Augmented Reality Visualization of Autonomous Mobile Robot Change Detection in Uninstrumented Environments

CHRISTOPHER REARDON, University of Denver, USA
JASON M. GREGORY, DEVCOM Army Research Laboratory, USA
KERSTIN S. HARING, University of Denver, USA
BENJAMIN DOSSETT, University of Denver, USA
ORI MILLER, University of Denver, USA
ANIEKAN INYANG, University of Denver, USA

The creation of information transparency solutions to enable humans to understand robot perception is a challenging requirement for autonomous and artificially intelligent robots to impact a multitude of domains. By taking advantage of comprehensive and high-volume data from robot teammates' advanced perception and reasoning capabilities, humans will be able to make better decisions, with significant impacts from safety to functionality. We present a solution to this challenge by coupling augmented reality (AR) with an intelligent mobile robot that is autonomously detecting novel changes in an environment. We show that the human teammate can understand and make decisions based on information shared via AR by the robot. Sharing of robot-perceived information is enabled by the robot's online calculation of the human's relative position, making the system robust to environments without external instrumentation such as GPS. Our robotic system performs change detection by comparing current metric sensor readings against a previous reading to identify differences. We experimentally explore the design of change detection visualizations and the aggregation of information, the impact of instruction on communication understanding, the effects of visualization and alignment error, and the relationship between situated 3D visualization in AR and human movement in the operational environment on shared situational awareness in human-robot teams. We demonstrate this novel capability and assess the effectiveness of human-robot teaming in crowdsourced data-driven studies, as well as an in-person study where participants are equipped with a commercial off-the-shelf AR headset and teamed with a small ground robot which maneuvers through the environment. The mobile robot scans for changes, which are visualized via AR to the participant. The effectiveness of this communication is evaluated through accuracy and subjective assessment metrics to provide insight into interpretation and experience.

CCS Concepts: • **Computer systems organization** → **Robotics**; • **Human-centered computing** → **Mixed / augmented reality**; • **Computing methodologies** → *Mobile agents.*

Additional Key Words and Phrases: human-robot interaction, augmented reality, mobile robots, change detection

Authors' addresses: Christopher Reardon, christopher.reardon@du.edu, University of Denver, 2155 E. Wesley Ave, Denver, CO, USA, 80208; Jason M. Gregory, jason.m.gregory1.civ@army.mil, DEVCOM Army Research Laboratory, 2800 Powder Mill Rd, Adelphi, MD, USA; Kerstin S. Haring, kerstin.haring@du.edu, University of Denver, 2155 E. Wesley Ave, Denver, CO, USA, 80208; Benjamin Dossett, benjamin.dossett@du.edu, University of Denver, 2155 E. Wesley Ave, Denver, CO, USA, 80208; Ori Miller, ori.miller@du.edu, University of Denver, 2155 E. Wesley Ave, Denver, CO, USA, 80208; Aniekan Inyang, aniekan.inyang@du.edu, University of Denver, 2155 E. Wesley Ave, Denver, CO, USA, 80208.

Manuscript submitted to ACM

ACM Reference Format:
Christopher Reardon, Jason M. Gregory, Kerstin S. Haring, Benjamin Dossett, Ori Miller, and Aniekan Inyang. 2023. Augmented Reality Visualization of Autonomous Mobile Robot Change Detection in Uninstrumented Environments. 1, 1 (February 2023), 31 pages. https://doi.org/XXXXXXX.XXXXXXX

1 INTRODUCTION

Novel applications in robotics require that robots work as capable team members alongside humans [28]. This requirement imposes rigorous and challenging requirements to the robot technology. For one, the technology itself needs to be significantly advanced to contribute to the team. This means that the robot-to-human information exchange either detects and provides relevant information through the robot technology in addition to the existing human knowledge in the team [22], increases the human situational awareness [44], or prevents the human from failing to recognize certain important events in their environment as it can happen for example with change blindness. Change blindness refers to a perceptual phenomenon when human observers fail to detect substantial changes to the visual details of objects and scenes [61]. A second requirement is that the robot needs to efficiently and effectively communicate this information to the human agents on the team. Efficient in this context refers to the timeliness and the speed a piece of information can be relayed to a human, and the effectiveness refers to how well the information is communicated, if the information is relevant, and how well and rapidly the human can comprehend the information. If a substantial change occurred in the environment, a robot-to-human information exchange must take place communicating the change in the environment and facilitate human decision making which requires representing robot data in ways that support fast and accurate analyses by humans [67]. For example, if an explosive device was added in a volatile environment the human should be notified of this addition even if it is a perceptual small change due to the potentially high consequences. If an item was removed from a supermarket environment, a notification of this change could lead the human to decide to re-order the item. If 50 items were removed and 20 items were added in a robot-patrolled warehouse, it likely would be helpful to the human to have an appropriate visualization of where the items were removed from and added to.

While both the robot technology to detect information and the communication between robot and human are major challenges in itself to be overcome if humans and robots are to form teams [17, 27], an additional third challenge is imposed by the high volumes of metric-based information that modern robots are becoming increasingly capable of producing from advanced sensing and perception systems. This poses a challenge to robot systems in mixed human-robot teams beyond robot capabilities and interaction, as the robot system also needs to intelligently reason about the relevance of information in order to down-select and prioritize what information to communicate and how to visualize dynamic and uncertain information to a human. These three challenges, robot capabilities, interaction, and information communication, highlight the current gaps in robotics and human-robot teaming research.

Notably in safety-critical domains, but also in other domains that require the inspection of the environment, robots can use their sensing and perception abilities to identify *changes* in the environment relative to their prior experience. This specific change detection application encapsulates all three of the above challenges. Robot facilitated change detection also highlights the potentially high quantity of information encoded in environment changes when compared to a previous point in time, and could be particularly useful for environments that are not highly structured. For example, the presence of a new object in an environment could indicate a survivor's presence or activity in a search and rescue mission. A change (addition) in the environment could also mean adversarial activity like an improvised explosive device (IED) that a military or humanitarian mission might encounter, or a bomb in a public place targeted by terrorists. In those critical scenarios, accurate detection and robot-to-human communication of changes could mean life or death. There are many other high impact applications in industry that also fit this paradigm while being not life-critical but likely occur more often. For example, inventory management applications such as identification of out-of-stock, in-stock, misplaced, and missing or damaged inventory in large-scale supermarkets or warehouses.

Even with the ability to intelligently detect changes, a robot might not understand which of those changes hold information that is important to humans. For example, a "wet floor" sign in a supermarket is a change to the environment, and it is important information to some humans (e.g., customers that pass by) but not to others (e.g., employees that need to determine where shelves were emptied and how much to re-order). The robots that detect changes might not understand the relevance or the implication of the saliency of these change detections, whereas a human that understands the mission or the context would interpret those changes appropriately. For example, IEDs are often hidden and their new addition to the environment represents a very small visual change, yet it is a highly critical one. The removal of a family-size cornflakes package is a larger visual change, yet not critical to the operation of the supermarket. A robot patrolling an environment can be better than humans at detecting changes as a human might not have memory of or notice a certain change especially when under high cognitive workload and in complex environments [62], whereas a robot would identify all changes no matter the scale. However, the importance of the change(s) and resulting actions are likely better to be determined by the human, and this process is guided by the robot-to-human information exchange. Both, the detection of a change as well as correctly identifying the meaning of a change, is essential to determine the appropriate action to follow and will determine the value of a robot in a team.

An integral part of all teams is their communication [13]. Therefore, possessing additional information is not the only necessity for a robot to make a good team member [79]; effectively communicating this information to the human team members is also critical. We believe that this capability is best solved via artificial intelligence (AI); that is, imbuing the robot with the ability to reason and select the best communication strategies. It has been shown that the quality of the communication is significantly more related to the the team performance than the frequency of communication [35]. An example of qualitative communication is when team members contribute unique information during a task (i.e., information elaboration [29]) and when team members contribute special or hard-to-find knowledge, expertise, or specialized skills to other team members (i.e., knowledge sharing [16]). The information flow in communication can be unidirectional (i.e. only from the robot to the human or vice versa) or bidirectional [32]. While there is general agreement that human-robot communication is a crucial part of effective human-robot teams, the form that communication will take is still undefined. In particular, recent research has shown that when asked to identify different targets in a collaboration with a robot, humans missed more targets when bidirectionally communicating compared to unidirectional communication where the robot pushed information to the human [34]. For the purpose of this research, we are avoiding confounding factors of bidirectional human-robot communication and focus on a unidirectional information flow pushed from the robot to the human. While there is a body of research that suggests that human-robots teams with natural language perform better [13], other research suggests that the decision of using verbal or non-verbal communication is highly task dependent [77]. In our work, the robot is identifying changes in a three-dimensional, natural environment. Utilizing natural language to describe where a change happened would require the robot to resolve spatial natural language references and parse its natural language utterances into action sequences that direct the human towards the changes in the environment [78]. Our approach avoids introducing additional algorithms for spatial reference resolution as well as the additional cognitive load that switching modalities would require, i.e., communicating a visually perceived change in the environment through natural language. Instead, we communicate the visually perceived change through a visual highlight projected directly into the human's perception of

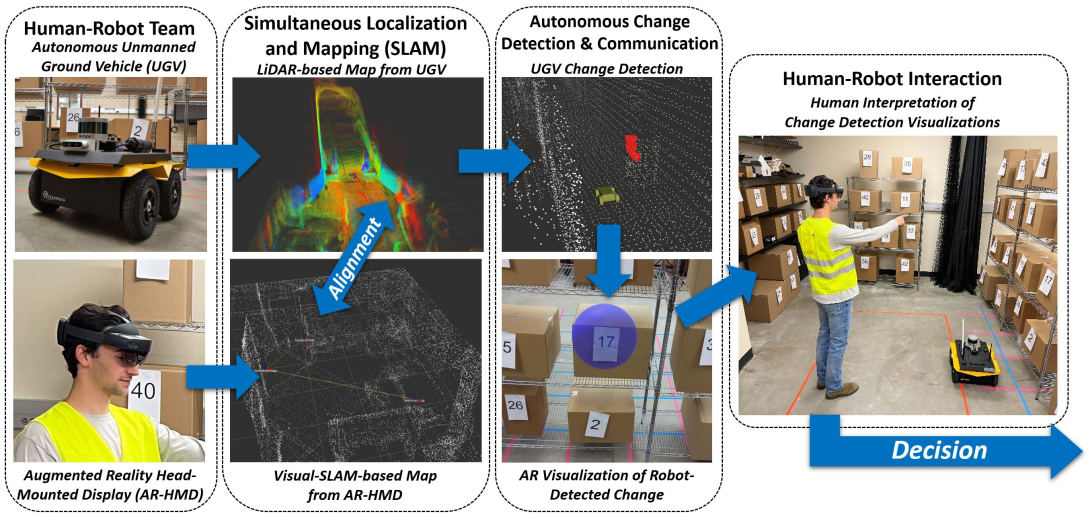


Fig. 1. End-to-end system operational concept. From left to right, top to bottom: A robot equipped with a LiDAR sensor is teamed with a human wearing an AR-HMD; the robot and the AR-HMD both perform SLAM and the robot consumes the AR-HMD map to calculate the live transformation between their respective maps; the robot maneuvers through the environment and performs change detection; alignment allows detected changes in the robot's coordinate system to be visualized live to the human via the AR-HMD; the human quickly interprets these changes to make a better-informed decision in the real world. Importantly, all processing (SLAM, change detection, communication, visualization) are done on-board the robot and AR-HMD, and this system is readily deployable to new environments as it requires no external instrumentation or engineering of the environment.

the environment through an AR headset (see Fig. 2). This interface is sufficient to communicate the acquired environment data on changes and disambiguate change information in the operational space [76].

One factor that occurs when humans form teams is that humans are continuously trying to make sense of other's behaviors. Similarly, when humans form teams with robots, they will try to make sense of robot behaviors. For example, people might form expectations towards a robot based on surface-level cues (e.g. robot appearance) but the robot does not have the capabilities to match those expectations [25]. It also has been shown that people infer more information than is specifically intended to be communicated, for example when robot actions or behaviors are interpreted even though not they are not relevant to the task [6]. It is expected that some of the robot behaviors or the robot communication is interpreted by the human participants in a way that was not anticipated or intended. For this reason, it is crucial that the robot system is being tested with naive human participants and that potential misinterpretations of robot behaviors or communications are avoided in future iterations to build a more effective and efficient system.

In addition to communication, it has been shown that trust plays a crucial role in teamwork and cooperation [30]. Similar, trust and especially calibrated trust are crucial factors in human-robot teams [33, 57]. While there currently are no comprehensive trust models for all robots in mixed human-robot teams, recent research points to robot characteristics like performance and appearance in a given environment [4, 24]. This research looks at robot as an interdependent team member as the robot has capabilities that the human does not (i.e. detect a change in the environment) and therefore increases human performance in the task (i.e. identify the change). This research did not manipulate the appearance of the robot across the studies, but rather looked at robot performance in terms of the quality of communicating changes to the human.

This work seeks to address the above challenges of effective communication of useful metric perception data from a robot to a human teammate to provide enhanced situational understanding of the environment (Fig. 1). The robotic system used in this work runs a full AI autonomy stack, including navigation, perception (including change detection and teammate localization), reasoning, and decision making. Given change information from the robot's perception system, we seek to improve the robot's intelligence in deciding how to most effectively communicate that change information. Our contributions therefore lie between the perception and action selection stages of traditional AI robotics. While we focus on change detection information specifically, we seek to answer questions that will ultimately enable a robot to more intelligently make communication decisions in human-robot interaction applications. First, we present the design of a novel robot-enabled change detection system where the changes in the environment are automatically detected and communicated to the human that can visualize the information in AR. To the best of our knowledge, this is the first system that couples a human wearing an AR-HMD with an autonomous mobile robot that possesses Simultaneous Localization and Mapping (SLAM), teammate localization, and change detection capabilities for improving information flow and situational awareness. Second, we evaluate our proposed system with multiple online and in-person user studies where humans identify what changes the robot detected. We test several postulates regarding effective communication via AR in this manner to examine the user performance and experience impacts of immersive visual AR effects, in-person experience, the creation of aggregated visualizations, instructional explanation of the system, prior exposure to video game systems, and visualization error on understanding and user experience. This is a major extension of the authors' prior early work and is the first comprehensive evaluation of augmented reality communication and user experience methods' effect on understanding mobile robot change detection. While we focus our experiments on industry-relevant representative scenarios, our results generalize to other application domains, particularly those where metric information such as change in uninstrumented environments is perceived by a robot and communicated to a human teammate.

1.1 Related Work

1.1.1 Reality-Modification Technologies for Human-Robot Teaming. Perhaps the single most important enabling capability for successful human-robot teams is effective and efficient communication. Recent advances in augmented reality technology have led to a rising popularity of developing mixed reality (MR), virtual reality (VR), and augmented reality (AR) communication methods to address this essential capability. It is now well-established that reality-modification technologies such as MR, VR, and AR have significant advantages as human-robot interaction-mediation technologies [23, 65] and have recently been used for immersive remote command-and-control supervision of robot teams [55, 72]. It has been found that in order to make human-robot teaming efficient, it is vital to establish grounding, situational awareness, a common frame of reference and spatial referencing and that Augmented Reality (AR), the overlaying of computer graphics onto the real worldview, can provide the necessary means for a human-robotic system to fulfill these requirements [19, 37]. For example, AR can be used to achieve intuitive and seamless human–robot-environment interactions by incorporating the collaborative robots' additional capabilities [2]. Recent work has found that the contribution of AR visualizations is so significant that even when combined with complex language and gestures creating high cognitive load, human-robot communication is significantly improved [69]. AR has recently been used both communicate robot motion intent in human-robot joint tasks [56] as well as providing sequential waypoints for UAVs' global path planning [3].

The authors have previously developed and studied AR, VR, and MR technologies used for enabling human-robot teaming in applications such as cooperative search [51], cooperative teammate-aware exploration [53], robot tasking and control [21], teaming from the tactical edge to the operations center [14], temporally-coordinated cooperative human-robot team search [42], and maintaining shared situational awareness with multi-robot systems [23]. This work builds upon previous efforts to enable human-robot teaming specifically through AR communication, with a detailed and comprehensive investigation of factors affecting human understanding of AR visualization of robot change detection information. Specifically, the work in this paper leverages previous efforts that created a fully autonomous, intelligent robot to navigate an environment with a human teammate and, e.g., share information about its search [51] and exploration [53] behaviors. The authors' prior work documented the system design of a robot capable of sharing raw change detection information [49] via AR (without a human study). Additional prior work [50] presented partial results from what is referred to as Study 1 in this paper, analysis of which is extended herein and placed in the context of a three-part study design.

1.1.2 Autonomous Robot-Based Environmental Change Detection. Novelty or change detection is the process of comparing current metric sensor readings against a previous reading or prior model to identify differences, which indicate change or novelty [45]. In this work the change detection is performed online by an autonomous robot, and the sensor space is all changes in the robot's environment; hence we term this environmental change detection. The robot's goal is to identify metric differences between its prior model and what it is currently sensing in the physical environment. The focus of this work is then how the robot can most effectively communicate that information.

There are numerous impactful applications for autonomous robot-based environmental change detection, such as robust outdoor navigation [63], inspection [36], surveillance [41, 71], and safety and security [64]. Preliminary work by the authors used autonomous robot-based environmental change detection to show the potential for human understanding of change detections communicated via AR [21, 52]. This work significantly expands upon these preliminary studies to provide a comprehensive quantitative analysis of communication strategies, instruction, errors, and human experience across prior and new crowdsourced and in-person human-subjects experiments. While the studies conducted for this research particularly demonstrate the utility of autonomous robot-based environmental change detection for the industry 4.0-relevant domain, we observe that the results generalize to other domains where a robot equipped with a depth sensor such as LiDAR can traverse and scan an uninstrumented environment.

1.1.3 Data Visualization and HRI. There currently is a large need to represent robot-to-human information exchange in a way that supports human decision making. However, the interfaces used to facilitate this decision making do not always follow best-practice principles of data visualization [67]. A full survey of data visualization used for human-robot-interaction is beyond the scope of this work. Relevant design guidelines from robotics research can be found in [1, 10, 12, 31, 39, 66, 80, 81] and from visualization research can be found in [5, 7, 40, 54, 60]. We briefly review here the research we grounded our visualization in. We decided to use AR glasses as means of communicating information from the robot to the human. We based our decision on the possible applications of such a human-robot team where it could be necessary that the human and the robot are spatially separated (e.g., exploration of a volatile environment), the human needs to maintain situational awareness of the environment they are in (e.g., hostile environment), and the increased ability to communicate uncertain, complex, or ambiguous data that might not be possible to communicate through a different modality (e.g., verbal cues). Bringing together two novel technologies, AR and robots, in a human-robot team scenario implies that an interdisciplinary evaluation is necessary to determine the the functionality of the robot system including SLAM and human detection, the connection to the AR system, the detection of changes, the visualization of changes, and ultimately, the capabilities of the overall team. AR-based interfaces have been proposed as efficient interfaces to facilitate human-robot interaction [8, 9, 15, 18, 46, 73]. It however remains challenging to establish AR systems due to environment complexities, and the difficulties in human–robot collaboration [74]. It further remains an open research question how AR can significantly improve situational awareness [38] in human-robot teams and how AR can increase effective communication and informed decision making in uncertain environments [68]. This research explores the feasibility of a human-robot team assessing the robot-to-human communication of environment changes through AR.

2 METHODOLOGY

2.1 Point Cloud Methods for SLAM, Teammate Localization, and Change Detection

2.1.1 SLAM. Our robot performs an onboard SLAM process described in [48]. Briefly, the SLAM process utilizes a refined version [20] of *OmniMapper* [70]. A *pose graph* of point cloud measurements taken adjacently in space is constructed, and loop closures are used to refine the graph. A solution to the robot's trajectory is computed via least squares using the GTSAM nonlinear optimization framework [11] based upon square root smoothing and mapping with point cloud measurements as input. To build a model point cloud representing the environment (Fig. 1), point clouds along the optimized trajectory are sampled and projected into a common frame of reference.

2.1.2 Teammate Localization via Frame Alignment. In order to share information like change detection, the robot must perceive the human's position and orientation. We achieve this by ensuring the robot and human-worn AR-HMD share a reference frame. We use the frame alignment process presented in [51] to accomplish this task online. In this process, the point cloud representation generated by the robot using the SLAM process above is aligned with a point cloud representation of the AR-HMD's map (Fig. 1). The AR-HMD's map is generated by proprietary stereo vision-based mapping software, which we treat as a black box and directly utilize the mesh map output representation as input into our alignment algorithm. We translate the AR-HMD's mesh-based representation into a point cloud format and then compute the heterogeneous transformation matrix between the robot and AR-HMD point clouds using the Iterative Closest Point (ICP) algorithm [58]. The human participant provides an initial coarse estimate via the AR interface that is used for the transformation matrix calculation, which is then updated as the human and robot's maps update. This transformation matrix allows us to seamlessly translate any position in the robot's coordinate frame into the human's coordinate frame, and vice-versa.

2.1.3 Change Detection. The change detection process involves the comparison of two models: a *prior* of the "clean" environmental state is compared against a *test* model of the current state. Both of these models are built by the robot using the SLAM process described above. The *test* model is either built completely prior to processing, or can be incrementally processed online to identify changes on the fly as the robot patrols. Analysis of changes between the *prior* and *test* point cloud model require an approximate reference frame generated by either a coarse alignment similar to the reference frame alignment above, or physical co-origination of the human and robot agents. This approximately aligned reference frame is then refined online using generalized ICP [59]. Proper alignment of the *prior* and *test* models allows for change detection to take place. Change detection is implemented using PCL [58] using difference segmentation routine that builds a KD-tree of the model, reducing the quadratic search complexity to $n \log n$. This allows us to compare each point in the *test* cloud with the *model* cloud to find the nearest point. We define three hyperparameters for this algorithm. If the distance to the nearest point exceeds a threshold δ, that point is copied into a *change* point cloud for further filtering. In the filtering step, we attempt to filter out detections from the *change* model caused by range error noise in the sensor itself or quantitation error caused by the mapper. This is accomplished by requiring λ supporting detection points within a radius r.

While our approach is capable of detecting changes of addition (objects added to the scene) and deletion (objects removed from the scene), for this work we choose to focus on changes of addition. We justify this decision based on 1) clarity, i.e. study participants are most easily able to clearly identify objects present as being changed (as opposed to objects missing), 2) simplicity, i.e. to support our experimental design, where the participants do not have prior knowledge of the environment over which they could reason about changes of deletion. We also note that other types of change exist such as relocation, which requires object tracking to create a correspondence between a deletion and addition, and modification, where an object has changed, e.g., in appearance or orientation. We do not address these perception problems in this work, but examining differences in communication of these types of change would certainly be suitable for future work.

2.2 Change Visualization

Even with changes detected and filtered, there still exists a significant challenge to be overcome. While presenting the resulting point-changes directly to a human via AR is possible [48], even in controlled test environments there could be hundreds to tens of thousands of individual point changes, depending on the scale of the environment and the resolution of the sensor. This high volume of change information presents a challenge to effective communication to the user. Therefore, the robot should take steps to aggregate change information before presentation to the user. This process scales down the number of visualizations that the human may interpret and interact with to a tractable quantity.

We perform this aggregation using the Density-Based Spatial Clustering of Applications with Noise (DBSCAN) [43] clustering technique applied to the *change* model. Each resulting non-noise cluster with over a threshold member quantity is then published as a visualization object by the robot. The threshold value, along with the ϵ parameter of DBSCAN, which is the maximum distance between two samples for one to be considered as in the neighborhood of the other, are the primary hyperparmeters for this method. Both are adjusted empirically and dependent on the environment, sensor resolution, and size of expected change. The AR-HMD consumes the published message and then creates a corresponding visualization in AR to communicate the change with the human. Because the individual change points represent the edge of the change facing the robot, we create a transparent spherical AR visualization centered on then centroid of the change cluster with radius scaled by the intracluster distance to roughly encompass the region containing the changes.

3 EXPERIMENTS AND RESULTS

We begin with a discussion of hardware and software used in these experiments in Sec. 3.1, and then discuss studies, corresponding results, and finding-based motivations for the next study. Over the course of our three studies, we pose 10 postulates to explore and assess the performance of the system and the human's ability to understand information curated by the robot. Our postulates are motivated by the vision of deploying intelligent, mobile robots that effectively communicate salient information to humans for situational awareness with minimal user training or technical overhead. We pursue these postulates because they serve to develop a deeper understanding of the role and characteristics of AR visualization for conveying robot change detection to humans as it relates to situational awareness. Given the prominence of robotics in industrial settings and, to the best of our knowledge, that no such AR-based human-robot teaming technology has been deployed to date, we believe that studies of relevant systems, such as the one presented in this work, are warranted.

In Sec. 3.2, we discuss the first two video-based studies focused on determining interpretability of change detection information. Study 1 is an initial evaluation of our system that focuses on the intuitiveness of the visualizations. Study 2 collects additional data and adds simple prior instructions to attempt to improve performance and dynamic perspective to better understand how a more immersive quality affects the human's comprehension. Study 3 (Sec. 3.3) incorporates the lessons learned from Studies 1 and 2 and evaluates the AR-HMD and robot during an in-person study to explore the impact of a fully-immersive, live system. Across all three studies we collected survey data from the participants to determine the impact of prior experience in the vein of real-world applicability. We enumerate and present findings with respect to unique postulates, $\mathcal{P}_i$, since some research questions span multiple studies. An overview of all 10 postulates collectively posed in this work, and the corresponding study in which they were investigated, is shown in Table 1.

Table 1. Overview of postulates explored in our three studies. Support findings discussed in Sec. 4.

$\mathcal{P}_i$	*Overview*	*Impact of*	*Postulate* *Study 1*	*Study 2*	*Study 3*	*Support*
$\mathcal{P}_1$	Changes are interpretable without instruction	Instruction	$Study1_1$	$Study2_1$		Y
$\mathcal{P}_2$	Clusters decrease participant's accuracy	Display type	$Study1_2$	$Study2_2$		N
$\mathcal{P}_3$	Original and mirrored videos are equivalent	Video	$Study1_3$	$Study2_3$		Y
$\mathcal{P}_4$	Accuracy increases by adding instructions	Instruction		$Study2_4$		Y
$\mathcal{P}_5$	Accuracy increases with dynamic perspective	Immersion		$Study2_5$		Y
$\mathcal{P}_6$	System correctly displays ≥ 90% of changes	Live system			$Study3_1$	N
$\mathcal{P}_7$	Participant's error decreases in-person	Live system			$Study3_2$	Y
$\mathcal{P}_8$	Ambiguities reduce understanding	Live system			$Study3_3$	N
$\mathcal{P}_9$	Fewer ambiguities correlate to better experiences	Live system			$Study3_4$	Y
$\mathcal{P}_{10}$	Gaming exposure does not impact performance	Experience	$Study1_4$	$Study2_6$	$Study3_5$	Y

3.1 Hardware and Software

Because our change detection algorithm is SLAM and pointcloud-based, it can be run on any robot that has a LiDAR or color-depth sensor, an IMU sensor, and sufficient computing resources onboard. For these experiments, we used a Clearpath Robotics Jackal mobile ground robot equipped with a Velodyne VLP-16 LiDAR, MicroStrain 3DM-GX4-25 IMU, and Intel Core i5 CPU (Fig. 2a). The robot is capable of movement at speeds up to $2m/s$ and for these experiments was teleoperated by the experimenters around the environment to build the initial *prior* and *test* models. Change detection was performed online by the robot and visualization was communicated via Wi-Fi to a human wearing an AR-HMD. For the first two studies in Secs. 3.2.1 and 3.2.3, a first-generation Microsoft Hololens was used as the AR-HMD (Fig. 2b). For the third study (Sec. 3.3), a Microsoft Hololens 2 was used (Fig. 2c). All software for these experiments is a combination of custom-developed software written in C++, C#, and Python (e.g., for SLAM, change detection and teammate localization via frame alignment, as described in Sec. 2) and builds upon the Robot Operating System (ROS) middleware [47].

(a) (b) (c)

Fig. 2. Experimental hardware used: **(a)** A Clearpath Robotics Jackal; **(b)** Microsoft Hololens; **(c)** Microsoft Hololens 2.

3.2 Interpretability of Change Detection via Video Capture Studies

In order to evaluate the interpretability of change detection information as well as the aggregation of change detection information, we conducted two studies. Because of restrictions for in-person experiments imposed by the COVID-19 pandemic, video-based studies using crowdsourced participants were performed. While this imposed limitations on the user experience (discussed in Sec. 4.1), it also allowed us to test a variety of postulate in support of the objectives over a large study base. The studies' objectives were generally directed at understanding how participants interpret the information presented to allow for further testing and refining of visualization methods prior to an in-person study.

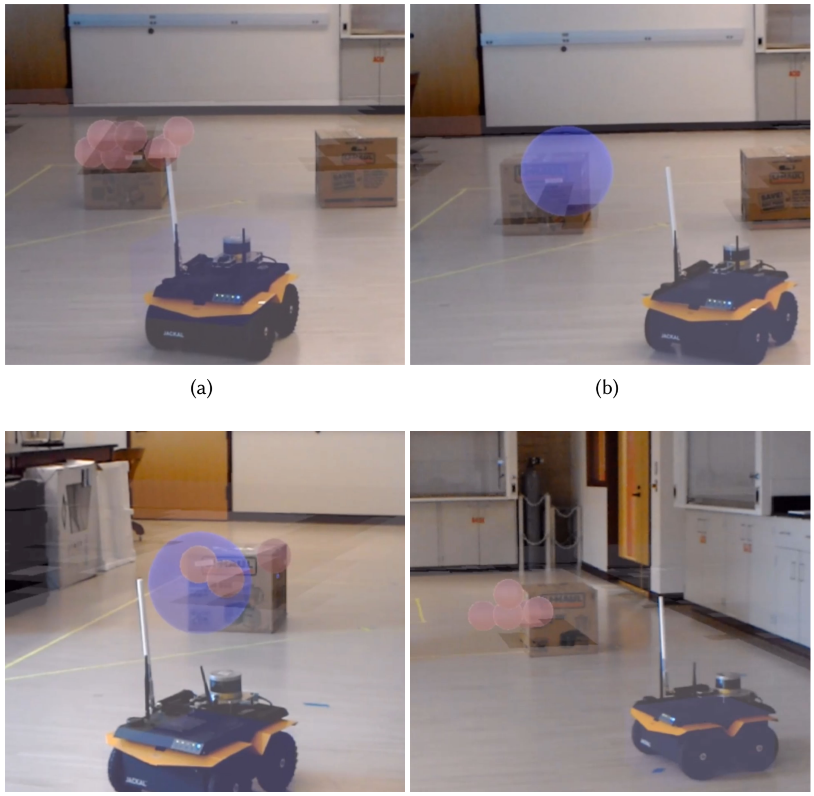
Experiments took place in an indoor laboratory setting approximately $200m^2$ in size. For these initial studies, a less-cluttered environment was tested before moving to a more challenging environment in the in-person experiments. In each experiment, one to three cardboard boxes of identical size (~$0.5m$) were used as change objects. In each experimental trial, the change relative to the *prior* setting was that either none, one, or two boxes were added to the room to predetermined positions. Boxes were located evenly spaced in a line perpendicular to the starting position and orientation of the participant. Distractor objects in the room such as cabinets and doors were included in the survey. While these distractor objects were further outside the field of attention of the participants than the boxes, they did elicit false positives (see Sec. 3.2.2), and significantly more distracting distractors were incorporated into Study 3 (Sec. 3.3). No changes occurred outside of the experimentally induced changes.

In both video-based studies, the robot performed change detection and visualization online. Live videos were recorded of communicated change visualizations in real time onboard the AR-HMD using the device's built-in video capture capability. While not the full AR experience, because these videos were taken online from the perspective of a user wearing the AR-HMD, this allows us to evaluate the AR visualizations' interpretability. These videos were ultimately presented to the participants online and their interpretations of the change detections were recorded.

As described in Sec. 2.1.3, an experiment consists of two phases. First, the robot explores the environment traversing a pre-computed path for the first time and builds a *prior* model. In this initial phase, one, two, or three boxes are already present in the environment. In between the first and second phases, the experimenters make one, two, or no box additions to the environment. Then, in the second phase, the robot again surveys the environment and builds a *test* model. Per our change detection process, corresponding points between the *prior* and *test* models with distances over a threshold value were treated as changes by the robot, filtered to reduce noise as described in Sec. 2.1.3. Changes were then communicated to the human via the AR-HMD (Sec. 2.2). As noted in Sec. 2.1.3 it is possible to either compile a complete model and then detect changes, or to run the change detection online and display changes detected incrementally, on the fly. In our case, the robot performed the latter, and video was recorded of the AR-HMD displaying changes as they were detected and communicated by the robot. For the hyperparameters described in Sec. 2.1.3 we have empirically found via real-world experimentation using different LiDAR sensors with varying fields of view that $\delta = 0.1m$, $\lambda = 10$ and $r = 0.3m$ work well for the online, incremental communication strategy [48].

3.2.1 Study 1: Aggregation of Change Information. To address the challenge of presenting large volumes of change data to the human teammate in an understandable and interetable way, the first study focused on a comparison of visualizations of aggregation of change data against raw change data visualizations. Using the methods in Sec. 2.2, three categories of visualizations were created for this study: 1) individual point-change "raw" data, illustrated by red spheres of a small, fixed size ($0.1m$ in our experiment, empirically tuned based on the average inter-change distance given the sensor resolution and expected robot distance from changes), 2) visualization objects represented clustered individual point-changes, and 3) both individual and clustered visualizations combined. Examples of these visualization methods are shown in Figs. 3a–3c. We note that for 1) the quantity of visualization objects is equal to the quantity of points in the filtered *change* model; for 2) the quantity of objects is substantially less; for 3) the quantity of objects is the sum of 1) and 2). Videos for each of these three categories, for all combinations of box locations, with zero (baseline), one, and two boxes changed were generated, resulting in 19 videos total: 1 (no change) + 3^2 (one change) + 3^2 (two changes). Of the 19 videos, six were generated from their "mirror-image" videos, e.g., a change of the left-most object was mirrored to create a video with a change of the right-most. These were included as a control, as any difference between them and the non-mirrored versions might indicate a design error or crowdsourced population issue. Videos used in this study were taken from a fixed perspective; dynamic perspective was investigated in Study 2 (Sec. 3.2.3).

The study was determined as exempt by the Institutional Review Board (IRB) of the University of Denver. Participants received monetary compensation for their time. Participants ($N = 21$) were recruited online and shown the 19 videos, one at a time. Videos contained 0, 1, or 2 true positive changes. Each of the video with changes showed each of the three visualization categories (3^2 with one change + 3^2 with two changes). Videos were shown in randomized order to counter potential learning effects introduced by order. Because the videos were collected from a real-world system, occasional uncorrected false positives and infrequent small errors in the placement of the visualization objects were present. A representative example of these small visualization errors is shown in Fig. 3d, where a slight misalignment between the robot and the AR-HMD causes the visualization markers to be displaced to the left of the changed object. As the goal of Study 1 is to evaluate the current version of the change detection visualization, those errors remained in the videos. The degree of alignment error is consistent with what is shown in Fig. 3d and occurs in less than 10% of the videos. Two participants were excluded: one for failing to complete the survey on-time, and the second for declining to participate, leaving $N = 19$ participants ($M_{age} = 36.1, SD_{age} = 7.7$, 12 Male, 7 Female).



(a) (b)

(c) (d)

Fig. 3. Examples of the visualization categories: **(a)** individual point changes; **(b)** clustered changes; **(c)** both points and clustered changes; **(d)** and representative visualization error, in this case where visualizations are offset slightly to the left relative to the viewer's perspective.

For Study 1, we postulated that:

$Study1_1$: Changes will be interpretable without instruction.

$Study1_2$: Participant's accuracy in recognizing change will decrease when visualizing clustered data relative to visualizing discrete point data.

$Study1_3$: There will be no significant difference in participant's accuracy to recognize change when they view original or mirrored videos.

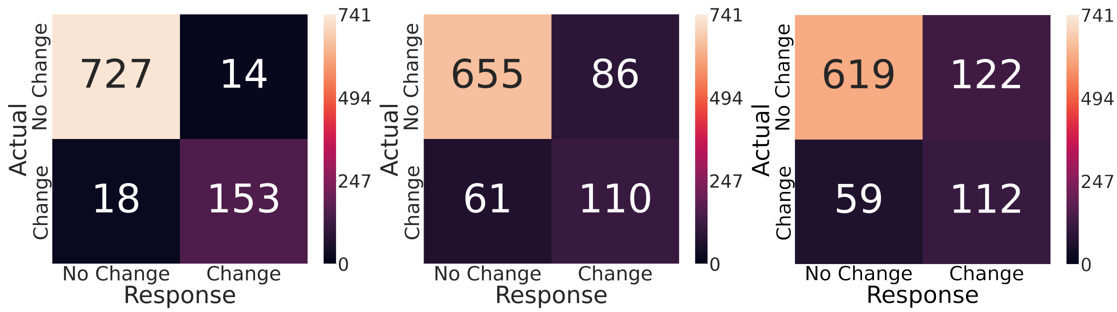
$Study1_4$: Prior exposure to video games will not impact performance

3.2.2 Results Study 1. In Study 1 the participants were shown a series of videos, each containing zero, one, or two changes. After viewing a video, the participant was asked to indicate their level of confidence as to whether a particular object changed by selecting from four options: definitely changed, probably changed, probably did not change, and definitely did not change. In addition to the three possible objects that could have changed in any given video, the participants were also asked about five additional objects within the field of view. These five objects never changed in any video and were included to measure the participant's attention and understanding of each scene. We computed confusion matrices using the participant's responses, as shown in Figure 4, where a true positive corresponds to the sum of participants that stated an object definitely or probably changed when it had, a false positive corresponds to the number of participants that stated an object definitely or probably changed when there was no change, a false negative corresponds to the number of participants that thought an object definitely or probably did not change when there was a change, and a true negative corresponds to the sum of participants that stated an object definitely or probably hadn't changed when there was no change.

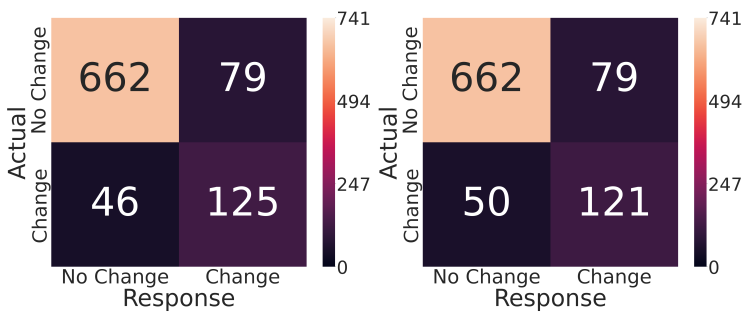
We are primarily focused on the effect of aggregation of change information in Study 1. As such, we computed confusion matrices, accuracy, and error rate based on the control variable of visualization marker type. We use accuracy as a measure for how well participants correctly identify change detections and calculate this metric as the sum of true positives and true negatives divided by the total number of responses. Similarly, we use error rate (sometimes referred to as misclassification rate) as a metric for how many participant responses are incorrect, which is calculated as the sum of false positives and false negatives divided by the total number of responses. From Figures 4a–4c, we see the cluster, points, and both marker types led to a participant accuracy of 0.965, 0.839, and 0.802, respectively. We also consider the effect of showing participants mirrored versions of the original videos. Figures 4d and 4e show that the participants scored an accuracy of 0.863 and 0.859 for the original and mirrored videos, respectively.

3.2.3 Study 2: Situated Dynamic Perspective and Instruction. Where the first study utilized a fixed visualization perspective and examined the interpretability of visualizations without prior instruction, the second study added additional conditions. Specifically, the effects of the following were examined by the second study: 1) dynamic perspective caused by moving viewpoint on the interpretability of the change visualizations (Figs. 5 and 6), and 2) instructional information regarding the meaning of visualizations provided to the participant prior to the study (Figs. 7 and 8). The motivation for 1) was to incorporate the sort of realistic, immersive perspective change that would be experienced by a participant in an actual live user experience that was not present in the first video study. The purpose of 2) was to evaluate if a short instruction video elicited a difference in participant accuracy. For example it was noted in the first study that in the absence of any instructions participants ascribed their own meanings and interpretations to the color and size of visualizations, when in reality the colors distinguished between point (red) of fixed-size and aggregated (blue) change detections of size proportional to the intracluster distance (discussed further in Sec. 4).

To represent 1) above, two sets of videos were taken from dynamic perspectives, i.e. with the user in motion. The motion patterns demonstrated were *panning* and left-right *traversal*, keeping the objects in the center of the field of view (a motion similar to side-to-side "strafing" in modern first-person video games), as shown in Figs. 5 and 6, respectively. One video set for each perspective change motions were collected, with corresponding videos for each of the 19 conditions from Study 1 (Sec. 3.2.1) where 0, 1, or 2 boxes were changed, for a total of $2 * 19 = 38$ videos, which included 12 mirrored videos (six for each motion pattern) created as described in Sec. 3.2.1. For 2), four new videos were recorded. One instructional video, shown in Figs. 7 and 8, was created as well as demonstration videos for each of the three visualization types (point, cluster, both) using a $\sim 0.7m$ tall cabinet of similar size and shape to the boxes in the experiment videos as a demonstration object (see Figs. 8a–8e). The experimental design for Study 2 showed all 38 videos to each of the participants (within condition) with one group of participants having video instructions vs. no instructions (between conditions).



(a) Cluster markers (b) Point markers (c) Both markers

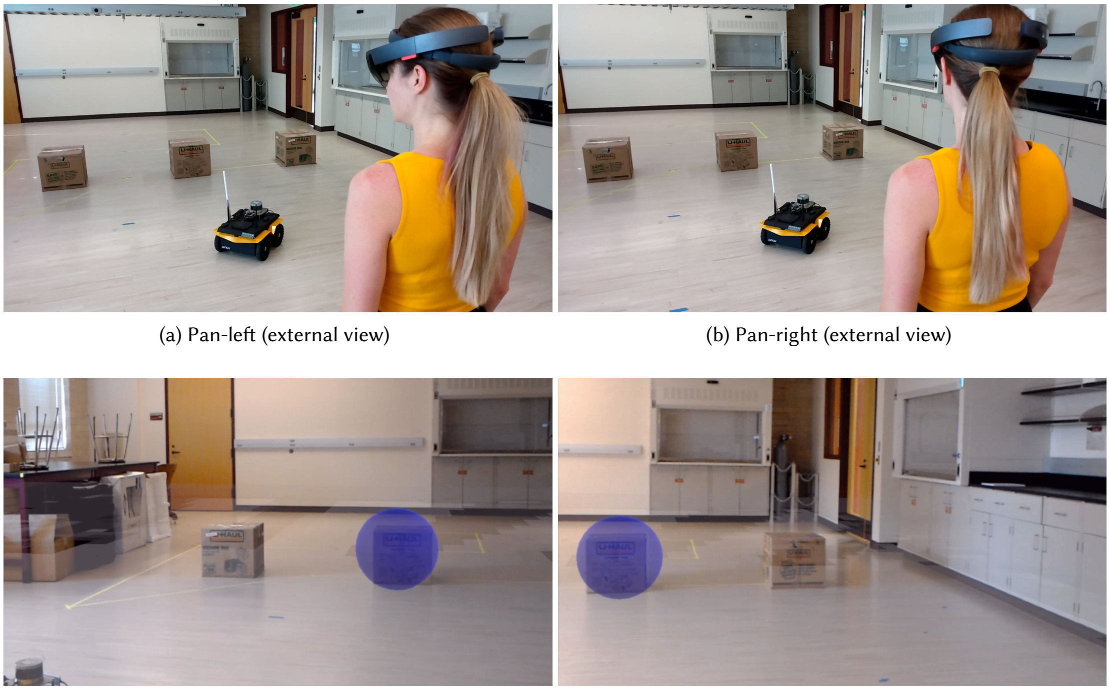


Control Variable	*Accuracy*	*Error Rate*
Cluster	**0.965**	0.035
Points	0.839	0.161
Both	0.802	0.198
Original	0.863	0.137
Mirror	0.859	0.141

(d) Original video (e) Mirrored video (f) Statistics

Fig. 4. Results from Study 1. **(a)–(c)** confusion matrices for different visualization markers; **(d)** and **(e)** confusion matrices for video type; **(f)** accuracy and error rates.

The study was determined as exempt by the IRB of the University of Denver. Participants received monetary compensation for their time. Participants ($N = 120$) were recruited online and each of the four conditions (no instruction vs. instructions × regular vs. mirrored video) had ($N = 30$) participants assigned in a between experimental design. Each participant was shown the 38 videos, one at a time, as in Sec. 3.2.1, and the same question format was used to allow for a direct comparison of effects. No participants were excluded from the study ($M_{age} = 42.9, SD_{age} = 9.7$, 55 Male, 65 Female).



(a) Pan-left (external view)

(b) Pan-right (external view)

(c) Pan-left (internal view)

(d) Pan-right (internal view)

Fig. 5. Examples of the dynamic motion patterns used in the Study 2 videos. **(a)** and **(b)** show the external and **(c)** and **(d)** show the corresponding first-person internal views of the *pan* movement, where the perspective is dynamically varied by the user rotating her head from left to right.

For Study 2, we postulated that:

$Study2_1$: Changes will be interpretable without instruction.

$Study2_2$: Participant's accuracy in recognizing change will decrease when visualizing clustered data relative to visualizing discrete point data.

$Study2_3$: There will be no significant difference in participant's accuracy to recognize change when they view original or mirrored videos.

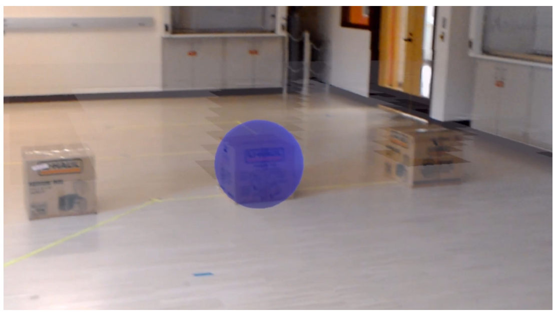
$Study2_4$: Participant's accuracy in recognizing change will increase when participants are provided with instructions.

$Study2_5$: Participant's accuracy in recognizing change will increase when crowdsourced videos include dynamic viewing angles, compared to videos with stationary viewpoints.

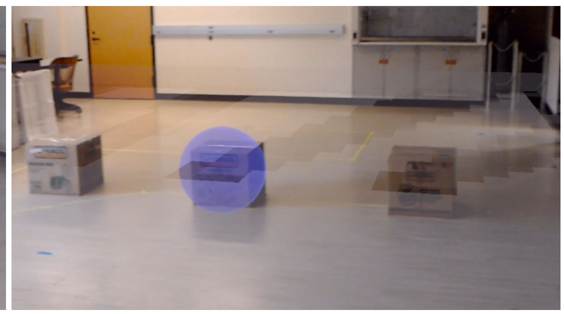
$Study2_6$: Prior exposure to video games will not impact performance.

(a) Traverse-left (external view)

(b) Traverse-right (external view)



(c) Traverse-left (internal view)



(d) Traverse-right (internal view)

Fig. 6. Examples of the dynamic motion patterns used in the Study 2 videos. **(a)** and **(b)** show the external and **(c)** and **(d)** show the corresponding first-person internal views of the *traversal* movement, where the perspective is dynamically varied by the user walking to the left and right extremes of the environment while keeping the room centered on her field-of-view.

3.2.4 Results Study 2. Study 2 placed emphasis on the aspects of situated dynamic perspective as well as the role of instruction in human understandability of robot-based changed detection. Similar to Study 1, participants were shown short videos with zero, one, or two changes – this time the viewpoint included pan or left-to-right traversal movement – and then they were asked to provide their confidence in the presence of change for the eight possible objects. Confusion matrices comparable to Study 1 are shown in Figures 9a–9c, which highlight the participants' understanding of the videos controlled for when they saw changes indicated by clusters, discrete points, or both marker types. Here, participants achieved an accuracy of 0.934, 0.910, 0.933, respectively. Confusion matrices for the original and mirrored videos in Study 2 were also calculated and shown to have accuracy scores of 0.925 and 0.924, respectively.

In terms of the inclusion of dynamic movement in the change detection videos shown to participants, confusion matrices for the pan and left-right traversal videos are presented in Figures 10b and 10c. While the stationary videos from Study 1 led to a participant accuracy of 0.868, the addition of pan and strafe movement in the group of Study 2 participants who did not receive video instruction increased accuracy to 0.921, and 0.933, respectively. Lastly, confusion matrices showing the resulting responses for participants that received instructions, compared to those participants that received no instructions in Study 2, are provided in Figures 10d and 10e. Under these conditions, the participants' accuracy was measured to be 0.951 and 0.898, respectively.

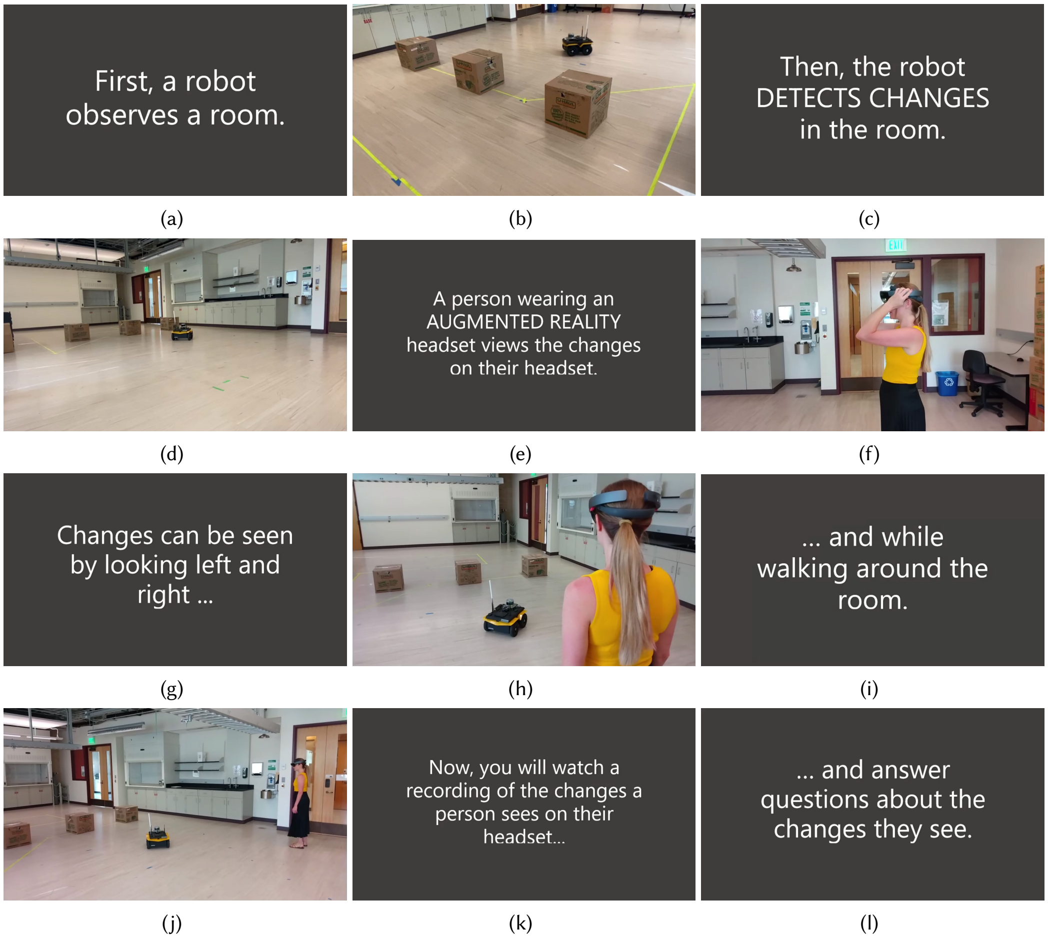


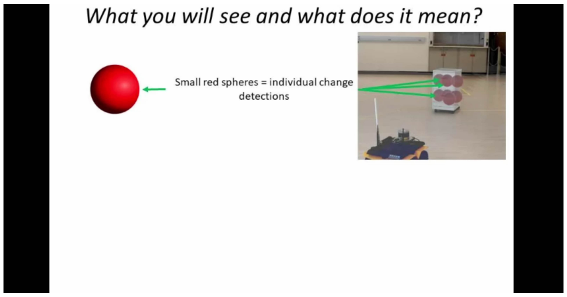
Fig. 7. Screen captures from the segments 1-12 of the instructional video for Study 2.

3.3 Study 3: In-Person Inventory Management User Study

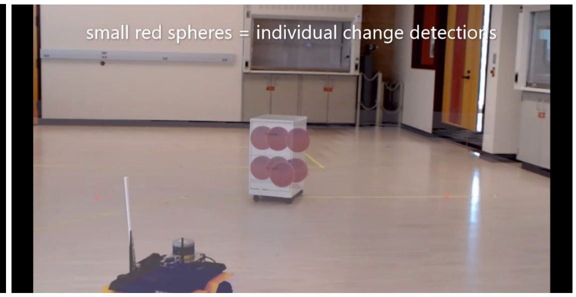
The third and final study incorporated the lessons learned from the prior two online studies, especially the aggregation of data, the use of instruction, and perspective change. The motivation for this study was to evaluate the system in a realistic, in-person experiment with a significant number of participants in order to 1) demonstrate the viability of the approach, 2) evaluate the impact of error in visualization placement on interpretability with in-person user experience, 3) examine the user experience.

The scenario for this experiment was designed to replicate warehouse workers searching for missing/changed inventory around a storage warehouse environment. The environment was set up using 3 walls of $2.5m$ high shelves in a U-shape, with each shelf holding boxes of varying sizes. Each box was labeled with a unique number for unambiguous identification, as shown in Fig. 11a.

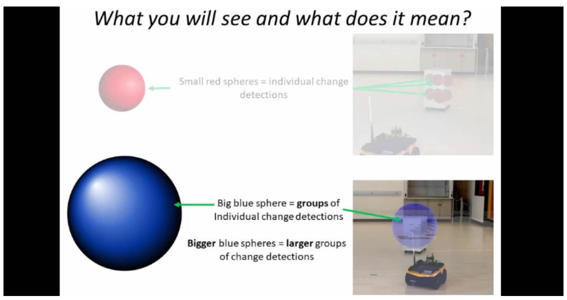
The general outline of a study session is as follows. When a participant arrives for the study, the experimenters provide the written consent form and an oral overview of the experiment. The experiment begins with a participant



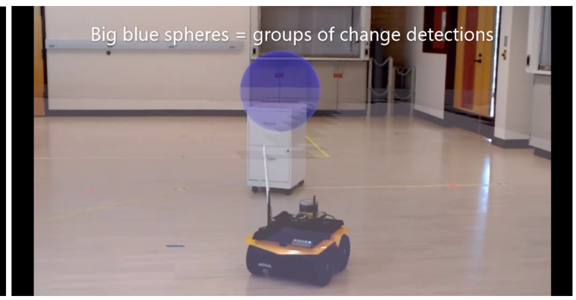
(a)



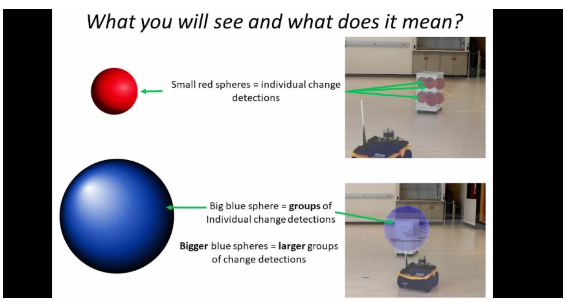
(b)



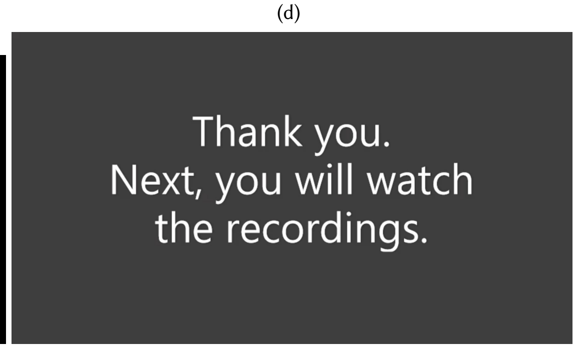
(c)



(d)



(e)

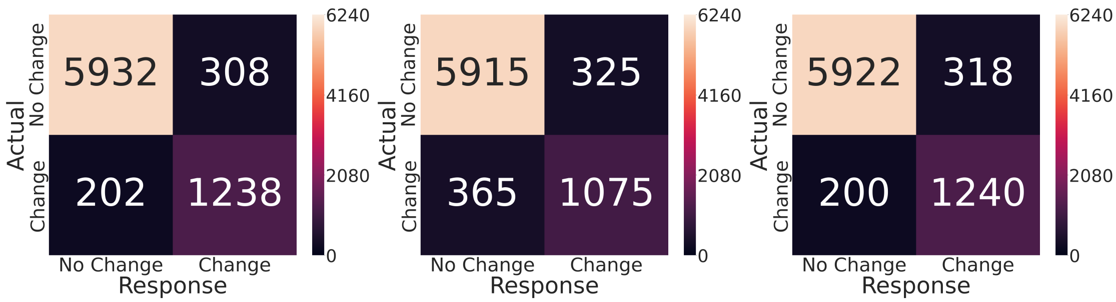


(f)

Fig. 8. Screen captures from segments 13-18 of the instructional video for Study 2.

entering the simulated "warehouse" environment with the AR-HMD on. Participants are asked to start on a mark in the middle of the environment, and are told they can walk anywhere within a space indicated by taped lines on the floor. This region was created to ensure the participants maintained enough distance from the shelves so that they could see the visualizations. A photograph of the environment with a person and robot observing the shelves is shown in Fig. 11b.

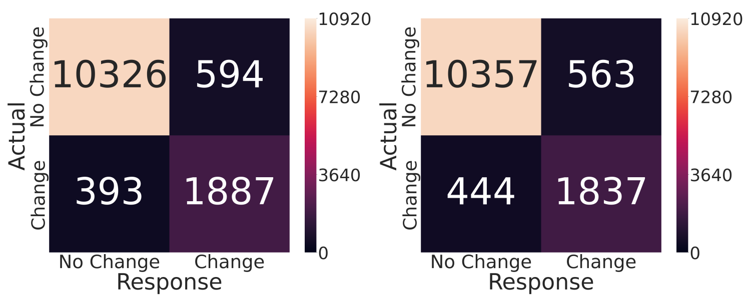
The experimenters then navigate the robot into the environment. While the change detection and visualization system is individually stable, as noted in Sec. 3.2.1, small errors in visualization do occur. Therefore, to ensure a uniform user experience, instead of detecting changes live, the robot plays back a set of change detections recorded during a prior live change detection by the robot. These detections reflect the actual changes in the room in the robot's perception coordinate frame and have been saved in a ROS `.bag` file. During the experiment, the robot performs online frame alignment to perceive the human's pose relative to its own (see Sec. 2.1.2). Using this alignment, the robot is able to



(a) Cluster markers

(b) Point markers

(c) Both markers



(d) Original video

(e) Mirrored video

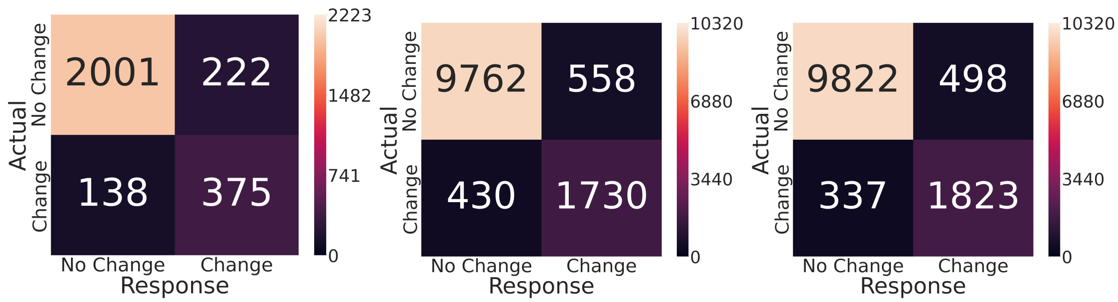
Control Variable	*Accuracy*	*Error Rate*
Cluster	**0.934**	0.066
Points	0.910	0.090
Both	0.933	0.067
Original	0.925	0.075
Mirror	0.924	0.076

(f) Statistics

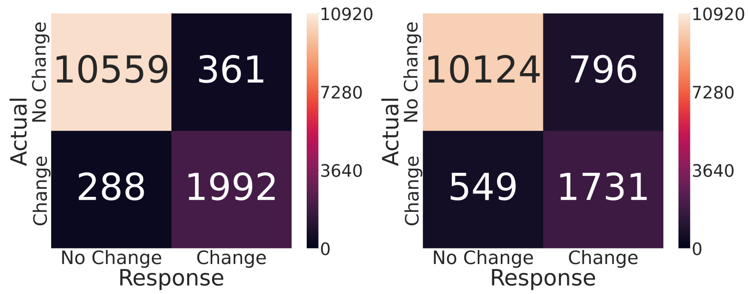
Fig. 9. Results from Study 2. **(a)**–**(c)** confusion matrices for different visualization markers; **(d)** and **(e)** confusion matrices for video type; **(f)** accuracy and error rates..

project visualizations of the detected changes into the correct locations in the AR-HMD. Thus, the only portion of this experiment that was not fully live was the use of live-recorded data of the exact same environment configuration. As the robot completes its patrol, it communicates the changes detected via AR through the AR-HMD, as illustrated in Fig. 11c. The participants then proceed to identify changes to the experimenters by verbalizing the numbers of the boxes that have changed. Once the participant believes they have seen all the visualized changes, they inform the experimenters. This process repeated for four additional change scenarios with between one and three different boxes changed in each, for a total of five scenarios. For control, the same first scenario is played for each participant. Then, the subsequent four change scenarios are played in a different, non-repeating order. During the study, the Hololens 2 AR-HMD also records video from its forward-facing camera for later analysis.

The study was determined as exempt by the IRB of the University of Denver. Participants ($N = 22$) were recruited through an e-mail newsletter at the University of Denver. Participants volunteered their time and did not receive compensation. Participants spent on average ~15 minutes in the lab where the study took place. After completing the steps described above, participants were asked to fill in a brief survey. No participants were excluded from the study ($M_{age} = 25.0, SD_{age} = 9.1$, 12 Male, 10 Female).



(a) Stationary (b) Pan movement (c) Traversal movement



(d) Instructions (e) No instructions

Control Variable	*Accuracy*	*Error Rate*
Stationary	0.868	0.132
Pan	0.921	0.079
Strafe	**0.933**	0.067
Instr.	**0.951**	0.049
No Instr.	0.898	0.102

(f) Statistics

Fig. 10. Aggregate results from Study 1 and 2. **(a)** confusion matrix for the stationary viewpoint from Study 1; **(b)** and **(c)** confusion matrices for the dynamic perspectives from Study 2; **(d)** and **(e)** confusion matrices for instructions; **(f)** accuracy and error rates.

For Study 3, we postulated that:

$Study3_1$: Our system can correctly identify and display detected changes with a live human for at least 90% of actual changes.

$Study3_2$: Participant's error rates will decrease for in-person tests, compared to the crowdsourced videos.

$Study3_3$: Ambiguities in visualization will reduce understanding of change detection.

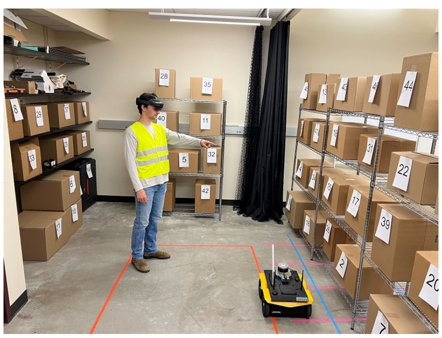
$Study3_4$: Users that experience fewer ambiguities will indicate a more favorable user experience

$Study3_5$: Prior exposure to video games will not impact performance.

3.4 Results Study 3

The 22 participants in Study 3 saw three different types of scenarios, which included one change, two changes, or three changes, and the participants did not know how many changes were present at any given time. Each participant saw a total of five scenarios (two scenarios with one change, two scenarios with two changes, and one scenario with three changes) for a total of nine changes during their trial. In aggregate, our study evaluated human understanding over 110 scenarios with a total of 198 changes.

(a)



(b)

(c)

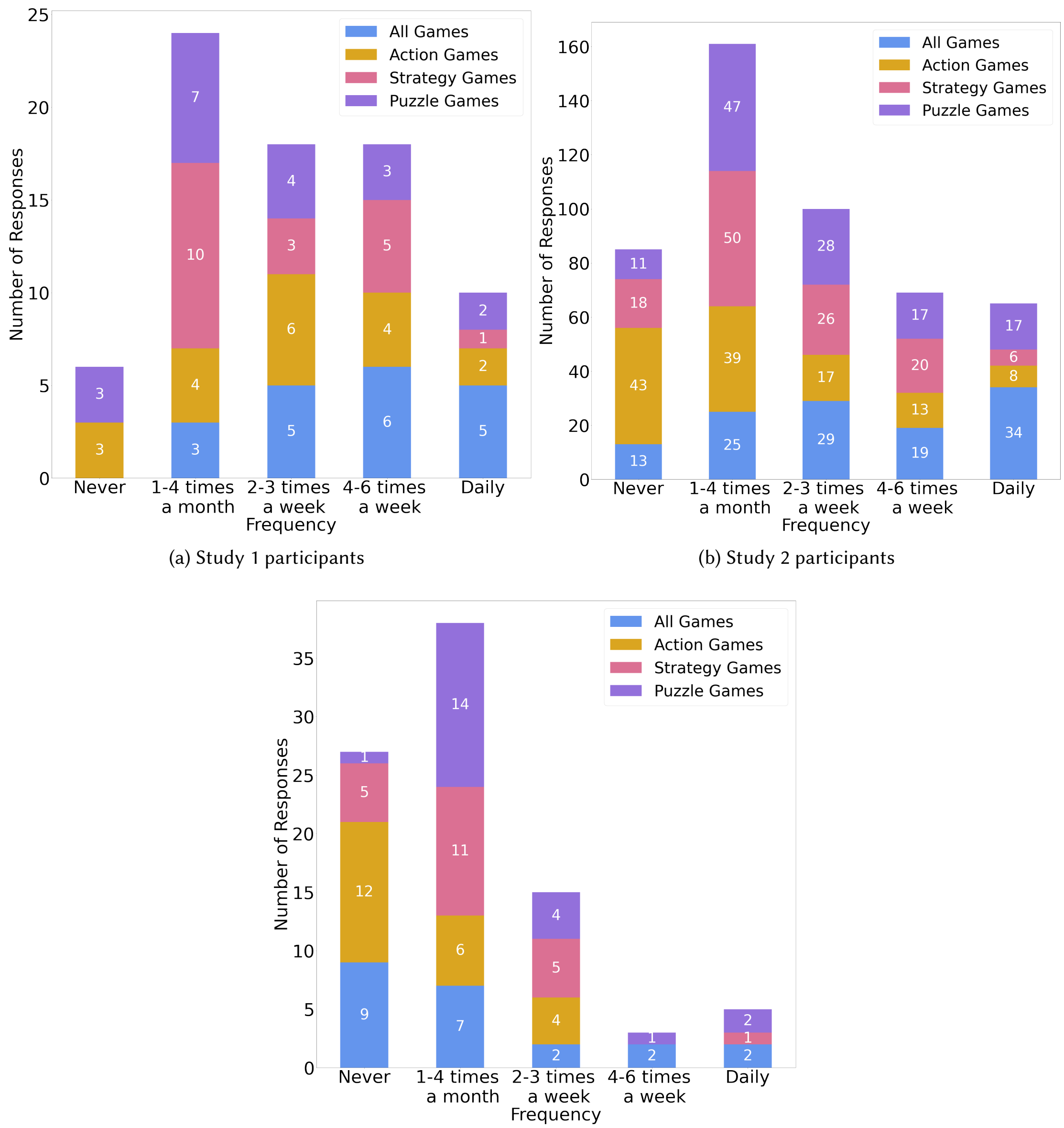
Fig. 11. In-person user study examples (Study 3 - Sec. 3.3). **(a)** panoramic view of the experimental environment; **(b)** user study experience; **(c)** change visualization shown via AR to participant.

For the purposes of analyzing system performance and human understandability, we define three important concepts present in this study. *Display error* corresponds to when our system erroneously shows a marker on a box to indicate a change that does not correspond to the ground truth, i.e., a system-introduced false positive. Display errors encompass both true false positives (no change occurred), as well as errors that occur due to imperfect frame alignment between the robot and AR-HMD as well as inevitable noise in the robot's SLAM subsystem. A display error is therefore an error that is introduced by the system and not the human. *Response error* corresponds to when a participant selects the incorrect box or misses one or several boxes that are correctly highlighted as a change by the system when asked what has changed, i.e., a participant-introduced false positive. A response error is therefore an error that is introduced by the human and not the system. Importantly, the participants did not know if they were experiencing any display error because they received no feedback with respect to the ground truth. Therefore, participants might be unaware of either their own error or display errors when they were asked to assess the usability of the system (see 4 for discussion). An additional type of error was when our system did not show a marker exactly on a single box, but rather off-center; in this case the participant experienced *ambiguity*. We note that this type of ambiguity error results from a combination of small system error and the environmental complexity. Because change visualizations are produced by the robot and overlaid on the participant's view in the AR-HMD, it is difficult to evaluate absolute error relative to the ground truth location of the boxes. Therefore, to determine if a participant experienced ambiguity error, we manually reviewed the first-person video recording from the AR-HMD across all tests and visually inspected the location of the change detection marker with respect to the box. We marked the presence of ambiguity error if the center of the change visualization appeared on the outer edge of the box, which would correspond to at least ~0.25m error relative to the ~0.5m box. Note, this error could be much larger, including the occasional ambiguity error corresponding to visualization markers shown in-between two different boxes. We measured the user experience along six questions, three investigating the ease of use (i.e. if the system enabled them to see the changes) and three investigating the difficulties (i.e. if the system was challenging to interpret). The answers were given for the overall study after the task, and not for individual scenarios. A rating scale from one to five shows how strongly participants agree. Results consistently showed a high ease of use ($M = 4.62, SD = 0.53$) and a low difficulty ($M = 1.3, SD = 0.52$). Out of the $N = 22$ participants, four felt strong enough to add additional comments to their experience mentioning ambiguity for some (bit never all) scenarios. Our post-experiment video check revealed that all four participants experienced ambiguity where the visualization was off-center.

Post-experimental analysis of the first-person AR-HMD video recordings showed that, of the 110 scenarios, 85 were shown precisely (i.e., no display error and no ambiguity), 12 were shown with display error only, 10 were shown with ambiguity only, and three were shown with both display error and ambiguity. Eight of the 10 scenarios with only ambiguity consisted of one ambiguous visualization while the remaining two scenarios exhibited two ambiguous visualizations. 11 participants completed the study without experiencing any ambiguity, nine participants experienced one scenario with ambiguous visualization, and two participants experienced two scenarios with ambiguous visualization.

3.5 Results Game Exposure

For each Study 1, 2, and 3 we asked participants to indicate how often they play video games, including generally playing and specifically playing action, strategy, and puzzle games. Options for measuring frequency of playtime ranged from never, rarely (measured as 1-4 times a month), sometimes (measured as 2-3 times a week), often (measured as 4-6 times a week) and daily. The purpose of this pre-study questionnaire was to gauge the amount of video game exposure, exposure to new technologies, and experience with egocentric vision the populations in our studies have as this might influence how they perceive spatial presence and act in the environment [26]. The reported video game frequency for participants for all three studies can be seen in Figure 12. In Study 1, the majority of participants reported rarely playing strategy or puzzle games, sometimes playing action games, and oftentimes playing general games and in Study 2 more people indicated they "never" play video games, however this study had a much lager number of participants ($N = 19$ in Study 1 vs. $N = 120$ in Study 2). Most of the participants in Study 1 and 2 had some regular exposure to games, and both studies were conducted online. Study 3 was an in-person study that recruited from the University of Denver's student population. We found that participants' experience was somewhat balanced between a low and no regular exposure to games. In our evaluation, we found no significant differences in video game exposure and performance the task across all three studies.



(a) Study 1 participants

(b) Study 2 participants

(c) Study 3 participants

Fig. 12. Video game exposure reported by participants.

4 DISCUSSION

In this section we provide analysis of the three studies and discuss trends with respect to our ten postulates that are summarized in Tab. 1, which also includes an indication of whether our findings support each postulate in the final column. First, we briefly summarize our key take-away observations. Regarding the impact of instruction, instruction was not strictly necessary for successful understanding of the visualization communications ($\mathcal{P}_1$); however, addition of instruction increased accuracy ($\mathcal{P}_4$). For the impact of the selection of display type (cluster aggregation or raw point visualizations, aggregation of data into cluster visualizations did not decrease accuracy ($\mathcal{P}_2$) and may indeed increase accuracy. Regarding the use of mirrored videos, original and mirrored videos yielded equivalent accuracy ($\mathcal{P}_3$). For the impact of a more immersive dynamic perspective in video, this appears to have increased accuracy ($\mathcal{P}_5$). Regarding the impact of the live, in-person experience, we were unable to achieve the 90% accuracy threshold (84.8% was achieved) ($\mathcal{P}_6$); but, we did see a decrease in error in the in-person experiment ($\mathcal{P}_7$). Additionally for the in-person experiment, ambiguity errors decreased accuracy as one might expect ($\mathcal{P}_8$), and fewer such errors appeared to contribute more positively towards user experience ($\mathcal{P}_9$). Finally, gaming exposure did not appear to have any impact on performance, which is positive for the usability of the system ($\mathcal{P}_{10}$). A full, detailed discussion follows.

With $\mathcal{P}_1$, we postulated that our system provided robot-based change detections in a way that was intuitive for humans such that they would not require instructions to understand what the robot was indicating. The participants in Study 1 collectively achieved an accuracy of 0.868 when identifying changes using the stationary videos, none of which included instructions on the system display. We further investigated this in Study 2 by intentionally not providing a subset of the participants with instructions and this group achieved an accuracy of 0.898. From the results of these two studies, we conclude that instructions are not strictly necessary for a layman to successfully use our system to effectively understand robot-based change detection visualizations.

Our second postulate, $\mathcal{P}_2$, explored the impact of display type (i.e. cluster vs. points vs. a combination of both) and posited that the cluster marker type would lead to a decrease in understanding a change, as measured by the participant's accuracy. The results of Study 1 and 2 do not support this postulate and reveal that participants actually performed the best when changes were visualized as cluster markers compared to discrete point data or both visualization types combined. More specifically, the participants' accuracy in Study 1 was 0.965 for changes indicated by cluster markers whereas accuracy decreased to 0.839 for points and to 0.802 both visualization types. In Study 2, change detection accuracy conveyed by clusters, points, and both was measured to be 0.934, 0.910, and 0.933, respectively . We therefore were able to reject $\mathcal{P}_2$ and this motivated using the cluster visualization for in-person experimentation in Study 3.

To ensure that our study fairly constructed supplemental videos to assess human understanding of change detection, we explored the impact of showing participant's mirrored versions of original videos. As postulated by $\mathcal{P}_3$, there was no significant difference in the participant's accuracy when viewing original or mirrored videos. In Study 1 the participants' accuracy was measured to be 0.863 and 0.859 in original and mirrored videos, respectively, and in Study 2 the difference was even smaller (0.925 and 0.924). From these results, we deduce that mirrored videos can be fairly included in our analysis as these videos did not accidentally introduce bias or misunderstanding that is not attributable to our system.

While our results from Study 1 and 2 suggest that participants can understand our robot-based change detection without requiring instructions, we postulated that accuracy would increase when participants are provided with instructions. We found that the participants in Study 2 that were explicitly given instructions regarding the system visualization reported an accuracy of 0.951, which is notably higher than the participants that received no instructions (0.898). The instructions explained some of the variations in the visualizations and provided one example with the goal to provide one training example and to reduce misinterpretations (Figs. 7 and 8). In fact, the participants from Study 1 provided comments in their exit survey to indicate that they assigned meaning to the visualizations in the absence of instruction. One participant stated, "Red shows slightly, or potential of change, a larger red and or small blue means probably, and a big blue means definitely" while a second participant thought, "the robot marked objects that it thought definitely changed with red circles...the purple circles indicated objects that might have potentially changed."

Unbeknownst to the participants that received no instructions, the meaning of the colors was not germane to the users' understanding (red were individual changes detected in the point cloud; blue were aggregated clusters of change points; there were no purple visualizations) and the robot was not communicating any notion of certainty nor were participants given any notion that this might be the case. This shows how much humans are trying to cognitively understand and make sense of the information given to them. Participants over-interpreted the meaning of the visualization and added meaning where these was none. While this tendency did not lead to a breakdown of understanding (e.g., misidentifying changes), a situation where humans are under high cognitive workload or time constraints, this additional cognitive effort could lead to errors. It is therefore advisable to be very cognizant of the visualization and how, if any, training is provided.

In the same vein of adding instructions to improve human understanding, we postulated with $\mathcal{P}_5$ that the participant's accuracy will increase when videos include dynamic viewing angles, compared to videos with stationary viewing angles. Study 1 only included videos with stationary viewpoints and the population's collective accuracy was 0.868. Accuracy improved to 0.921 when videos included a panning motion (Fig. 5), and improved further to .933 with a left-right translation (Fig. 6). We believe that the dynamic viewing angle provides additional spatial information and a more immersive experience. The the increase in change detection accuracy seems to support the notion that a dynamic view allows the human to more accurately understand where the visualization marker is in 3D space with respect to the physical environment.

Study 3 was the first of our studies to assess human understanding of robot-based change detection using a live and in-person system. As a result, we formulated four postulates to glean the performance of our human-robot system. Of all 198 total changes in Study 3, 168 were shown correctly with respect to the ground truth; this 0.848 accuracy does not support postulate $\mathcal{P}_6$, namely that $\geq 90\%$ would be correctly identified by the robot system. Of the remaining 30 changes, 14 changes were shown with display error, and 16 changes were shown with ambiguity. It is worth noting that 12 of the 14 changes that were shown with display error corresponded to the same change in the same scenario, meaning the error was consistent and not random. Because this error was consistent and can be therefore reproduced, we postulate that such systems errors have a high probability of being resolved in the next technical iterations of the robot system. This would indicate that the overall results of the human-robot system are a more conservative estimate of our system performance. Along these lines, 15 of the 16 changes shown with ambiguity were in close proximity of the correct box (i.e., up to $\sim 30cm$ off-center of the correct box or in between the correct box and an incorrect box). We believe this type of system limitation can likely be mitigated by engineering efforts to optimize frame alignment and reduce noise, resulting in decreased ambiguity and better overall human-robot system performance.

Study 1 and 2 consisted of video-based change detection evaluation where all changes were accurately shown to participants and the only source of display error was in the form of rare false positives that were not coincident with any changeable objects. Therefore, we compare these two error-free studies to the set of 85 scenarios shown in Study 3 that also did not include any display error or ambiguities. We see that there were 146 possible changes in these scenarios and only four response errors equating to an accuracy of 0.973. This error rate of 0.027 for the in-person Study 3 is strictly lower than the error rates for the stationary, pan, and left-right translation perspectives in the video-based Study 1 and 2 (0.132, 0.079, and 0.067, respectively). This provides support for our postulate $\mathcal{P}_7$ which states participant's error will decrease for in-person tests. We suspect that in-person evaluation, and likewise real-world deployment, is the most immersive experience because humans can choose how to move through out the environment. This can improve their understanding of where visualization markers are located in the virtual space and with which changed objects these markers are likely to be associated in the real space. What also might be playing a role is the shift from a 2D experience (i.e., video) showing changes in 3D space to a 3D experience (i.e., AR) showing changes in 3D space. It has been shown that adding depth and multiple viewpoints [75] improves performance further. Cognitively, this translation requires human effort and while we consider the initial accuracy rate excellent, we showed that the human-robot system can even perform better when being *in situ*.

An identified issue with our live system is the presence of visualization error in the form of ambiguities. With $\mathcal{P}_8$, we postulated that ambiguous visualizations (i.e., those not displayed directly over a single object) would reduce participants' understanding of the change information being communicated. 11 participants were unintentionally shown 16 instances of ambiguity in Study 3, of which eight participants correctly identified nine of the changes. This finding suggests that ambiguity does not reduce the correct identification of change detection altogether given that over 72% of participants that were shown ambiguity correctly identified the change; however, it does not provide strong support that participants can effectively mitigate the effect of ambiguous visualisations and discern the robot's intended change detection.

In addition to the participant's quantitative performance in the presence of ambiguous visualizations, we are also interested in their qualitative evaluation. We postulated with $\mathcal{P}_9$ that users presented with fewer ambiguities will enjoy a more favorable experience with our live system. The 11 participants that saw no ambiguity in any scenarios provided no negative comments while three participants of the 11 that saw any ambiguous visualizations described a worsened user experience. Two of these participants explicitly highlighted how ambiguous visualizations complicate reasoning by stating, "The orbs were sometimes not on the boxes and in between, which could make choosing the right box not always the most straight forward." and "Sometimes the spheres were between or to the side of the boxes which made it hard to tell which object actually had the sphere on them." Interestingly, one participant described their strategy for moving within the environment to deduce the intention of the robot-based change detection and overcome the ambiguous visualization. They stated, "The images were, at times, difficult to interpret from a particular angle/position within the workspace. If I moved around enough it became fairly clear what the images referred to. For example, in one case the image appeared in front of the box, and if there had been a box behind it I may have interpreted it as that box instead. Because the box was the last in the row and against a wall, it was the only box that made sense." This participant's explanation shows a deductive reasoning process and shows again that humans try to cognitively make sense of the information given to them. We speculate that this process could lead to a higher cognitive workload and could significantly decrease system performance if the user is pressed for time or otherwise under a high cognitive load.

We believe an AR-based change detection system designed to improve human-robot teaming, such as the one demonstrated in this work, has wide applicability in many domains. Given the potentially large and diverse set of operators, and to maximize utility and minimize overhead of deploying such a system, we assert that such a system should not require extensive training. This leads us to formulate postulate $\mathcal{P}_{10}$. We postulate prior exposure to video games will not impact the participant's ability to understand changes shown by the robot and correctly identify changes. The user responses from Study 1, 2, and 3 indicate that the minority of participants play action, strategy, puzzle, or general video games often (4-6 times a week) or daily; in fact, the majority of participants play these games rarely (1-4 times a month). Based on this self-reporting we believe that the populations in our studies are representative of the layman or non-expert. Furthermore, the lack of video game exposure suggests that prior knowledge with comparable virtual systems did not play a significant role in the participant's understanding of our system.

4.1 Limitations

Despite the significance of the findings of the studies in this system, we have identified several limitations that should be taken into consideration for understanding these results and designing similar future studies.

- Display errors and ambiguities should be mitigated further for the purposes of both analyzing human understanding of robot-based change detection and the realization of a fieldable solution. These inaccuracies occur because the system is constantly performing online alignment, SLAM, and change detection (Fig. 1). An error in any of the subsystems can create small displacements in positioning of visualizations. While this is infrequent (see $\mathcal{P}_8$ and $\mathcal{P}_9$ discussion) and the error is small, if the error exceeds the width of the objects subject to change, the visualization is placed ambiguously between objects, on the wrong object, or in empty space (see Sec. 3.4).
- In the course of the research presented, we had the opportunity to iteratively identify and address several limitations between studies. Study 1 presented videos AR visualizations from a static camera perspective that lacked an immersive quality. In Study 2, we introduced a dynamic perspective to replicate this immersive quality. Similarly, instructions were added in Study 2. However, both studies were conducted online due to the limitations imposed by COVID-19.
- Study 3 capitalized on these additions and adjustments in an in-person experiment, however this reduced or ability to recruit a larger number of participants.
- No feedback was given if the human's understanding of the change was correct in Study 3. This perhaps falsely inflates subjective user evaluations of the system, as the study participants did not experience *display errors*, only *ambiguity* (see Sec. 3.4). Feedback would allow for more insight for understanding human trust, communication quality, and overall system performance.
- While the current system does include bi-directional flow of information in the SLAM subsystems, e.g., for alignment as described in Fig. 1, as far as change information goes the current system is designed to study the effect of change information communicated to the human. The system does not allow for feedback from the human to the robot regarding the quality of the information or the decisions the human makes using that information. Future directions will incorporate bi-directional exchange regarding the human's interpretation of the change detections.
- This work focused on changes of addition only, as participants do not have prior knowledge of the environment over which they could reason about changes of deletion. In future work, we intend to explore changes of deletion, as well as changes of relocation with the incorporation of object tracking.
- While Studies 1 and 2 were crowdsourced online and paid, recruitment for Study 3 was limited to university students recruited as volunteers from the student population. Participants with a more diverse background or application specific experience would strengthen the generalization and insight of the findings.

5 CONCLUSION

In this paper, we present an end-to-end system for information transparency-enabled intelligent human-robot teaming. Our novel system couples a human-worn augmented reality headset with an intelligent, autonomous mobile robot that, as part of its perception capabilities includes both detecting changes in the environment and localizing its human teammate. This work additionally focuses on improving the robot's intelligence through identifying strategies for effective communication of this change information. We present extensive results from three explorative studies showing that using this system, the human teammate can understand information shared by via AR by the robot about changes in the environment in order to make appropriate decisions. We explore an array of design and experiential features with multiple studies yielding informative findings covering the ability to aggregate of change information without loss of human understanding, the positive impact of instruction on information interpretation, the effects of visualization error on user experience, and the importance of immersive visualization. We believe an AR-based change detection system designed to improve human-robot teaming, such as the one demonstrated in this work, has wide applicability in many domains. While this work focuses on communication of change detection information, we believe the findings contain more broad implications for improving robot decision making, e.g. with machine learning techniques, in any robot-to-human communication scenario where the robot has access to large volumes of perception data. Such improved decision making via machine learning methods is a target for future work.

REFERENCES

[1] Julie A Adams. 2002. Critical considerations for human-robot interface development. In *Proceedings of 2002 AAAI Fall Symposium*. 1–8.

[2] Arash Ajoudani, Andrea Maria Zanchettin, Serena Ivaldi, Alin Albu-Schäffer, Kazuhiro Kosuge, and Oussama Khatib. 2018. Progress and prospects of the human–robot collaboration. *Autonomous Robots* 42, 5 (2018), 957–975.

[3] Angelos Angelopoulos, Austin Hale, Husam Shaik, Akshay Paruchuri, Ken Liu, Randal Tuggle, and Daniel Szafir. 2022. Drone Brush: Mixed Reality Drone Path Planning. In *Proceedings of the 2022 ACM/IEEE International Conference on Human-Robot Interaction*. 678–682.

[4] Deborah R Billings, Kristin E Schaefer, Jessie YC Chen, and Peter A Hancock. 2012. Human-robot interaction: developing trust in robots. In *Proceedings of the seventh annual ACM/IEEE international conference on Human-Robot Interaction*. 109–110.

[5] David Borland and Russell M Taylor Ii. 2007. Rainbow color map (still) considered harmful. *IEEE computer graphics and applications* 27, 2 (2007), 14–17.

[6] Cynthia Breazeal, Cory D Kidd, Andrea Lockerd Thomaz, Guy Hoffman, and Matt Berlin. 2005. Effects of nonverbal communication on efficiency and robustness in human-robot teamwork. In *2005 IEEE/RSJ international conference on intelligent robots and systems*. IEEE, 708–713.

[7] Stuart K Card and Jock Mackinlay. 1997. The structure of the information visualization design space. In *Proceedings of VIZ'97: Visualization Conference, Information Visualization Symposium and Parallel Rendering Symposium*. IEEE, 92–99.

[8] Sonia Mary Chacko and Vikram Kapila. 2019. An augmented reality interface for human-robot interaction in unconstrained environments. In *2019 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 3222–3228.

[9] Kishan Chandan, Vidisha Kudalkar, Xiang Li, and Shiqi Zhang. 2021. ARROCH: Augmented reality for robots collaborating with a human. In *2021 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 3787–3793.

[10] Jessie YC Chen, Ellen C Haas, and Michael J Barnes. 2007. Human performance issues and user interface design for teleoperated robots. *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)* 37, 6 (2007), 1231–1245.

[11] Frank Dellaert. 2012. *Factor graphs and GTSAM: A hands-on introduction*. Technical Report. Georgia Institute of Technology.

[12] Jeffrey Delmerico, Stefano Mintchev, Alessandro Giusti, Boris Gromov, Kamilo Melo, Tomislav Horvat, Cesar Cadena, Marco Hutter, Auke Ijspeert, Dario Floreano, et al. 2019. The current state and future outlook of rescue robotics. *Journal of Field Robotics* 36, 7 (2019), 1171–1191.

[13] Mustafa Demir, Nathan J McNeese, and Nancy J Cooke. 2020. Understanding human-robot teams in light of all-human teams: Aspects of team interaction and shared cognition. *International Journal of Human-Computer Studies* 140 (2020), 102436.

[14] Mark Dennison, Christopher Reardon, Jason Gregory, Theron Trout, and John G Rogers III. 2020. Creating a mixed reality common operating picture across C2 echelons for human-autonomy teams. In *Virtual, Augmented, and Mixed Reality (XR) Technology for Multi-Domain Operations*, Vol. 11426. International Society for Optics and Photonics, 114260M.

[15] HC Fang, Soh-Khim Ong, and Andrew YC Nee. 2014. Novel AR-based interface for human-robot interaction and visualization. *Advances in Manufacturing* 2 (2014), 275–288.

[16] Samer Faraj and Lee Sproull. 2000. Coordinating expertise in software development teams. *Management science* 46, 12 (2000), 1554–1568.

[17] Terrence Fong and Illah Nourbakhsh. 2005. Interaction challenges in human-robot space exploration. *Interactions* 12, 2 (2005), 42–45.

[18] Scott A Green, Mark Billinghurst, XiaoQi Chen, and J Geoffrey Chase. 2007. Human robot collaboration: An augmented reality approach—a literature review and analysis. In *International Design Engineering Technical Conferences and Computers and Information in Engineering Conference*, Vol. 48051. 117–126.

[19] Scott A Green, Mark Billinghurst, XiaoQi Chen, and J Geoffrey Chase. 2008. Human-robot collaboration: A literature review and augmented reality approach in design. *International journal of advanced robotic systems* 5, 1 (2008), 1.

[20] Jason Gregory, Jonathan Fink, Ethan Stump, Jeffrey Twigg, John Rogers, David Baran, Nicholas Fung, and Stuart Young. 2016. Application of multi-robot systems to disaster-relief scenarios with limited communication. In *Field and Service Robotics*. Springer, 639–653.

[21] Jason M Gregory, Christopher Reardon, Kevin Lee, Geoffrey White, Ki Ng, and Caitlyn Sims. 2019. Enabling intuitive human-robot teaming using augmented reality and gesture control. *arXiv preprint arXiv:1909.06415* (2019).

[22] Victoria Groom and Clifford Nass. 2007. Can robots be teammates?: Benchmarks in human–robot teams. *Interaction studies* 8, 3 (2007), 483–500.

[23] Kerstin S Haring, Victor Finomore, Dylan Muramoto, Nathan L Tenhundfeld, J Wen, and B Tidball. 2018. Analysis of using virtual reality (vr) for command and control applications of multi-robot systems. In *Proceedings of the 1st International Workshop on Virtual, Augmented, and Mixed Reality for HRI (VAM-HRI)*.

[24] Kerstin S Haring, Elizabeth Phillips, Elizabeth H Lazzara, Daniel Ullman, Anthony L Baker, and Joseph R Keebler. 2021. Applying the swift trust model to human-robot teaming. In *Trust in Human-Robot Interaction*. Elsevier, 407–427.

[25] Kerstin S Haring, Katsumi Watanabe, Mari Velonaki, Chad C Tossell, and Victor Finomore. 2018. FFAB—The form function attribution bias in human–robot interaction. *IEEE Transactions on Cognitive and Developmental Systems* 10, 4 (2018), 843–851.

[26] Michael Havranek, Nicolas Langer, Marcus Cheetham, and Lutz Jäncke. 2012. Perspective and agency during video gaming influences spatial presence experience and brain activation patterns. *Behavioral and brain functions* 8, 1 (2012), 1–13.

[27] Bradley Hayes and Brian Scassellati. 2013. Challenges in shared-environment human-robot collaboration. *learning* 8, 9 (2013).

[28] Guy Hoffman and Cynthia Breazeal. 2004. Collaboration in human-robot teams. In *AIAA 1st intelligent systems technical conference*. 6434.

[29] Astrid C Homan, John R Hollenbeck, Stephen E Humphrey, Daan Van Knippenberg, Daniel R Ilgen, and Gerben A Van Kleef. 2008. Facing differences with an open mind: Openness to experience, salience of intragroup differences, and performance of diverse work groups. *Academy of Management Journal* 51, 6 (2008), 1204–1222.

[30] Gareth R Jones and Jennifer M George. 1998. The experience and evolution of trust: Implications for cooperation and teamwork. *Academy of management review* 23, 3 (1998), 531–546.

[31] M Waleed Kadous, Raymond Ka-Man Sheh, and Claude Sammut. 2006. Effective user interface design for rescue robotics. In *Proceedings of the 1st ACM SIGCHI/SIGART conference on Human-robot interaction*. 250–257.

[32] Tobias Kaupp, Alexei Makarenko, and Hugh Durrant-Whyte. 2010. Human–robot communication for collaborative decision making—A probabilistic approach. *Robotics and Autonomous Systems* 58, 5 (2010), 444–456.

[33] Bing Cai Kok and Harold Soh. 2020. Trust in robots: Challenges and opportunities. *Current Robotics Reports* 1, 4 (2020), 297–309.

[34] Shan G Lakhmani, Julia L Wright, Michael R Schwartz, and Daniel Barber. 2019. Exploring the effect of communication patterns and transparency on performance in a human-robot team. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, Vol. 63. SAGE Publications Sage CA: Los Angeles, CA, 160–164.

[35] Shannon L Marlow, Christina N Lacerenza, Jensine Paoletti, C Shawn Burke, and Eduardo Salas. 2018. Does team communication represent a one-size-fits-all approach?: A meta-analysis of team communication and performance. *Organizational behavior and human decision processes* 144 (2018), 145–170.

[36] Stephen Marsland, Ulrich Nehmzow, and Jonathan Shapiro. 2005. On-line novelty detection for autonomous mobile robots. *Robotics and Autonomous Systems* 51, 2-3 (2005), 191–206.

[37] Eloise Matheson, Riccardo Minto, Emanuele GG Zampieri, Maurizio Faccio, and Giulio Rosati. 2019. Human–robot collaboration in manufacturing applications: A review. *Robotics* 8, 4 (2019), 100.

[38] Terran Mott, Thomas Williams, Hao Zhang, and Christopher Reardon. 2021. You have time to explore over here!: Augmented reality for enhanced situation awareness in human-robot collaborative exploration. In *4th International Workshop on Virtual, Augmented, and Mixed Reality for HRI*.

[39] Robin R Murphy. 2014. *Disaster robotics*. MIT press.

[40] Chris North. 2006. Toward measuring visualization insight. *IEEE computer graphics and applications* 26, 3 (2006), 6–9.

[41] P Núñez, P Drews, R Rocha, M Campos, and Jorge Dias. 2009. Novelty detection and 3D shape retrieval based on Gaussian mixture models for autonomous surveillance robotics. In *2009 IEEE/RSJ International Conference on Intelligent Robots and Systems*. IEEE, 4724–4730.

[42] Savannah Paul, Christopher Reardon, Tom Williams, and Hao Zhang. 2020. Designing augmented reality visualizations for synchronized and time-dominant human-robot teaming. In *Virtual, Augmented, and Mixed Reality (XR) Technology for Multi-Domain Operations*, Vol. 11426. International Society for Optics and Photonics, 1142607.

[43] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay. 2011. Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research* 12 (2011), 2825–2830.

[44] Elizabeth Kathleen Phillips and Florian G. Jentsch. 2017. Supporting Situation Awareness through Robot-to-Human Information Exchanges under Conditions of Visuospatial Perspective Taking. *J. Hum.-Robot Interact.* 6, 3 (dec 2017), 92–117. https://doi.org/10.5898/JHRI.6.3.Phillips

[45] Marco AF Pimentel, David A Clifton, Lei Clifton, and Lionel Tarassenko. 2014. A review of novelty detection. *Signal Processing* 99 (2014), 215–249.

[46] Shuwen Qiu, Hangxin Liu, Zeyu Zhang, Yixin Zhu, and Song-Chun Zhu. 2020. Human-robot interaction in a shared augmented reality workspace. In *2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 11413–11418.

[47] Morgan Quigley, Josh Faust, Tully Foote, and Jeremy Leibs. 2009. ROS: an open-source Robot Operating System. In *International Conference on Robotics and Automation Workshop on Open Source Software*.

[48] Christopher Reardon, Jason Gregory, Carlos Nieto-Granda, and John G Rogers. 2020. Enabling Situational Awareness via Augmented Reality of Autonomous Robot-Based Environmental Change Detection. In *International Conference on Human-Computer Interaction: Virtual, Augmented, and Mixed Reality*. Springer, 611–628.

[49] Christopher Reardon, Jason Gregory, Carlos Nieto-Granda, and John G Rogers III. 2021. Designing a mixed reality interface for autonomous robot-based change detection. In *Virtual, Augmented, and Mixed Reality (XR) Technology for Multi-Domain Operations II*, Vol. 11759. International Society for Optics and Photonics, 117590J.

[50] Christopher Reardon, Kerstin Haring, Jason M Gregory, and John G Rogers. 2021. Evaluating Human Understanding of a Mixed Reality Interface for Autonomous Robot-Based Change Detection. In *2021 IEEE International Symposium on Safety, Security, and Rescue Robotics (SSRR)*. IEEE, 132–137.

[51] Christopher Reardon, Kevin Lee, and Jonathan Fink. 2018. Come See This! Augmented Reality to Enable Human-Robot Cooperative Search. In *Proceedings of the 2018 IEEE Symposium on Safety, Security, and Rescue Robotics*.

[52] Christopher Reardon, Kevin Lee, John G Rogers, and Jonathan Fink. 2019. Augmented Reality for Human-Robot Teaming in Field Environments. In *International Conference on Human-Computer Interaction: Virtual, Augmented, and Mixed Reality*. Springer, 79–92.

[53] Christopher Reardon, Kevin Lee, John G Rogers, and Jonathan Fink. 2019. Communicating via augmented reality for human-robot teaming in field environments. In *2019 IEEE International Symposium on Safety, Security, and Rescue Robotics (SSRR)*. IEEE, 94–101.

[54] Bernice E Rogowitz and Alan D Kalvin. 2001. The" which blair project": A quick visual method for evaluating perceptual color maps. In *Proceedings Visualization, 2001. VIS'01*. IEEE, 183–556.

[55] Juan Jesús Roldán, Elena Peña-Tapia, David Garzón-Ramos, Jorge de León, Mario Garzón, Jaime del Cerro, and Antonio Barrientos. 2019. Multi-robot systems, virtual reality and ROS: developing a new generation of operator interfaces. In *Robot operating system (ROS)*. Springer, 29–64.

[56] Eric Rosen, David Whitney, Elizabeth Phillips, Gary Chien, James Tompkin, George Konidaris, and Stefanie Tellex. 2020. Communicating robot arm motion intent through mixed reality head-mounted displays. In *Robotics research*. Springer, 301–316.

[57] Alessandra Rossi, Kerstin Dautenhahn, Kheng Lee Koay, and Joe Saunders. 2017. Investigating human perceptions of trust in robots for safe HRI in home environments. In *Proceedings of the Companion of the 2017 ACM/IEEE International Conference on Human-Robot Interaction*. 375–376.

[58] Radu Bogdan Rusu and Steve Cousins. 2011. 3D is here: Point cloud library PCL. In *2011 IEEE International Conference on Robotics and Automation*. IEEE, 1–4.

[59] Aleksandr Segal, Dirk Haehnel, and Sebastian Thrun. 2009. Generalized-ICP. In *Robotics: Science and Systems*, Vol. 2. 435.

[60] Ben Shneiderman. 1996. The eyes have it: A task by data type taxonomy for information visualizations. In *Proceedings 1996 IEEE symposium on visual languages*. IEEE, 336–343.

[61] Daniel J Simons. 2000. Current approaches to change blindness. *Visual cognition* 7, 1-3 (2000), 1–15.

[62] Daniel J Simons and Daniel T Levin. 1997. Change blindness. *Trends in cognitive sciences* 1, 7 (1997), 261–267.

[63] Boris Sofman, Bradford Neuman, Anthony Stentz, and J Andrew Bagnell. 2011. Anytime online novelty and change detection for mobile robots. *Journal of Field Robotics* 28, 4 (2011), 589–618.

[64] Mirco Sturari, Marina Paolanti, Emanuele Frontoni, Adriano Mancini, and Primo Zingaretti. 2017. Robotic platform for deep change detection for rail safety and security. In *2017 European Conference on Mobile Robots (ECMR)*. IEEE, 1–6.

[65] Daniel Szafir. 2019. Mediating Human-Robot Interactions with Virtual, Augmented, and Mixed Reality. In *International Conference on Human-Computer Interaction*. Springer, 124–149.

[66] Daniel Szafir, Bilge Mutlu, and Terrence Fong. 2017. Designing planning and control interfaces to support user collaboration with flying robots. *The International Journal of Robotics Research* 36, 5-7 (2017), 514–542.

[67] Daniel Szafir and Danielle Albers Szafir. 2021. Connecting human-robot interaction and data visualization. In *Proceedings of the 2021 ACM/IEEE International Conference on Human-Robot Interaction*. 281–292.

[68] Aaquib Tabrez, Matthew B Luebbers, and Bradley Hayes. 2022. Descriptive and prescriptive visual guidance to improve shared situational awareness in human-robot teaming. In *Proceedings of the 21st International Conference on Autonomous Agents and Multiagent Systems*. 1256–1264.

[69] Nhan Tran, Trevor Grant, Thao Phung, Leanne Hirshfield, Christopher Wickens, and Tom Williams. 2021. Get This!? Mixed Reality Improves Robot Communication Regardless of Mental Workload. In *Companion of the 2021 ACM/IEEE International Conference on Human-Robot Interaction*. 412–416.

[70] A. J. B. Trevor, J. G. Rogers, and H. I. Christensen. 2014. OmniMapper: A modular multimodal mapping framework. In *2014 IEEE International Conference on Robotics and Automation (ICRA)*. 1983–1990. https://doi.org/10.1109/ICRA.2014.6907122

[71] Antonio W Vieira, Paulo LJ Drews, and Mario FM Campos. 2014. Spatial density patterns for efficient change detection in 3D environment for autonomous surveillance robots. *IEEE Transactions on Automation Science and Engineering* 11, 3 (2014), 766–774.

[72] Michael Walker, Zhaozhong Chen, Matthew Whitlock, David Blair, Danielle Albers Szafir, Christoffer Heckman, and Daniel Szafir. 2021. A mixed reality supervision and telepresence interface for outdoor field robotics. In *2021 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2345–2352.

[73] Michael Walker, Hooman Hedayati, Jennifer Lee, and Daniel Szafir. 2018. Communicating robot motion intent with augmented reality. In *Proceedings of the 2018 ACM/IEEE International Conference on Human-Robot Interaction*. 316–324.

[74] Xi Vincent Wang and Lihui Wang. 2021. Augmented reality enabled human–robot collaboration. In *Advanced Human-Robot Collaboration in Manufacturing*. Springer, 395–411.

[75] Christopher D Wickens. 2000. The when and how of using 2-D and 3-D displays for operational tasks. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, Vol. 44. SAGE Publications Sage CA: Los Angeles, CA, 3–403.

[76] A William Evans, Matthew Marge, Ethan Stump, Garrett Warnell, Joseph Conroy, Douglas Summers-Stay, and David Baran. 2017. The future of human robot teams in the army: Factors affecting a model of human-system dialogue towards greater team collaboration. In *Advances in human factors in robots and unmanned systems*. Springer, 197–209.

[77] Tom Williams, Priscilla Briggs, and Matthias Scheutz. 2015. Covert robot-robot communication: Human perceptions and implications for human-robot interaction. *Journal of Human-Robot Interaction* 4, 2 (2015), 24–49.

[78] Thomas Emrys Williams, Rehj Cantrell, Gordon Briggs, Paul Schermerhorn, and Matthias Scheutz. 2013. Grounding natural language references to unvisited and hypothetical locations. In *Twenty-Seventh AAAI Conference on Artificial Intelligence*.

[79] Franziska Doris Wolf and Ruth Stock-Homburg. 2020. Human-Robot Teams: A Review. In *International Conference on Social Robotics*. Springer, 246–258.

[80] Holly A Yanco, Brenden Keyes, Jill L Drury, Curtis W Nielsen, Douglas A Few, and David J Bruemmer. 2007. Evolving interface design for robot search tasks. *Journal of Field Robotics* 24, 8-9 (2007), 779–799.

[81] Sierra N Young and Joshua M Peschel. 2020. Review of human–machine interfaces for small unmanned systems with robotic manipulators. *IEEE Transactions on Human-Machine Systems* 50, 2 (2020), 131–143.