IMU.

For these experiments, the human is paired with a robot teammate. A Clearpath Robotics Jackal (Fig. 1b) is equipped with a Velodyne VLP-16 LiDAR, Microstrain 3DM-GX4 inertial measurement unit (IMU) and an Orbbec Astra

---

[1] www.microsoft.com/en-us/hololens.

(a)   (b)   (c)

Fig. 1. The hardware used for these experiments: (a) Microsoft HoloLens AR-HMD, and (b) Clearpath Robotics Jackal. In (c), the human and robot teammates enter an experiment environment.

Pro camera. The robot is capable of both simultaneous localization and mapping as well as autonomous navigation as described in [6].

4.2 Understanding Teammate Position Enables Cooperative Navigation

Through communication of the mapping information from the AR-HMD and subsequent alignment of that map with the robot's map, the robot is able to understand where the human is relative to the robot. This understanding enables the human-robot team to perform cooperative navigation, where the robot is able to both share search results and provide navigation assistance to the human to each search target.

We demonstrated this ability in a search scenario, where the human and robot are searching an environment for targets, which could be disaster victims, infrastructure needing repairs, etc., which when found by the robot require the attention of the human teammate. Because this is not a visual classification task, in our case we use AprilTags[2] to represent these targets. The targets are hidden in uninstrumented indoor and outdoor field environments (Fig. 2).

The human and robot split up to search. As illustrated in Fig. 3, when the robot identifies and localizes a target, using its understanding of the human position, the robot is able to share the target location via AR. Using its onboard navigation planning, the robot is also able to provide navigation direction via AR to bring the human to the correct target. This cooperative navigation is possible even if the human has not explored the environment where the target is located – the human teammate receives easy to follow navigation direction in the form of a highlighted path through the environment that is updated as the human moves.

We believe these experiments in cooperative search illustrate the general potential impact for robot teammates providing increased situational awareness

---

[2] april.eecs.umich.edu/software/apriltag.

(a)   (b)

Fig. 2. Indoor/outdoor field environment (a) and example target placement in the environment (b).

to human teammates, through enabling the critical capability of robot understanding of human position.

4.3 Merging Map Information Enables Cooperative Exploration

Building upon understanding of human teammate position, the ability to merge heterogeneous human and robot maps allows *both* the human and the robot to reason about where the other has been.

With this capability, knowledge of each teammate's map is available to the other, but clearly each teammate is able to use this information in different ways. The robot is able to plan its exploration strategy over the map explored by the team, not just the robot. Likewise, the human is able to see detailed information about the team's exploration progress and the robot's reasoning, which we present via AR and illustrated in Fig. 4. This includes visualization of the unexplored regions (yellow spheres), unexplored frontiers and projected information gain for exploring them (orange polygons best seen in Fig. 4b), and the robot's plan (green path showing intended navigation route and purple discs showing navigation goals to explore corresponding frontiers in Fig. 4c).

Using this information, the robot is able to explore only areas that the human has not explored, increasing team exploration efficiency. Likewise, the human is able to do the same, using visualizations showing the explored and unexplored regions. In addition, as the human teammate also has access to the robot's planned path, he/she can select areas to explore that are not along the robot's route. Further, by being able to see the robot's *future* plans and expected information gains for each, the human can choose to exploit this information to further deconflict his/her exploration actions while maximizing exploration information gained, e.g., by exploring the frontier with the highest information gain that is not in the robot's immediate plan.

(a)

(b)

(c)
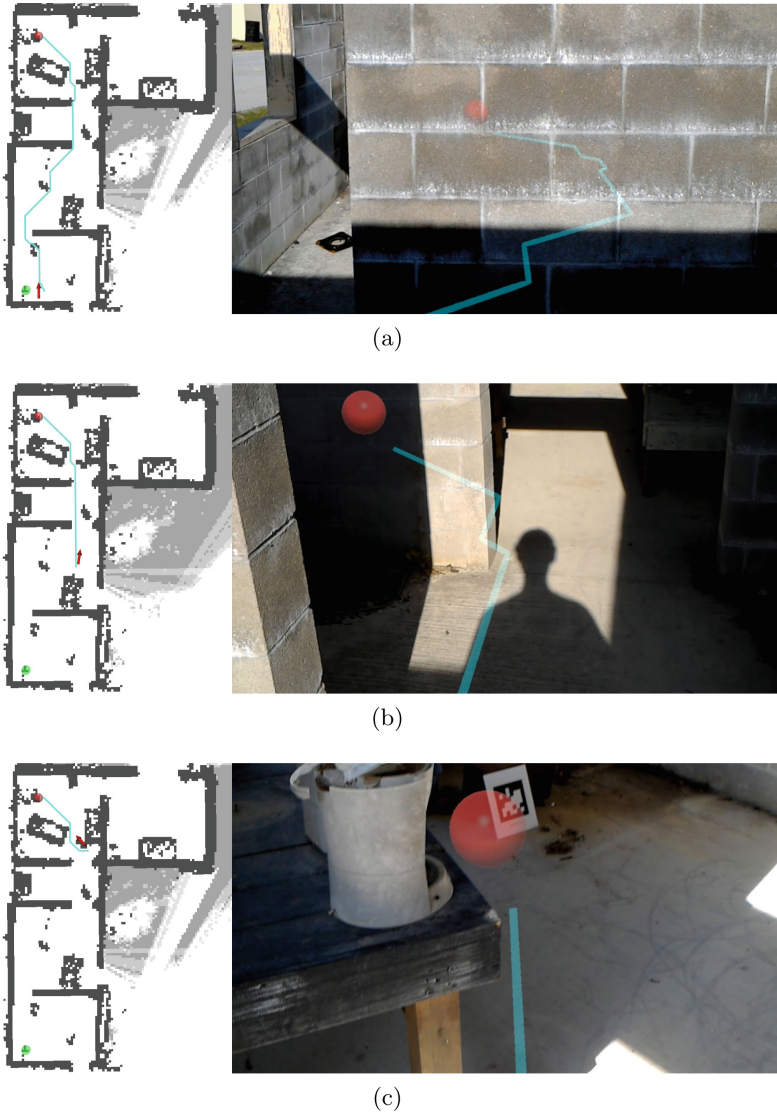
Fig. 3. Cooperative navigation in a search task: (a)–(c) show a sequence of human navigation to a target detected by the robot teammate. In this scenario, the human and robot search independently. When the robot discovers a target, it plans a path from the human's current position to the target. That navigation plan is communicated and visualized to the human (cyan blue line) as well as the target location (red sphere). The navigation plan is updated online as the human moves through the environment, as shown in the robot-generated map on the left, where the red arrow indicates the current location of the human. Through understanding of teammate position in this way, the robot is able to provide navigation assistance to the human teammate, even through parts of the environment that the human has not yet visited. (Color figure online)
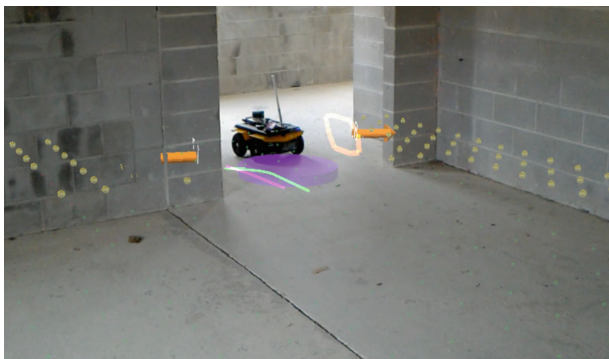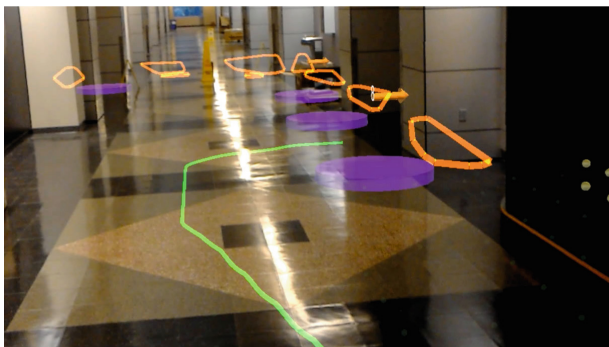
(a)

(b)

(c)

Fig. 4. Heterogeneous map composition allows sharing autonomous robot information-theoretic-based exploration plans, including visualizing unexplored regions (yellow spheres), exploration frontiers and estimated potential information gain (orange polygons with arrows and white numerical text), selected points to begin frontier exploration (purple circles), and robot planned path (green line) via AR for cooperative exploration. Using this information, the human teammate is able to understand the robot's reasoning about the exploration task and explore areas of high reward (information gain) that do not conflict with the robot teammate's plan. (Color figure online)
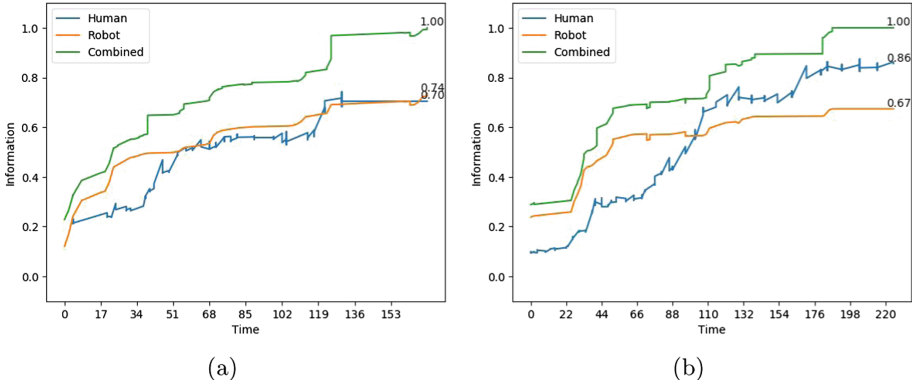
Fig. 5. Information gain from human, robot, and combined exploration. (a) shows information gain in the presence of communication of robot intent only (green lines in Fig. 4). (b) shows information gain from communication of all information (Fig. 4c). Note that overlaps in exploration result in combined information gain that is not the sum of human and robot information gain. (Color figure online)

Our preliminary experiments show that sharing this sort of past and future information can be used to inform and shape human behavior in human-robot collaborative tasks. For example, Fig. 5 shows an example outcome of sharing information about robot intent (Fig. 5a), as in the robot's current planned path, compared with showing full information about the robot's reasoning (Fig. 5b), including frontiers, projected information gain, and potential goal positions as illustrated in Fig. 4c. While this evidence is anecdotal, we believe the observation that sharing the robot's reasoning about complex tasks such as information-based exploration can shape the behavior of human teammates and thereby improve task performance will be supported by planned future studies.

4.4   Shared Object Recognition Enables Semantic Understanding

The ability to provide immediate, transparent insight into the robot teammate's semantic understanding of its environment is a powerful tool to enable team coordination. We validate this capability through proof-of-concept experiments, where the human teammate is able to select, via the AR interface, objects of interest for the robot to visit.

The robot detects and classifies objects of interest in its surroundings. Again building upon understanding of the human location, the robot is then able to provide information about those objects (location, size, orientation) so that the objects can be highlighted (e.g., with a bounding box) and annotated (e.g., with text) via AR. Then, the human teammate is able to use the AR-HMD interface to select an object for the robot to visit. This is done using the cursor in the center of the HMD field of view and the "click" gesture provided by the HoloLens interface. The selected object is set as a goal position for the robot to visit using its onboard autonomous navigation.

(a) (b)
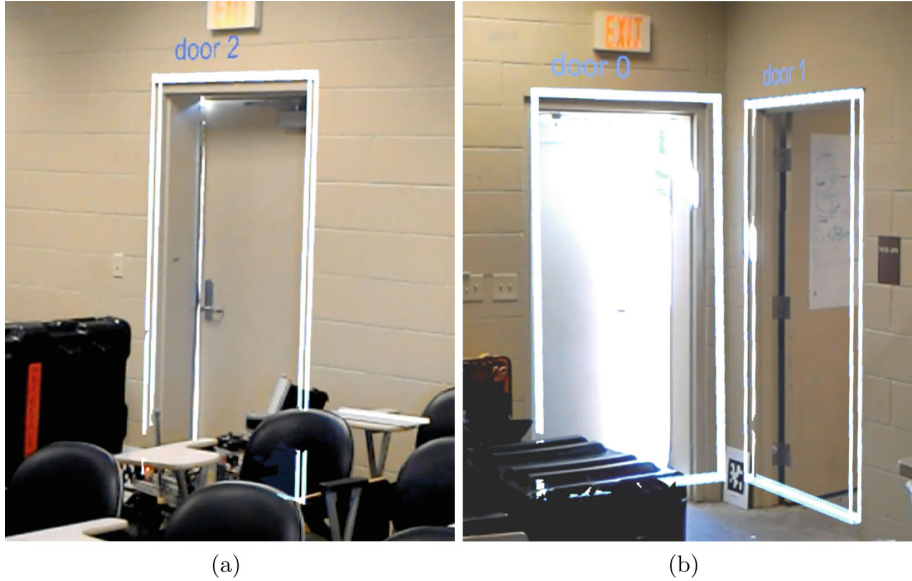
Fig. 6. Shared object recognition allows object (doorway) location and semantic labeling. (a) shows the case of a single door label and index ("door 2"). (b) shows the case of two doors detected in the same vicinity ("door 0" and "door 1"). Note that the bounding box in these images have been manually adjusted to best account for the AR-HMD viewing angle. Automatically finding and displaying the most useful and human-interpretable placement of such 3D visualizations from the human/AR-HMD perspective is an open research question.

For example, illustrated in Fig. 6a, the human is able to give direction to visit the doorway ("door 2"), knowing that the robot has detected and labeled a doorway. The robot can then take appropriate action. While in this case that action is to simply approach the doorway opening, one could imagine more complicated robot behaviors tied to object classification types, commands, and the combinations thereof.

For example, if the robot recognized a potentially dangerous object (e.g., fire, leaking pipe, explosive device), it could be instructed to *take measurements* of that object. Because the human teammate is aware that the robot has classified the object as dangerous, he or she would know that the *take measurements* command would result in the robot keeping a safe distance while recording sensor data, without having to instruct the robot to maintain a safe distance explicitly, thus simplifying the interaction requirements.

We envision this sort of mutual semantic understanding through shared object recognition, which is readily enabled by AR, will improve human-robot team coordination. By knowing what the robot teammate has seen because he or she can see the same information visualized in augmented reality, the human

Table 1. List of simple commands enabled by AR-enhanced dialogue. Table contains spoken command (italics) and non-verbal component (left column), interpretation by robot teammate (center column), and the input modality enabled by the AR-HMD for effecting the command (right column).

Command	Meaning	AR-enabled input	Source
Go to + cursor	Navigate to a position	Human gaze	AR-HMD cursor location
Go to [object]	Approach an object & orient towards that object	Human command	AR-HMD Speech recognition
Follow + human	Actively maintain a minimum safe distance from the human teammate	Human position	AR-HMD location
Explore + direction	Begin information-based exploration in direction	Human orientation	AR-HMD orientation
Return + human	Navigate to the human teammate	Human location	AR-HMD location
Stop	Cease movement		

will be able to communicate to the robot in that context and know if he or she can expect contextually appropriate behavior.

4.5   AR-Enhanced Dialogue Enables Team Decision Making

Finally, we further build upon the previous capabilities to show how AR-enhanced dialogue can be used for team decision making.

Using a basic fixed dialogue consisting of commands and object annotations as discussed in Sect. 3.4, the human teammate is able to give the robot basic commands. The semantic understanding of objects (Sect. 4.4) is particularly valuable in that the human is able to use the classifications of objects in commands, e.g., to "go to door 2" as in Sect. 4.4, except using spoken dialogue instead of pointing and clicking. This dialogue is made more impactful by both the visualization of information from the robot via AR (semantic labeling of objects), and by the use of human teammate information (position and head orientation from the AR-HMD). A table of basic commands is shown in Table 1, showing the command, command meanings, and input modality from the AR-HMD used in the command context.

While this is indeed a useful, if basic, capability, the significance of this capability is even more apparent when dialogue can be used to clarify ambiguity in the robot's understanding of human commands.

For example, Fig. 6b shows a situation where the robot detects multiple doorways. In this situation, if the human directs the robot to go to the doorway, the

robot must make a decision about which doorway. Using the AR-HMD and audio and visual text prompts, the robot can ask the human teammate for clarification – i.e., which doorway. The human can respond with a simple vocal command (e.g., "door 0"), to disambiguate and identify the correct doorway object.

This capability, combined in this way with the object recognition capability from Sect. 4.5, results in a powerful method of rapidly disambiguating communication, through communicating the presence of and resolution to the ambiguity. We believe this capability will be particularly impactful as human-robot teams operate in increasingly complex, cluttered environments.

5   Conclusion

Motivated by the vision of humans and robots operating together as teammates in challenging field environments, we have presented four approaches to using AR to share metric and symbolic information between human and robot teammates: understanding of teammate position, composition of heterogeneous mapping information from teammates, shared understanding of object recognition, and AR-enhanced dialogue. The combination of these AR-enabled abilities beings to address critical requirements for human-robot teaming in field environments. While the results presented in this work are limited, and engineering effort is still required to create robust systems, the technological approaches presented can begin to enable these capabilities, and these limitations will be addressed with future studies. The realization of these capabilities supports the argument that current AR technology, combined with novel approaches, can enable teams of humans autonomous mobile robots to work together in field environments.

References

1. Andersen, R.S., Madsen, O., Moeslund, T.B., Amor, H.B.: Projecting robot intentions into human environments. In: 2016 25th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN), pp. 294–301. IEEE (2016)
2. Chadalavada, R.T., Andreasson, H., Krug, R., Lilienthal, A.J.: That's on my mind! robot to human intention communication through on-board projection on shared floor space. In: 2015 European Conference on Mobile Robots (ECMR), pp. 1–6. IEEE (2015)
3. Chakraborti, T., Sreedharan, S., Kulkarni, A., Kambhampati, S.: Alternative modes of interaction in proximal human-in-the-loop operation of robots. arXiv preprint arXiv:1703.08930 (2017)
4. Erat, O., Isop, W.A., Kalkofen, D., Schmalstieg, D.: Drone-augmented human vision: Exocentric control for drones exploring hidden areas. IEEE Trans. Vis. Comput. Graph. **24**(4), 1437–1446 (2018)
5. Frank, J.A., Moorhead, M., Kapila, V.: Mobile mixed-reality interfaces that enhance human-robot interaction in shared spaces. Front. Robot. AI **4**, 20 (2017)

6. Gregory, J., et al.: Application of multi-robot systems to disaster-relief scenarios with limited communication. In: Wettergreen, D.S., Barfoot, T.D. (eds.) Field and Service Robotics. STAR, vol. 113, pp. 639–653. Springer, Cham (2016). https:// doi.org/10.1007/978-3-319-27702-8_42

7. Hedayati, H., Walker, M., Szafir, D.: Improving collocated robot teleoperation with augmented reality. In: Proceedings of the 2018 ACM/IEEE International Conference on Human-Robot Interaction, pp. 78–86. ACM (2018)

8. Murphy, R.R.: Disaster Robotics. MIT Press, Cambridge (2014)

9. Palmarini, R., Erkoyuncu, J.A., Roy, R., Torabmostaedi, H.: A systematic review of augmented reality applications in maintenance. Robot. Comput.-Integr. Manuf. **49**, 215–228 (2018)

10. Park, S., Oh, Y., Hong, D.: Disaster response and recovery from the perspective of robotics. Int. J. Precis. Eng. Manuf. **18**(10), 1475–1482 (2017)

11. Reardon, C., Lee, K., Fink, J.: Come see this! augmented reality to enable human-robot cooperative search. In: Proceedings of the 2018 IEEE Symposium on Safety, Security, and Rescue Robotics (2018)

12. Ren, S., He, K., Girshick, R., Sun, J.: Faster R-CNN: towards real-time object detection with region proposal networks. In: Advances in Neural Information Processing Systems, pp. 91–99 (2015)

13. Rosen, E., et al.: Communicating robot arm motion intent through mixed reality head-mounted displays. In: Proceedings of the 2017 International Symposium on Robotics Research (ISRR) (2017)

14. Segal, A., Haehnel, D., Thrun, S.: Generalized-ICP. In: Robotics: Science and Systems, vol. 2, p. 435 (2009)

15. Trevor, A.J.B., Rogers, J.G., Christensen, H.I.: OmniMapper: a modular multi-modal mapping framework. In: 2014 IEEE International Conference on Robotics and Automation (ICRA), pp. 1983–1990, May 2014. https://doi.org/10.1109/ ICRA.2014.6907122

16. Walker, M., Hedayati, H., Lee, J., Szafir, D.: Communicating robot motion intent with augmented reality. In: Proceedings of the 2018 ACM/IEEE International Conference on Human-Robot Interaction, pp. 316–324. ACM (2018)

17. Williams, T., Tran, N., Rands, J., Dantam, N.T.: Augmented, mixed, and virtual reality enabling of robot deixis. In: Proceedings of the 10th International Conference on Virtual, Augmented, and Mixed Reality (2018)

18. Zevallos, N., et al.: A real-time augmented reality surgical system for overlaying stiffness information. In: Robotics Science and Systems (RSS) (2018)

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