A Robotic Augmented Reality Virtual Window for Law Enforcement Operations

Nate Phillips, Brady Kruse, Farzana Alam Khan, J. Edward Swan II, and Cindy L. Bethel

Mississippi State University, Mississippi State, MS 39762, USA
Nathaniel.C.Phillips@ieee.org, brady.kruse@ieee.org, fk141@msstate.edu, swan@acm.org, cbethel@cse.msstate.edu

Abstract. In room clearing tasks, SWAT team members suffer from a lack of initial environmental information: knowledge about what is in a room and what relevance or threat level it represents for mission parameters. Normally this gap in situation awareness is rectified only upon room entry, forcing SWAT team members to rely on quick responses and near-instinctual reactions. This can lead to dangerously escalating situations or important missed information which, in turn, can increase the likelihood of injury and even mortality. Thus, we present an *x-ray vision* system for the dynamic scanning and display of room content, using a robotic platform to mitigate operator risk. This system maps a room using a robot-equipped stereo depth camera and, using an augmented reality (AR) system, presents the resulting geographic information according to the perspective of each officer. This intervention has the potential to notably lower risk and increase officer situation awareness, all while team members are in the relative safety of cover. With these potential stakes, it is important to test the viability of this system natively and in an operational SWAT team context.

Keywords: Augmented Reality; Situation Awareness; X-Ray Vision; Robotics

1 Introduction and Motivation

A major concern in our country and around the world is the challenges law enforcement officers face during incident responses and the attention they receive as a result of the actions and decisions made during these responses [3]. The focus of this paper is to describe the design and development process for an augmented reality virtual window to provide *x-ray vision*, through which otherwise occluded objects can be seen in unknown, potentially dangerous, and stressful environments in the context of a common SWAT (special weapons and tactics) team task.

As an example, a SWAT team is standing outside of an apartment during a high stress, slow, and methodical building search operation as part of serving a drug search warrant. They have a non-weaponized tactical robot, and they send it into the apartment to gain situation awareness and reconnaissance information prior to entry. The SWAT team breacher is able to open any door to provide access for the robot. The team members wear augmented reality (AR) head-mounted displays integrated with their helmets [47]. Visual images are provided to SWAT team members from both a through-the-camera view, seen from the robot's perspective, and an AR virtual window view, seen from the point of view of each team member wearing an AR head-mounted display. Providing video information to SWAT officers by sending a robot into the environment

Phillips N., Kruse B., Khan F.A., Swan II J.E., Bethel C.L. (2020) A Robotic Augmented Reality Virtual Window for Law Enforcement Operations. In: Chen J., Fragomeni G. (eds) *Virtual, Augmented and Mixed Reality. Design and Interaction. HCII 2020. Lecture Notes in Computer Science, vol 12190.* Springer, Cham, pp. 591–610. https://doi.org/10.1007/978-3-030-49695-1_40

This is the authors' version of the article. The final authenticated version is available online at https://doi.org/10.1007/978-3-030-49695-1_40.

prior to entry improves planning and decision making by the officers. With more information about the situation, better decisions can be made and more lives can be protected, both officers and public. It is expected that overall team performance, situation awareness, and safety will be enhanced through this *x-ray vision* approach.

2 Background

In order to accomplish this approach, we use *augmented reality* (AR). AR, sometimes also known as *mixed reality*, supplements our natural senses by overlaying additional information on top of what is natively sensed [12]. However, this approach does have some limitations and downsides; human attention is limited, and additional informational inputs can quickly surpass that limit [15, 16]. Further, cognitive processing for certain types of information, particularly as it relates to depth, may interact negatively with what can be seen in the real world [33]. These problems lead to two main issues for *X-Ray Vision*: perceptual problems associated with AR and situation awareness concerns inherent to human psychology. Even beyond these factors, human-robot integration also represents a significant challenge for this sort of task.

2.1 Perception and Augmented Reality

Successful AR x-ray vision, which has been a significant area of interest [1, 17, 25, 32, 33], is perhaps the most important component of this task. Accurate and accessible augmentation of the environment is a necessity in order to improve outcomes for SWAT teams. As such, this task requires the creation of 3D images that appear to occupy specific locations concurrent with the measured position of those objects in the real world. To do this, we make use of *depth cues*, environmental variables used by the human perceptual system to generate a sense of depth. These include occlusion, stereopsis, accommodation, contrast, motion parallax, familiar size, and many more [11, 12]. Significant research has been done to determine the relative effectiveness of each of these cues and how they affect augmented reality depth perception [11, 12, 22, 23, 31, 40, 41, 44]; however, a comprehensive approach for accurate depth perception within AR has been an elusive goal [12, 22, 31].

This is further complicated by the limited nature of most AR displays. While displays have improved dramatically over the last decade, they are still fundamentally limited. Few or no commercially available AR displays allow the adjustment of each depth perception parameter [40]. Thus, parameters like multi-focal accommodation, which are relatively difficult to design into mobile display technology, are often unavailable [27, 30]. Other parameters, such as brightness, can generally be adjusted, but current hardware limits fall short of the brightness of a typical sunny day [48]. Fortunately, this particular application is able to make use of several important depth perception cues, including stereopsis and motion parallax.

Stereopsis, or the combination of stereo images in human eyes to create a depth effect, has been found to have a strong effect on depth perception within arm's reach [11, 21, 23]. Beyond that, however, the effectiveness of this depth cue deteriorates with distance [11, 21]. On its own, stereopsis would likely represent a potentially fallible depth cue for this application [37, 42]. Fortunately, it is also paired with motion parallax, which has been found in several studies to notably increase depth estimation accuracy [8, 9, 42]. Even in real world studies, motion parallax is an important depth cue; its availability improves depth estimation and undermines several types of optical depth illusions [8, 9]. Other depth cues, such as familiar size, contrast, or texture, may have an impact on depth perception [35, 36], but those effects, with one exception, are notably less significant than the effects of stereopsis and motion parallax.

In a vacuum, we would expect stereopsis and motion parallax to be generally sufficient for locating objects in depth [37, 42]. However, there is a complication: *occlusion*, in which an object in front blocks the view of an object behind, is one of the most salient visual cues for human vision [11, 24]. In this task, where occlusion is necessarily ignored, there is a significant perceptual conflict between seeing an opaque wall and perceiving virtual 3D data beyond it. In some cases, this can cause the depth of the virtual data to be perturbed [5, 33]. In order to deprecate this perceptual conflict, we make use of the *window metaphor*. The window metaphor is a way of contextualizing the display of data past an apparently solid object [5]. Generally, humans have no experience with or grasp of seeing a collection of images through an opaque object. However, we do understand the context of a window, a bounded box that lets us see beyond some barrier. Where a boundaryless mass of floating stimuli beyond a wall might be confusing and bewildering, a neat, bounded window through a wall could be expected to be more perceptually understandable and approachable. Indeed, research on this window metaphor or similar concepts has generally found improvement in depth perception estimations, though further research should be done to confirm this phenomenon [5, 33]. Thus, we expect that the window metaphor will increase user understanding of depth and help prevent information overload.

2.2 SWAT Team Integration and Deployment

Even with the carefully engineered approach described above, this system would be ineffective if it could not be integrated with the target audience, a SWAT team. The complex integration of robotics and augmented reality has the potential to be either a successful extension of SWAT team awareness or a dangerous distraction that yields little operational value. While we do not yet have data on this use case, extrapolation from previous research can help us determine which characteristics are likely to lead to positive outcomes and which are not.

First, it is important to note that tactical robot integration is a challenging task and is difficult to do well [50]. It is easy to create system features that almost inherently detract from team goals: embedded operator requirements, robots that interfere with team performance, or distracting control schemes [2, 29]. However, positive responses have been garnered from some system features [2, 29]. Broadly speaking, these seem to point toward a desire for greater robot autonomy, robot features that draw fire away from (and not toward) team members, and robot features that allow potential offloading of team tasks onto the robot [2]. While these characteristics don't necessarily map 1-to-1 with this proposed system, there are still many general lessons to learn from these results.

The clear first lesson is that the proposed system needs to be easy to use. Team members are unlikely to want to spend valuable time trying to troubleshoot or set up a system that is buggy or difficult to use. Such a distraction would detract from their core task and increase the likelihood of negative outcomes. SWAT team members are also unlikely to want to dedicate a full-time operator to managing the robot [2]. Further, a robot with greater independence would also be better able to draw fire away from team members and help the team more efficiently clear rooms. As such, while it is beyond the purview of this project, developing limited or supervised autonomy that allows a robot to move between rooms, following police search protocols, would likely be a useful addition to this set of tools.

Finally, there are a few attributes of this system that have not been previously investigated and which may be difficult to integrate into tactical teams. A primary concern would be how officers will react to wearing mobile AR displays. Clearly, in the long term and for live use cases, wearing anything that interferes with body armor or head protection would be met with an almost immediate rejection. However, if the technology otherwise proves worthwhile and usable, it is possible that hardened versions of AR displays could be produced and integrated into body armor. Another potential factor would be the degree to which the AR display becomes a distraction when officers are in the process of clearing a room. This, however, seems like it could be easily counteracted with a minimal display footprint and/or with a manual AR-on/AR-off switch. It is expected that our evaluation of this system will determine how to best support tactical team performance and usability.

2.3 Situation Awareness

The crux of the issue, however, is *situation awareness*: the general understanding and grasp of a given task environment. This is the very quality that is expected to improve with this system [14, 15, 43]. As previous research has discovered, it is entirely possible to reduce situation awareness with some types of seemingly positive interventions. In particular, it is surprisingly easy to overwhelm or distort human cognition with too much information, particularly extraneous or difficult-to-incorporate information [14, 15]. Thus, in the design and testing of our system, we should consider whether it truly improves situation awareness or whether it, in fact, reduces it.

First, it is important to more fully examine the underlying concept of situation awareness (See Figure 1). Situation awareness informs human decision making by encompassing underlying perceived data; in a very real sense, situation awareness can be interpreted as our own internal model of our environment. For 'good' situation awareness, the internal model of the environment would be both accurate and focused on the features that actually matter; for 'poor' situation awareness, the internal model might be inaccurate or not useful for the situation. As such, situation awareness is composed of three main components: perception, comprehension, and prediction [15, 18]. In order to build an accurate model, information must be perceived and its implications understood. For complex scenarios, better understanding of the underlying information leads to improved predictions about the state of the environment.

However, it is worth noting that better situation awareness does not inherently result in better task performance [15, 16]. It is entirely possible for an operator to have excellent situation awareness and make a mistake or for an operator to have poor situation awareness and blunder into a better performance. It is even possible for an operator to be very confident in his or her situation awareness, but actually have a poor understanding of the situation [15, 16]. As such, while task performance or user self-judgment and situation awareness are likely highly correlated, these measures are not themselves ideal tests of situation awareness.

What, then, does represent an exemplary test for situation awareness? Several measurement methods have been drawn from multiple disciplines, including air traffic control, military command, and robotic operations [16, 18, 50]. There are three main methodical approaches: *explicit*

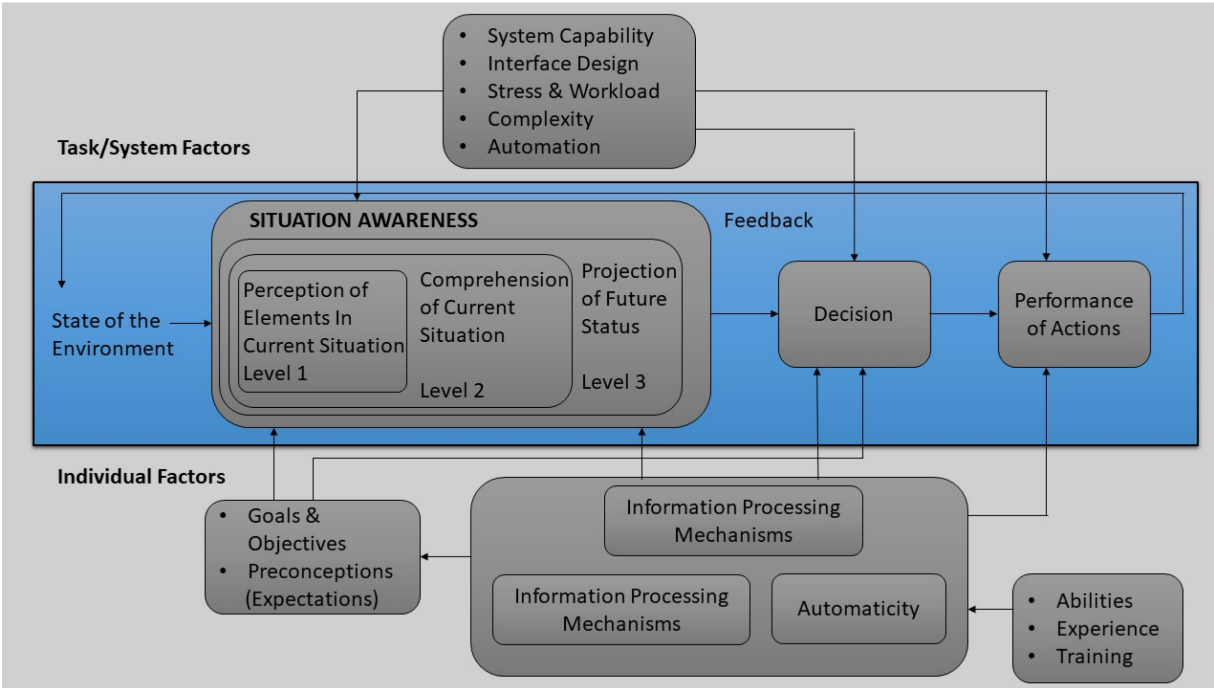

Fig. 1: Situation awareness model for dynamic decision making [15].

tests, *implicit* tests, and subjective tests. We mainly focus on explicit and implicit methodology, since explicit tests are both direct and objective and implicit tests are directly connected to task performance, and, as such, are generally considered sound methods [18]. SAGAT (Situation Awareness Global Assessment Technique), ASAGAT (Analog SAGAT), and QASAGAT (Quantitative Analog SAGAT) are three very similar approaches that use mid-intervention 'freezes' to assess a user's knowledge in the moment [16, 18]. CARS (Crew Awareness Rating Scale), PASA (Post Assessment of Situation Awareness), and SPASA (Short Post Assessment of Situation Awareness) are examples of retrospective techniques, which measure a user's understanding of the task after completion, usually through a short survey [16, 18]. Both of these approaches have their own shortcomings. SAGAT-based methods require frequent and intrusive task freezes that may affect the actual cognitive task approach or may bias the operator toward focusing specifically on what is being tested [16, 18]. The retrospective techniques, in contrast, occur post-task and so do not interfere with task cognition; however, they produce results that may not be indicative of an operator's effective situation awareness within the task [16, 18]. All of these approaches focus on the functional application of the task and individual parameters relating to task performance. However, the primary aspect of situation awareness that this system improves is perception. As such, one of the key questions we must ask in our research is: *How can we measure the effect of this system on an operator's spatial perception?*

3 Approach

The design approach for this project can be viewed as occurring in three distinct steps: (1) determining ideal system parameters from research and task stakeholders, (2) implementing the x-ray vision system, and (3) gauging the system's effectiveness. The following section contains explanations and information pertinent to the current methods used by SWAT teams, the technical execution of our project thus far, and potential future experiments meant to measure memory retention, general performance, and perceptual accuracy.

3.1 Tactical Room Clearing

SWAT police teams are called in to safely and effectively resolve certain situations where the threat of violence is considered high: counterterrorism operations, drug raids, high-risk search warrants, hostage situations, or apprehension and arrest of armed suspects. These situations are extremely stressful and can easily lead to unnecessary loss of life, as in the several recorded instances of swatting deaths. In particular, room clearing operations require team members to enter a novel environment/situation and evaluate it for potential threats. There are currently two methods for entering a room—blind entry and mirror entry. In blind entry, team members typically go into a room with no visual knowledge of what is on the other side. In mirror entry (See Figure 2), a team member uses a mirror on a pole to examine the contents of the room from cover before entering. Both of these approaches have notable problems; both leave team members with an imperfect awareness of the environment. In mirror entry, even though some environment information is imparted, this information is distinctly limited. The field of view is small, the mirror itself is awkward and difficult to use, and only one member of the team is able to scan the room at a time. The information from the mirror is also displaced with respect to the team and requires mental rotation and translation to transform it into the team member's own point of view. This displacement, in particular, has been shown to reduce task performance and increase cognitive load [13, 26, 46], which likely is related to significantly reduced situation awareness.

In contrast to these methods, our system offers a less cognitively demanding and potentially less dangerous approach. Officers will be able to send a robot, equipped with a stereo camera, in ahead of the team. The robot will stream the depth image of the room to team members wearing mobile head-mounted AR displays. This information will be presented to the team members from their own perspective through an adjustable window. As such, the final result will be an egocentric view of the room, from each operator's perspective—no risk to officers and no awkward mental re-orientation required! Clearly, such an outcome (if it can be successfully incorporated into SWAT practice) could potentially, and dramatically, reduce the risk associated with standard room clearing operations.

3.2 Apparatus

The Depth Camera The Intel RealSense D435 stereo depth camera is designed specifically for low-light environments and supports a wide field of view out to ten meters, which is ideal for this use case [20]. The depth camera is used in tandem with the robot controlled by the SWAT team; this is comparable to approaches used in previous work [6, 7]. In this system, the depth camera will generate RGB point cloud data (PCD) in real time, which is immediately transferred to the operators' head-mounted displays.

The depth camera is rigged with a combination of Robot Operating System (ROS) and SLAM software and is equipped with gyroscopic hardware [4, 45]. Using these features, depth data

Fig. 2: A team member performing mirror reconnaissance. Note that the view the team member sees here is rotated and displaced from his current perspective.

can be either taken in independent 'snapshots' or can be used to continuously build a single environmental map. Interestingly, this data can also be stored for later retrieval and analysis, which may be useful for certain law enforcement applications.

The Magic Leap AR System The Magic Leap headset is a self-contained mobile AR display device. It supports automatic SLAM (simultaneous localization and mapping) environment and position mapping; speech, gesture, and peripheral-based control schemes; standard network protocols; approximately three hours worth of battery life; a 40° horizontal field of view; and a target refresh rate of 60 frames per second. Once the system program has been uploaded to each Magic Leap, they will operate completely independently (aside from syncing and depth data delivered over the network). This is particularly key in that the device will not limit team mobility while in use, nor will charging issues represent a significant constraint for most scenarios.

In interacting with the stereo depth camera, Unity, aided by certain tools and libraries, converts the output into a format which can be directly uploaded to the Magic Leap [38, 39]. The output is then displayed to operators as a mesh (see Figures 5 to 7).

As previously stated, by generating and displaying these environments, we are supplying SWAT teams with a safer and more efficient room surveying approach. The additional information these methods provide, combined with the intuitive perceptual understanding that AR supports, represents a potential advancement over the state of the art—provided that the positive aspects of this system outweigh any hidden costs.

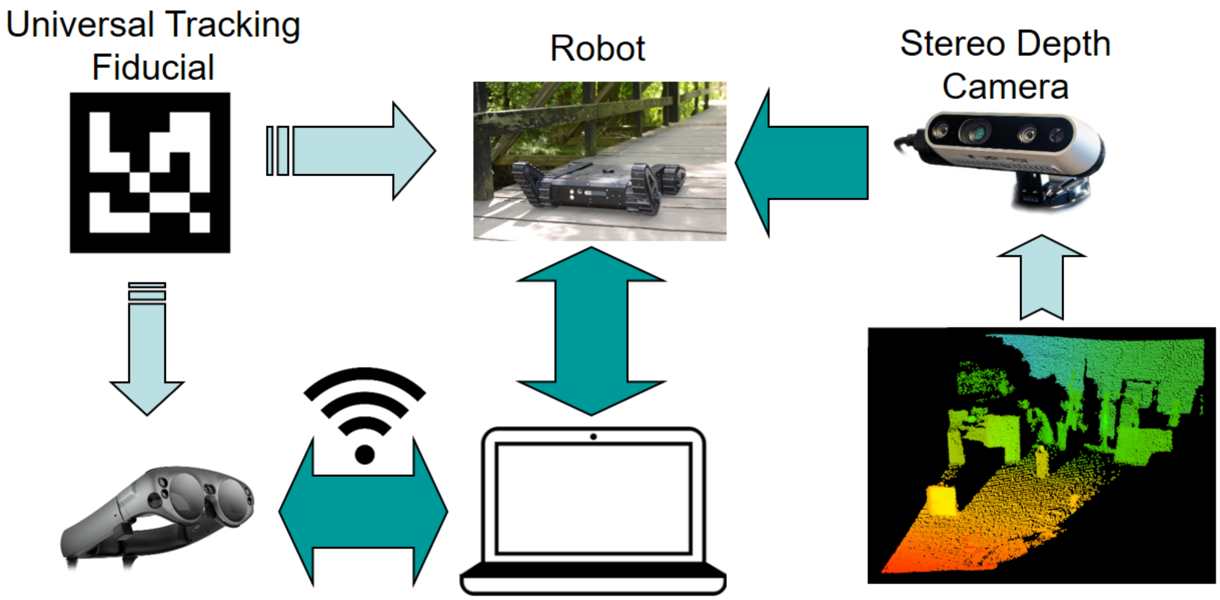

Fig. 3: Connection flow chart for the X-Ray Vision System. Depth images are taken by the stereo depth camera, processed, and delivered to a subscriber AR system through a secure network. Note that the robot and AR system positions are synchronized through the use of a universal tracking fiducial.

3.3 Implementation

In order to implement this system (See Figure 3), we use several connected components. First, we make use of a Jaguar V4 robot chassis, equipped with an Intel RealSense D435 camera. This robot will contain a computing system (supported by a secure wireless network) that allows the transfer of the point cloud room layout to team members. Each team member will be equipped with a Magic Leap AR system that provides stable augmentation and virtual images. Finally, the relative positions of both the robot and the Magic Leap systems will be synchronized and tracked using a printed fiducial. As an example of the system in action, see Figures 8 and 9.

The current system development is focused on the structure and design of the robotic platform. Interface improvements might also help improve the system's effectiveness and usability, while adding color and meshing to the displayed point cloud output may improve scene recognition. Certain practical issues, such as the frequency of displayed mesh updates, are also being investigated. These issues should be relatively easy to mitigate or solve with additional analysis. Finally, the important issue of testing and verification is being addressed.

3.4 System Effectiveness

This brings us to evaluating our system's effectiveness—what traits are worthy of analysis? Considering the nature of this intervention and the lack of predictive features, it seems best to evaluate

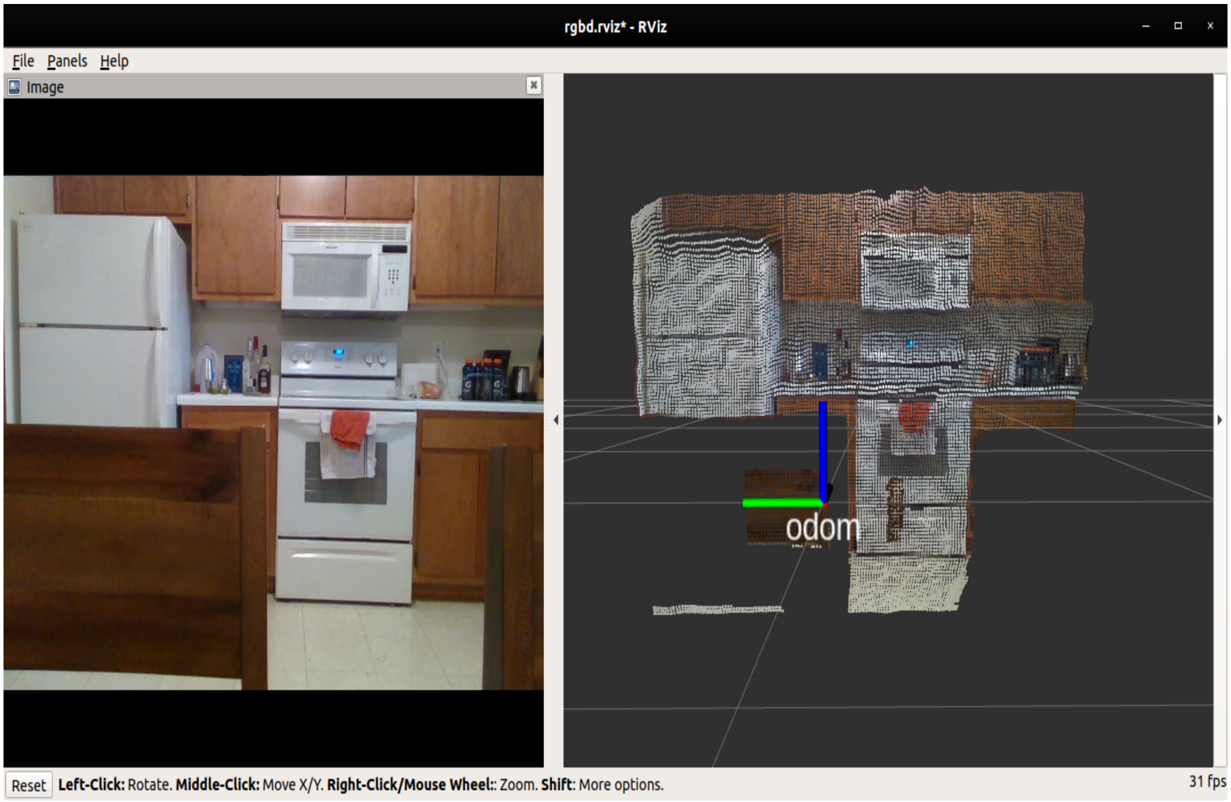

Fig. 4: A screenshot of a scanned room in Rviz

the perceptual validity of this system. Even though perception alone is not comprehensive situation awareness [15, 18], it is a key component and represents a notable step forward for team members who otherwise might not be able to know anything about the task environment prior to entry.

As such, there are at least three possible aspects of perception to consider: *memory*, *task performance*, and *depth perception*. Each of these represents, or is affected by, a part of our overall perception and, especially for the latter two, may be relevant to task performance and situation awareness.

Testing For each of the approaches listed below, testing would occur in a broadly similar manner. Each experimental approach would have three separate conditions: a real world condition using an actual, physical window; a mirror entry condition; and an x-ray vision condition. This allows us to consider each condition both on its own merits and in comparison to real world performance, to determine which approach is superior. This experimental data will be supplemented by a post-experiment survey.

Memory Testing In order to test operators' memory, we propose a simple environmental reconstruction experiment. In this task, an operator would observe and be encouraged to explore an adjacent room by looking in through a fixed window. After a short period of free-moving

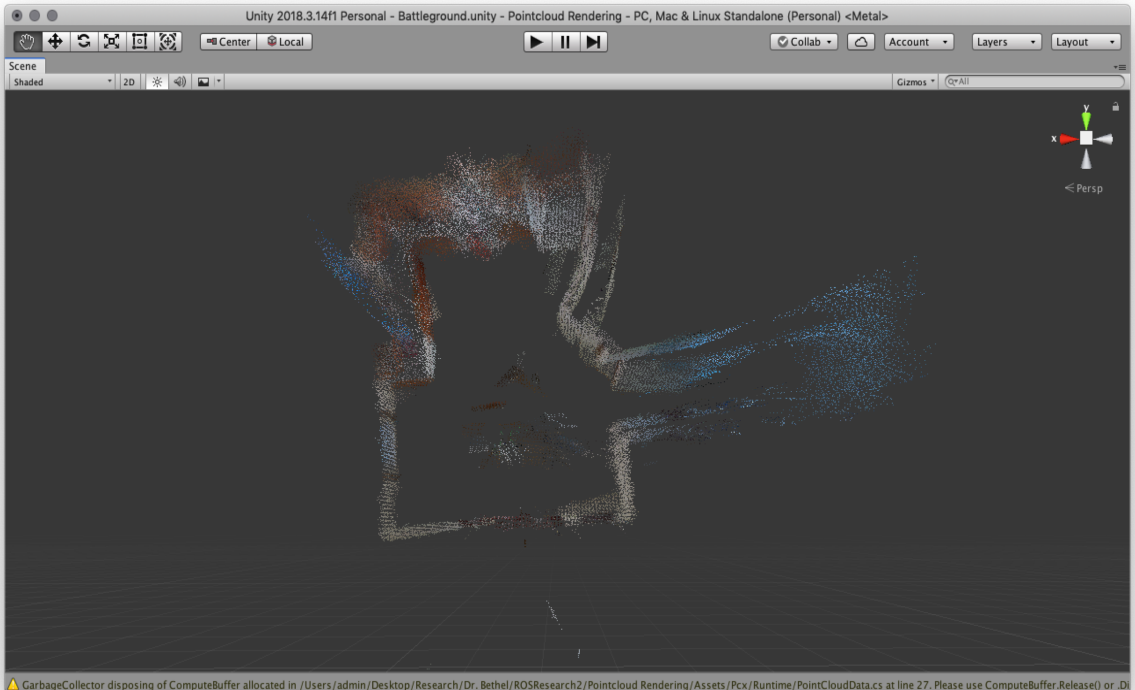

Fig. 5: An example converted point cloud data (PCD) file in Unity (top-down view).

exploration, the operator will be taken to a new area and tasked with reconstructing the explored scene. This reconstruction can be approached from a variety of angles; the operator could draw the scene from memory on a sketchpad or position physical objects identical to the objects in the observed environment. Further investigation is warranted to determine which of these approaches is best for this application. On the one hand, free-style drawing and spatial reasoning is definitively a practiced skill, with some operators being more proficient and some being dramatically less proficient. On the other hand, allowing the physical adjustment of room objects may provide additional feedback to the operator that they would not otherwise receive.

This experimental approach would allow the investigation of memory decay and accuracy. In particular, it will be interesting to observe how working memory functions when using the x-ray vision system, as opposed to strictly real world stimuli. These results will help determine how effectively the system replicates real world memory effects.

Task Performance Testing Task performance does not necessarily correlate directly with situation awareness [15, 16]. However, it may be possible to create a task that specifically is novel enough that training and performance differences don't apply and simple enough that completion is readily achievable.

In particular, team members during room clearing operations desire to have an accurate and complete model of the room prior to entry so that it can be navigated more easily and

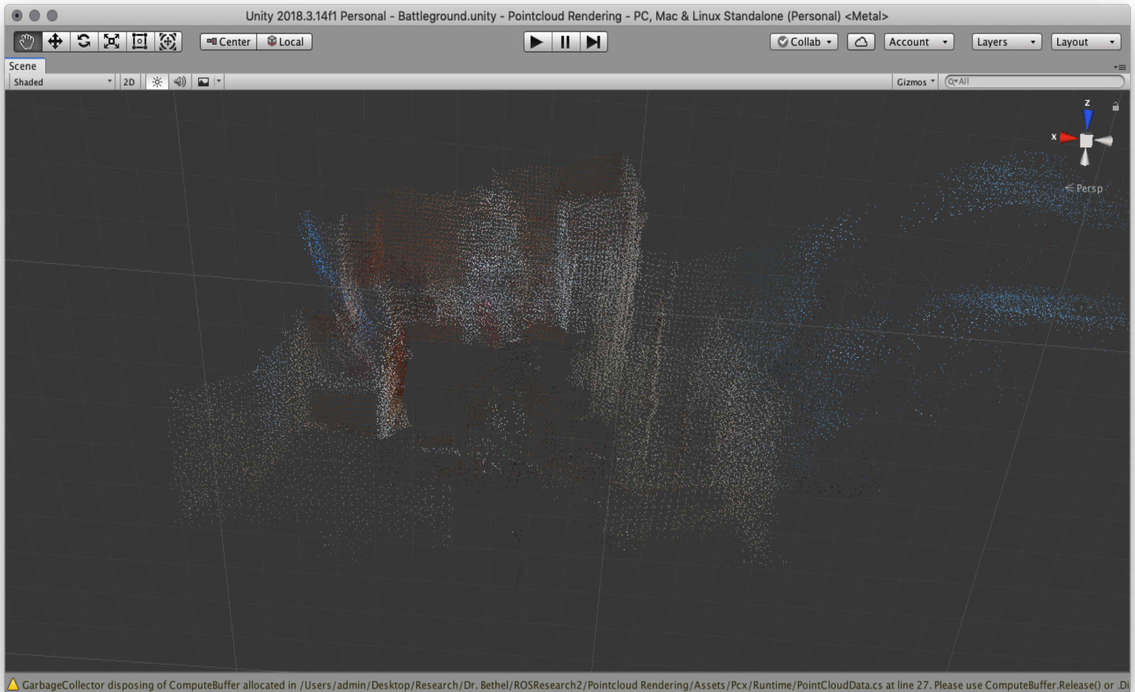

Fig. 6: An example converted point cloud data (PCD) file in Unity (camera view).

potential hiding spots identified and cleared. While effectively simulating a room clearing task with professional SWAT teams may be difficult, room navigation itself is much simpler—though the task becomes trivially easy upon entry into the room. However, if we require operators to rely solely on information acquired from the intervention, the outcome can be expected to rely heavily upon how well the system supports perception and memory. Thus, it may be useful to employ a monitored blind walking navigation task, where users observe the simulated room and are then instructed to enter it without sight. At that time, they would be given a simple navigation task in order to test their perceptual memory of the room layout. This approach has the advantage of testing task performance in a way that is both relevant to the room clearing operation and that is, at least at the beginning of the experiment, somewhat unpredictable to operators. As such, this approach may be a better measure of how a SWAT team member interacts with the system than an intervention where they know beforehand how they will be tested. To maintain safety during this experiment, a dedicated observer would be present to ensure that the participant does not run into any physical objects.

The results of this approach would clearly represent the expected effect of situation awareness prior to the room clearing task. Data could easily be gathered from positional reports by the Magic Leap and could consist of a variety of variables including: elapsed task time; task accuracy; and task errors, for any time a user attempts to walk through a location containing a displayed virtual object.

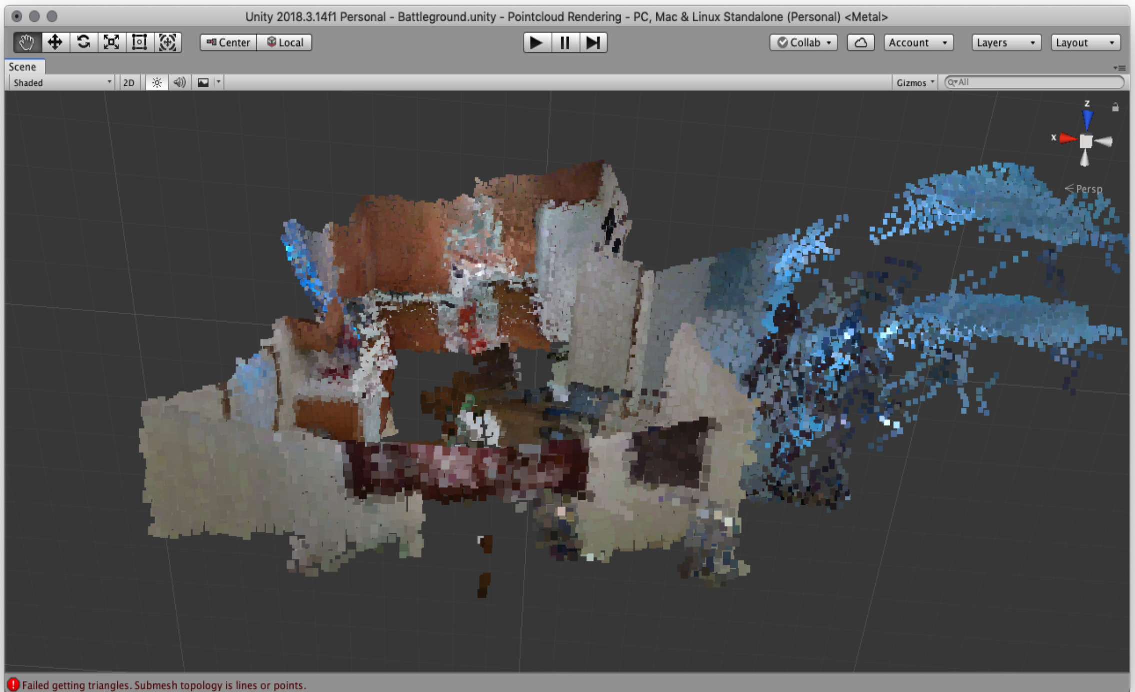

Fig. 7: The same data as Figure 6, but with point size increased, giving the mesh a blocky appearance.

Depth Perception Testing Depth perception, in this case, is an important aspect of situation awareness, though a high degree of accuracy isn't necessarily required for a room clearing task. One way that this might be measured is through a version of triangulated walking [34]. In classical depth perception experiments, triangulated walking tasks involve observing a virtual object that is observed briefly and then disappears. After it disappears, the subject is instructed to walk obliquely toward it while continuously pointing to its perceived location. This task has been observed to be an accurate experimental depth perception measurement and has regularly been used in scientific research [34].

This task could easily be modified for use in a hall; indeed, by having the subject move to several different locations in a hallway, both the perceived distance to the object and the associated memory footprint could be evaluated. This task might also indicate whether being subjected to movement/optical flow in the real world could distort the location of the remembered virtual object(s).

4 Discussion

As this project is still untested, this section primarily focuses on our expected results and any potential confounds. We discuss the expected general results of this system for improving situation

(a) The scanned depth data, as viewed through a closed door.

(b) The scanned depth data, with the scanned room visible.

Fig. 8: The X-Ray Vision System, displaying a standard office chair.

Fig. 9: The X-Ray Vision System, from several perspectives.

awareness, the anticipated use and benefit of AR and the window metaphor, and the possible challenges for the integration of the system into SWAT operations.

4.1 Hypotheses and Predictions

Some of the proposed tests have not been previously tried in this sort of venue before (e.g., memory reconstruction). Other tests either have not been applied in exactly this way before (e.g., blind walking), or dramatically differing technologies and levels of virtuality/depth cuing have been used in previous experiments (e.g., triangulated walking) [28, 51]. As such, expected results are hard to predict. We would expect results generated from the system to be approximately equivalent to results generated under real world conditions—which, for blind and triangulated walking, would indicate generally veridical depth judgments. On the other hand, we would expect results from the mirror test to be significantly worse than the real world for depth tasks and likely also worse for memory reconstruction. The relationship between memory reconstruction and the system may also be more complicated than anticipated; some previous research has theorized that presenting items in augmented reality may have novelty or arousal-inducing effects, which could potentially modify or strengthen the operator's memory of item locations [19]. This effect might be caused by the novelty inherent for most users in AR displays; it might also represent a more general tendency for operators to fixate on virtual content over real world content. Even so, we would expect that this effect would be relatively minor and that memory reconstruction would be approximately equivalent in both the system and the real world.

4.2 The Window Metaphor—Effectiveness and Usability

Previous research has noted a discrepancy between perception in egocentric and allocentric reference frames, causing cognitive displacement and confusion when participants attempt to rectify their location in the physical and virtual worlds [10, 49]. This phenomenon has been known to cause simulator sickness (or a general feeling of nausea induced by exposure to VR or AR) and other negative effects. In addition, views from non-egocentric frames of reference, referred to as displaced perspective, often exhibit increased cognitive demands for users [13, 26, 46]. This increased attentional overhead may have a negative effect on situation awareness and, thus, likely negative effects on performance in a standard room clearing task. Thus, this system focuses on egocentric perception from the ground up, sidestepping all of the significant costs of non-user-centered projections.

Another area that is important to consider is the effectiveness of the window metaphor. While we have not yet collected concrete data on this issue, subjective experience seems clear; the window metaphor appears to be an effective way to help operators mentally categorize the perceptual data. Without the window, the displayed data can appear chaotic and difficult to place, but, with the window, analysis and examination seem more intuitive. This may also be reinforced by the need to move. In the absence of a window, all the data can be viewed from a single point, while, with the window, movement or repositioning of the operator or window is necessary to view more than a small portion of data. Thus, in the presence of the window, operators are incentivized to move about, thus introducing motion parallax depth cues, which help more concretely establish and reinforce distance judgments to the displayed data. It might be interesting to more fully investigate this idea in future experimentation, but, for now, it seems subjectively clear that the window metaphor results in improved situation awareness and depth judgments.

4.3 SWAT Team Contextualization

In contrast, the reception of this system among SWAT team members is likely to be a more complicated variable. On the one hand, pre-task intelligence is expected to be highly valuable to SWAT team members, in terms of both its extrinsic value and its regard among the team members. However, this system will also be associated with several cognitive costs, some of which team members may be aware and some of which they may not be. In particular, the deployment and set up of this system is expected to take a few minutes of up-front time. In most contexts, this will be trivial, but in an incident where speed is extremely important, deployment of this technology could be stressful or result in increased cognitive demands. Therefore, system deployment procedures, and deployment training, are an important consideration. In addition, this task requires integrating a robotic system into SWAT team procedures. While several such integrations have proven successful and popular, it does still remain that current robotics platforms can be noisy, slow, and cumbersome, as compared to team members [2]. Therefore, while it seems likely that the addition of valuable intelligence and the lower risk of fatalities will offset these characteristics in the mind of SWAT team members, it remains to be seen how exactly they will interact with, and react to, this new technology.

5 Future Work

The clear next step for this project is to polish and finalize the various details of the x-ray vision system and begin testing its effectiveness, first with untrained users and then with SWAT team members. These details, while relatively minor, may have a significant impact on performance and operator reception. Once these are finalized, proper testing and experimentation can begin (see Section 4).

5.1 Environment Generation & Integration

One of the primary forthcoming issues to consider is how to display the point cloud room data in the Magic Leap. The underlying problem of getting the data to each Leap is largely solved, but determining the best parameters for displaying it is more open ended. How often should the displayed scene be refreshed from the camera? Updating too often risks bogging down the Magic Leap's processing power and cause rubber banding; updating too rarely, on the other hand, might cause unexpected perceptual effects or somewhat delay the receipt of critical mission data. Similar questions can be asked about ideal point count to be displayed, the usefulness of meshing or color for operators, and whether previously viewed areas should continue to be displayed or only the area currently visible to the depth camera.

5.2 Robot Design and Performance

Functionally, it is also important to finalize the production design of the robotic platform. The physical location of components and wiring all have to be analyzed for faults or weaknesses before the robot can be used in practical applications. In addition, as mentioned previously, expanding robot autonomy would be an interesting next step for this project. It is well beyond the current scope, but expanding the system to potentially encompass multiple semiautonomous, independent robots promises to be an exceptional step forward in large-scale room clearing tasks. If such a robotic fleet could effectively communicate with each other and the tactical team, it would be possible to dramatically reduce the time required to clear large buildings or even urban environments.

5.3 Testing and Experimentation

Finally, the most important future work is, arguably, testing and experimentation. As mentioned in Section 4, there are a number of experiments that can be used to determine if the x-ray vision system is an effective way to improve perception. However, situation awareness goes beyond more than just perception—an important next step would be scenario-based testing, where team members might have to comprehend or predict based on what they see. These scenarios would be fairly challenging to implement and would have to be presented in a surprising manner so as to better test the team's unprepared reactions and awareness. Other tests might include data-based analysis of various smaller components of the x-ray vision system: the effect of motion parallax on depth judgments, the effectiveness of the window metaphor vs. an unbounded display, or potentially several other system components.

6 Conclusion

We expect that this x-ray vision system will be an effective aid to tactical SWAT teams in room clearing tasks. It seems likely that presenting previously unknown room information from an egocentric frame of reference will allow team members to easily and intuitively understand, plan, and react to it. These benefits, too, seem to be free of major downsides that could bog down or hinder team efficiency. As such, it is expected that the x-ray vision system will significantly aid in team situation awareness, though further testing is required.

Acknowledgements

This material is based upon work supported by the National Science Foundation, under awards IIS-1937565, to J.E. Swan II and C.L. Bethel, and IIS-1320909, to J.E. Swan II. We acknowledge a productive collaboration with Mark Ballard, Chief of Police, Police Department, City of Starkville, MS, USA. We also acknowledge the contributions of Mohammed Safayet Arefin.

References

[1] R Bane and T Hollerer. "Interactive tools for virtual x-ray vision in mobile augmented reality". In: *Third IEEE and ACM international symposium on mixed and augmented reality*. IEEE. 2004, pp. 231–239.

[2] CL Bethel, D Carruth, and T Garrison. "Discoveries from integrating robots into SWAT team training exercises". In: *2012 IEEE International Symposium on Safety, Security, and Rescue Robotics (SSRR)*. IEEE. 2012, pp. 1–8.

[3] CL Bethel, DC May, and J Louine. "Use of Technology by Law Enforcement". In: *2016 International Police Executive Symposium (IPES)*. Washington, DC, USA: IPES, 2016.

[4] Y Bi, J Li, H Qin, M Lan, M Shan, F Lin, and BM Chen. "An MAV localization and mapping system based on dual realsense cameras". In: *Int. Micro Air Vehicles, Conf. Competitions, Nat. Univ. Singapore, Singapore, Tech. Rep.* 2016.

[5] C Bichlmeier, T Sielhorst, SM Heining, and N Navab. "Improving depth perception in medical ar". In: *Bildverarbeitung für die Medizin 2007*. Springer, 2007, pp. 217–221.

[6] J Biswas and M Veloso. "Depth camera based indoor mobile robot localization and navigation". In: *2012 IEEE International Conference on Robotics and Automation*. May 2012, pp. 1697–1702. DOI: 10.1109/ICRA.2012.6224766.

[7] J Biswas and M Veloso. "Depth camera based indoor mobile robot localization and navigation". In: *2012 IEEE International Conference on Robotics and Automation*. IEEE. 2012, pp. 1697–1702.

[8] MF Bradshaw, AD Parton, and RA Eagle. "The interaction of binocular disparity and motion parallax in determining perceived depth and perceived size". In: *Perception* 27.11 (1998), pp. 1317–1331.

[9] MF Bradshaw, AD Parton, and A Glennerster. "The task-dependent use of binocular disparity and motion parallax information". In: *Vision research* 40.27 (2000), pp. 3725–3734.

[10] A Byagowi and Z Moussavi. "Design of a virtual reality navigational (VRN) experiment for assessment of egocentric spatial cognition". In: *2012 Annual International Conference of the IEEE Engineering in Medicine and Biology Society*. IEEE. 2012, pp. 4812–4815.

[11] J Cutting and P Vishton. *Perception of space and motion. Handbook of perception and cognition*. 1st. San Diego: Academic Press, Inc., 1995.

[12] D Drascic and P Milgram. "Perceptual Issues in Augmented Reality". In: *Proceedings-SPIE The International Society For Optical Engineering*. Vol. 2653. SPIE International Society for Optical Engineering. 1996, pp. 123–134.

[13] SR Ellis, BD Adelstein, and K Yeom. "Human control in rotated frames: anisotropies in the misalignment disturbance function of pitch, roll, and yaw". In: *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*. Vol. 56. SAGE Publications Sage CA: Los Angeles, CA. 2012, pp. 1336–1340.

[14] MR Endsley. "Measurement of situation awareness in dynamic systems". In: *Human factors* 37 (1995), pp. 65–84.

[15] MR Endsley, DJ Garland, et al. "Theoretical underpinnings of situation awareness: A critical review". In: *Situation awareness analysis and measurement* 1 (2000), p. 24.

[16] MR Endsley, SJ Selcon, TD Hardiman, and DG Croft. "A comparative analysis of SAGAT and SART for evaluations of situation awareness". In: *Proceedings of the human factors and ergonomics society annual meeting*. Vol. 42. SAGE Publications Sage CA: Los Angeles, CA. 1998, pp. 82–86.

[17] L Furness. "The application of head-mounted displays to airborne reconnaissance and weapon delivery". In: *Wright-Patterson Air Force Base, Ohio, USA* (1969).

[18] Y Gatsoulis, GS Virk, and AA Dehghani-Sanij. "On the measurement of situation awareness for effective human-robot interaction in teleoperated systems". In: *Journal of cognitive engineering and decision making* 4.1 (2010), pp. 69–98.

[19] T Hopp and H Gangadharbatla. "Novelty effects in augmented reality advertising environments: The influence of exposure time and self-efficacy". In: *Journal of Current Issues & Research in Advertising* 37.2 (2016), pp. 113–130.

[20] *Intel Realsense D435*. 2019. URL: https://www.intelrealsense.com/depth-%20camera-d435/ (visited on 02/06/2020).

[21] EB Johnston, BG Cumming, and MS Landy. "Integration of stereopsis and motion shape cues". In: *Vision research* 34.17 (1994), pp. 2259–2275.

[22] JA Jones, JE Swan II, G Singh, E Kolstad, and SR Ellis. "The effects of virtual reality, augmented reality, and motion parallax on egocentric depth perception". In: *Proceedings of the 5th symposium on Applied perception in graphics and visualization.* ACM. 2008, pp. 9–14.

[23] B Julesz. "Binocular Depth Perception without Familiarity Cues: Random-dot stereo images with controlled spatial and temporal properties clarify problems in stereopsis". In: *Science* 145.3630 (1964), pp. 356–362.

[24] V Jurgens, A Cockburn, and M Billinghurst. "Depth cues for augmented reality stakeout". In: *Proceedings of the 7th ACM SIGCHI New Zealand chapter's international conference on Computer-human interaction: design centered HCI.* 2006, pp. 117–124.

[25] M Kersten-Oertel, P Jannin, and DL Collins. "The state of the art of visualization in mixed reality image guided surgery". In: *Computerized Medical Imaging and Graphics* 37.2 (2013), pp. 98–112.

[26] RL Klatzky, B Wu, and G Stetten. "The disembodied eye: Consequences of displacing perception from action". In: *Vision research* 50.24 (2010), pp. 2618–2626.

[27] R Konrad, N Padmanaban, E Cooper, and G Wetzstein. "Computational Focus-Tunable Near-Eye Displays". In: *ACM SIGGRAPH 2016 Emerging Technologies.* ACM. 2016, p. 3.

[28] BR Kunz, L Wouters, D Smith, WB Thompson, and SH Creem-Regehr. "Revisiting the effect of quality of graphics on distance judgments in virtual environments: A comparison of verbal reports and blind walking". In: *Attention, Perception, & Psychophysics* 71.6 (2009), pp. 1284–1293.

[29] A Lalejini, D Duckworth, R Sween, CL Bethel, and D Carruth. "Evaluation of supervisory control interfaces for mobile robot integration with tactical teams". In: *2014 IEEE International Workshop on Advanced Robotics and its Social Impacts.* IEEE. 2014, pp. 1–6.

[30] S Liu, Y Li, P Zhou, X Li, N Rong, S Huang, W Lu, and Y Su. "A Multi-Plane Optical See-Through Head Mounted Display Design for Augmented Reality Applications". In: *Journal of the Society for Information Display* 24.4 (2016), pp. 246–251. issn: 10710922.

[31] MA Livingston, Z Ai, JE Swan II, and HS Smallman. "Indoor vs. outdoor depth perception for mobile augmented reality". In: *2009 IEEE Virtual Reality Conference.* IEEE. 2009, pp. 55–62.

[32] MA Livingston, A Dey, C Sandor, and BH Thomas. "Pursuit of "X-ray vision" for augmented reality". In: *Human Factors in Augmented Reality Environments.* Springer, 2013, pp. 67–107.

[33] MA Livingston, JE Swan II, JL Gabbard, TH Hollerer, D Hix, SJ Julier, Y Baillot, and D Brown. "Resolving multiple occluded layers in augmented reality". In: *The Second IEEE and ACM International Symposium on Mixed and Augmented Reality, 2003. Proceedings.* IEEE. 2003, pp. 56–65.

[34] JM Loomis and JM Knapp. “Visual perception of egocentric distance in real and virtual environments”. In: *Virtual and Adaptive Environments*. Ed. by LJ Hettinger and MW Haas. Mahwah, Jan. 1, 2003.

[35] RD McIntosh and G Lashley. “Matching boxes: Familiar size influences action programming”. In: *Neuropsychologia* 46.9 (2008), pp. 2441–2444.

[36] RP O’Shea, SG Blackburn, and H Ono. “Contrast as a depth cue”. In: *Vision research* 34.12 (1994), pp. 1595–1604.

[37] S Palmisano. “Perceiving self-motion in depth: The role of stereoscopic motion and changing-size cues”. In: *Perception & Psychophysics* 58.8 (1996), pp. 1168–1176.

[38] *PCLLibrary*. 2020. URL: `http://pointclouds.org` (visited on 02/06/2020).

[39] *PCX*. 2020. URL: `https://github.com/keijiro/Pcx` (visited on 02/06/2020).

[40] N Phillips, K Massey, MS Arefin, and JE Swan II. “Design, Assembly, Calibration, and Measurement of an Augmented Reality Haploscope”. In: *2019 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*. IEEE. 2019, pp. 1770–1774.

[41] JJ Rieser, HL Pick, DH Ashmead, and AE Garing. “Calibration of human locomotion and models of perceptual-motor organization.” In: *Journal of Experimental Psychology: Human Perception and Performance* 21.3 (1995), p. 480.

[42] B Rogers and M Graham. “Motion parallax as an independent cue for depth perception”. In: *Perception* 8.2 (1979), pp. 125–134.

[43] NB Sarter and DD Woods. “Situation awareness: A critical but ill-defined phenomenon”. In: *The International Journal of Aviation Psychology* 1.1 (1991), pp. 45–57.

[44] G Singh, SR Ellis, and JE Swan II. “The effect of focal distance, age, and brightness on near-field augmented reality depth matching”. In: *IEEE transactions on visualization and computer graphics* (2018).

[45] *SLAM with D435i*. 2019. URL: `https://github.com/IntelRealSense/realsens%20e-ros/wiki/SLAM-with-D435i` (visited on 02/06/2020).

[46] KU Smith and WM Smith. “Perception and motion.” In: *American Psychological Association* (1962).

[47] J Stevens and L Eifert. “Augmented reality technology in U.S. army training (WIP)”. In: *Proceedings of the 2014 Summer Simulation Multiconference*. Monterey, California: Society for Computer Simulation International, 2014, pp. 1–6.

[48] AG Taylor. *Develop Microsoft hololens apps now*. Springer, 2016.

[49] M Vidal, MA Amorim, and A Berthoz. “Navigating in a virtual three-dimensional maze: how do egocentric and allocentric reference frames interact?” In: *Cognitive Brain Research* 19.3 (2004), pp. 244–258.

[50] M Voshell, DD Woods, and F Phillips. “Overcoming the keyhole in human-robot coordination: simulation and evaluation”. In: *Proceedings of the Human Factors and Er-*

gonomics Society Annual Meeting. Vol. 49. Sage Publications Sage CA: Los Angeles, CA. 2005, pp. 442–446.

[51] P Willemsen, MB Colton, SH Creem-Regehr, and WB Thompson. "The effects of head-mounted display mechanics on distance judgments in virtual environments". In: *Proceedings of the 1st Symposium on Applied perception in graphics and visualization*. ACM. 2004, pp. 35–38.