

## Spatial Accuracy Requirements for Visual Search Cues in Simulated Sparse and Dense Scenes

**John J. Graybeal**  
U.S. Army DEVCOM C5ISR  
Fort Belvoir, VA  
john.j.graybeal.civ@us.navy.mil

**Colleen E. Gerrity**  
MAG Aerospace  
Fairfax, VA  
colleen.e.gerrity.ctr@army.mil

**William Heath Sharp**  
Planned Systems International  
Arlington, VA  
william.h.sharp16.ctr@army.mil

**Emily Lasko**  
CACI International, Inc.  
Reston, VA  
emily.n.lasko.ctr@army.mil

**Todd W. Du Bosq**  
U.S. Army DEVCOM C5ISR Center  
Fort Belvoir, VA  
todd.w.dubosq.civ@army.mil

### ABSTRACT

Sensors and displays provide situational awareness to Soldiers and allow operators to locate allies, enemies, civilians, and other objects of interest. A variety of technologies aim to help Soldiers find threats and other objects in their environment by providing Soldiers with visual cues. Although these cues can be displayed in different ways, they all attempt to orient the user's gaze towards critical information. However, for these cues to be beneficial, they must be displayed with enough spatial accuracy to correctly draw the user's attention to the intended object. This research explores the spatial accuracy requirements for cues to improve human performance relative to unaided search. Past research has demonstrated that as the range to the cued object increases, visual cues can provide greater benefits to human performance, but the negative effects of imprecision also increase. This research expands on previous work by examining an additional dimension of task difficulty: the density of objects in the scene that must be searched. In sparse scenes, an imprecise visual cue may still be sufficient to orient the user to the correct object. However, in scenes with many objects, spatially imprecise cues may result in insufficient guidance. In this experiment, Soldiers repeatedly searched simulated scenes for a person carrying a weapon while receiving various types of assistance. To characterize the effects of spatial imprecision, Soldiers searched some scenes with imperfect visual cues. The experiment investigated the effects of multiple levels of spatial imprecision (i.e., angular error between the cued threat and the displayed cue) at multiple ranges. To characterize baseline performance, Soldiers searched other scenes without assistance. To establish the maximum possible benefits, Soldiers searched additional scenes with perfect cues. Collectively, the data demonstrate the substantial benefits of visual search cues while simultaneously demonstrating the negative effects of spatial imprecision.

**Keywords:** Visual search, human performance, human factors, target detection, augmented reality

### ABOUT THE AUTHORS

**John J. Graybeal** worked as a Lead Engineering Psychologist at the U.S. Army DEVCOM C5ISR Center. He received his Ph.D. in Psychology in 2017 from George Mason University, with a concentration in Cognitive and Behavioral Neuroscience. His research involved evaluating human performance with a wide range of sensor imagery and emerging technologies. He regularly conducted research with Soldiers using visual tasks (e.g., object detection, object identification) with both real and simulated imagery. During his tenure, he led efforts to modernize and refine human testing methodologies with a special focus on establishing virtual prototyping capabilities for the organization. His research also focused on human performance with augmented/virtual reality technologies and human use of imperfect systems. He is the recipient of the 2019 NATO Modelling and Simulation Group's Young

Scientist Award and multiple conference awards, including the 2019 I/ITSEC Best Paper Award. He recently transferred to the Naval Air Systems Command, where he continues to research ways to improve warfighter technologies.

**Colleen E. Gerrity** is employed by MAG Aerospace and works as an Engineering Psychologist for the U.S. Army DEVCOM C5ISR Center. She earned a Master of Arts in Psychology with a Concentration in Cognitive and Behavioral Neuroscience from George Mason University in 2022. She also earned a Bachelor of Science in Psychology with a Minor in Neuroscience from George Mason University in 2021. At C5ISR Center, Colleen supports the Human Perception Laboratory on the study of how thermal sensors and advanced displays impact Soldiers and how they can be improved. She takes a “Soldier-centered” design approach to measure human performance and cognitive processing with sensor imagery to help inform and evaluate emerging technologies.

**William Heath Sharp** is employed by Planned Systems International, Inc. and works as a Research Scientist for the U.S. Army DEVCOM C5ISR Center. He earned his Master of Arts in Human Factors Psychology in 2019 from the Catholic University of America and is currently finishing his PhD in Human Factors and Applied Cognition from George Mason University. His research focus at GMU is on trust in automation, human-automation teaming, and longitudinal trust. At C5ISR Center, Heath supports the Human Perception Laboratory on the study of how Soldiers interact with thermal sensors and advanced displays, specifically accurately measuring and quantifying human performance and cognitive processing with sensor imagery to help inform and evaluate emerging technologies.

**Emily Lasko** is employed by CACI International, Inc. and works as a Research Psychologist for the U.S. Army DEVCOM C5ISR Center. She earned a Master of Science in Experimental Psychology in 2017 and a PhD in Social Psychology with a focus on Social Affective Neuroscience and Experimental Design in 2021. At C5ISR Center, Emily supports the Human Perception Laboratory on the study of how Soldiers interact with thermal sensors and advanced displays, specifically accurately measuring and quantifying human performance and cognitive processing with sensor imagery to help inform and evaluate emerging technologies.

**Todd W. Du Bosq** serves as the Chief Scientist for the Modeling and Simulation Division at the U.S. Army DEVCOM C5ISR Center. In this role, he brings extensive expertise in modeling and simulation for Electro-Optical/Infrared (EO/IR) sensors, contributing to advancements in defense technologies. Dr. Du Bosq received his Bachelor of Science in Physics from Stetson University in 2001, followed by a Master of Science and PhD in Physics from the University of Central Florida in 2003 and 2007, respectively.

## **Spatial Accuracy Requirements for Visual Search Cues in Simulated Sparse and Dense Scenes**

**John J. Graybeal**  
U.S. Army DEVCOM C5ISR Center  
Fort Belvoir, VA  
john.j.graybeal.civ@us.navy.mil

**Colleen E. Gerrity**  
MAG Aerospace  
Fairfax, VA  
colleen.e.gerrity.ctr@army.mil

**William Heath Sharp**  
Planned Systems International  
Arlington, VA  
william.h.sharp16.ctr@army.mil

**Emily Lasko**  
CACI International, Inc.  
Reston, VA  
emily.n.lasko.ctr@army.mil

**Todd W. Du Bosq**  
U.S. Army DEVCOM C5ISR Center  
Fort Belvoir, VA  
todd.w.dubosq.civ@army.mil

### **INTRODUCTION**

The U.S. Army DEVCOM C5ISR Center is a world-class leader in the development, evaluation, and integration of state-of-the-art electro-optical and infrared (EO/IR) sensors, which provide significant tactical advantages to enhance the effectiveness and safety of Soldiers in the field. The maturation of unobtrusive, mobile platforms, such as see-through and helmet-mounted display technologies, have made it possible to improve the way we present sensor information to a human operator. As such, research on augmented reality (AR) technologies that attempt to enhance human sensory experiences by inserting digital information into the user's experience of the "real world" (Wu et al., 2013) have become an important aspect of the C5ISR Center's mission, as part of the U.S. Army's focus on increasing Soldier situational awareness.

The goals of enhancing situational awareness and improving operational task performance by providing Soldiers with increasingly complex information is becoming attainable due to the research and development of AR technologies. However, AR technologies can also pose risks to human performance. AR may capture visual attention in undesirable ways, such that operators become inattentive to other critical information in the "real world," or the AR systems may increase cognitive load (Tang et al., 2003; Dixon et al., 2013; Dunleavy & Dede, 2014; Radu, 2014).

For AR technologies to be successful and for information to benefit Soldiers, it must be of sufficient quality, and there are multiple dimensions of informational quality. It must be perceivable, easily comprehensible, relevant, timely, and accurate. If any of these dimensions of informational quality are sufficiently lacking, the information will not improve human operator performance and may even impair it. Fortunately, simulation can be used to systematically vary the quality of information presented to Soldiers to provide insight into how improvements or degradations affect operator performance (Monfort et al., 2018; Graybeal et al., 2019a; Graybeal et al., 2023). The results of these simulations can help system engineers make informed decisions about information requirements for a given task, which can subsequently help ensure a system is built to provide a sufficient quality of information to improve Soldier performance.

Visual cues are designed to support a variety of tasks, such as navigation and target detection, by orienting the operator's attention in a way that supports task completion. However, the spatial registration of AR symbology remains a significant technical challenge (Brunyé et al., 2016). There are many reasons why a system may fail to present a perfectly placed visual cue to the operator's eye. The quality of information entering the system, the computational limitations of the system, limitations in the AR display, and imperfect use or calibration of the display are all potential sources of spatial error that could cause a visual cue to be perceived in a location that is either

slightly or severely different than its ideal or intended position. As such, it is important to understand the consequences of presenting an operator visual cues with imperfect spatial accuracy.

At the 2019 Interservice/Industry Training, Simulation, and Education Conference (IITSEC), the C5ISR Center presented the results of a simulation that explored the spatial accuracy of visual cue information (Graybeal et al., 2019b). Specifically, we examined the level of spatial accuracy required for visual cues to improve human performance as a function of range to target. The experiment required Soldiers to scan a virtual field of regard and to locate a target without any visual cue assistance, with perfectly accurate visual cue assistance, and with visual cue assistance degraded by fixed amounts of angular error (1°, 2°, 3°, or 4°). The simulation demonstrated that perfect visual cues improved the operator's ability to find targets at all ranges; these benefits increased as range to target increased, and the task became more difficult. Similarly, the degradations to visual cue accuracy had minimal impact on the operator's ability to find targets at the closest range, but incremental visual cue errors progressively increased target acquisition time as range to target increased.

The current research expands on our previous work (Graybeal et al., 2019b) by examining the density of targets in the scene that must be searched, which has the potential to impact the difficulty of the search task and, subsequently, the quality of information needed to improve human task performance. The purpose of the current investigation is to examine the additional impact that scene density has on the effects of visual cue angular error as a function of range to target. When utilizing visual-only AR techniques (i.e., visual cues), denser scenes may lead to degradation in task performance due to the introduction of clutter, which may interfere with decision-making processes, as well as limit object recognition and segmentation (Kruijff et al., 2010). Furthermore, a study that varied object density in an AR-based task found that task accuracy decreased while time to complete the task increased as object density increased (Chen et al., 2021). Consequently, we specifically investigate how scene density influences the effects of visual cue angular error and range to target. Our primary research questions explore how much visual cue accuracy in sparse and dense scenes is needed for visual cue information to 1) improve human performance, 2) harm human performance, and 3) decrease performance below that of perfect visual cue guidance. As the current research is a second study that builds on our previous experiment (Graybeal et al., 2019b), much of the background and methodology are shared. For context and for the reader's convenience, any parallels between the two experiments will be reiterated in this manuscript, often verbatim.

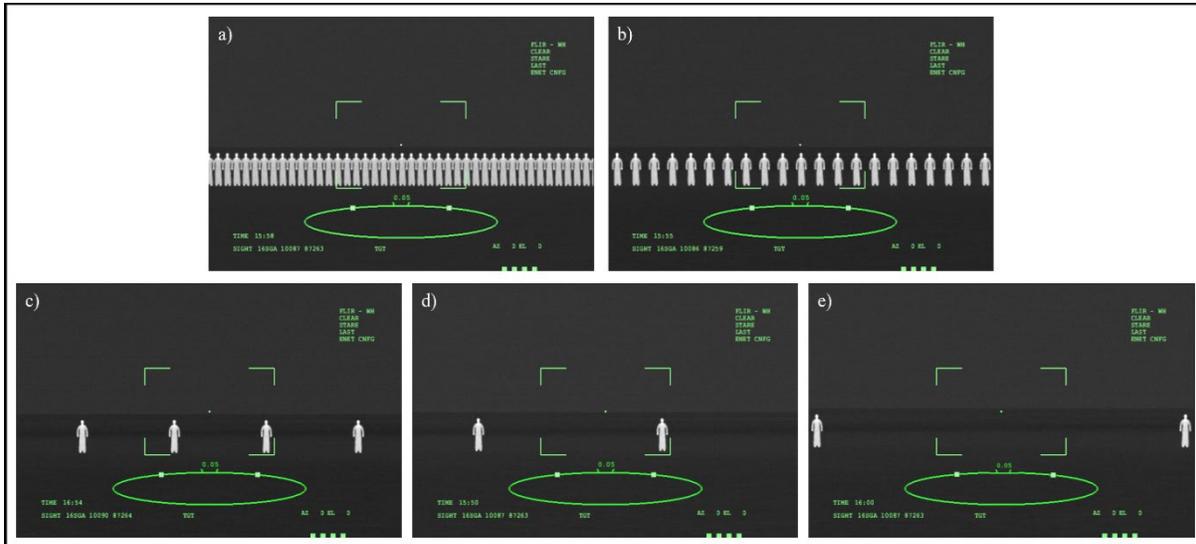
## METHODOLOGY

### Experimental Design and Hypotheses

We studied the effects of several independent variables on target acquisition performance using a within-subjects design. First, we studied the effects of visual cue accuracy. Participants experienced five categorical levels of the visual cue performance: No Visual Cue (i.e., a control condition where participants had to complete the task unaided by any visual cues), Perfect Visual Cue (i.e., no angular displacement between the visual cue and the true target), and three levels of Imperfect Visual Cue, consisting of 1°, 3°, and 5° of angular error between the visual cue and the true target. To keep the length of the experiment manageable while adding a new independent variable (i.e., scene density), this experiment used fewer levels of angular error than the previous experiment. Keeping with our past research, we hypothesized that greater amounts of visual cue error would increasingly impair target acquisition performance.

Second, we explored the effects of distance between the sensor and the target (i.e., range). We used the same methodological approach as the previous study by presenting targets at a Close range (where the target was easily visible without engaging the sensor's optical zoom), an Intermediate range (where the target was visible without engaging the optical zoom but optical zoom greatly aided target acquisition), and a Distant range (where the target was not detectable without engaging optical zoom). Based on theory and our past results, we hypothesized that target acquisition would take longer with targets at more distant ranges. We also hypothesized the previously observed interaction between range and visual cue information: compared to their effects at closer ranges, visual cues have more potential to improve target acquisition performance at more distant ranges, but the negative effects of angular error are also magnified at more distant ranges.

Third, and novel to the current experiment, we explored the effects of space between objects in the scene that must be searched (i.e., scene density). We studied the effects of five intervals of space between the virtual humans: 0.5 meter (m), 1 m, 5 m, 10 m, and 20 m (see **Figure 1**). Scenes with 20 m of space between the virtual humans were the sparsest, whereas scenes with 0.5 m of space between the virtual humans were the densest. First, we hypothesized that it would take longer to find and acquire targets in denser scenes compared to sparser scenes. Considering that manipulating scene density is another method of manipulating task difficulty, like manipulating range, we hypothesized that there would be an interaction between density and visual cue information, such that the Perfect Visual Cue Condition would provide more benefits in denser scenes (compared to the No Visual Cue Condition), but that angular error would also cause more harm in denser scenes. We also explored the interaction between density and range and whether the interaction between density and visual cue information differed based on range.



**Figure 1.** Sample scene densities from the simulation: a) 0.5 m, b) 1 m, c) 5 m, d) 10 m, and e) 20 m.

### Scene Generation

Leveraging the previous simulation tools (Graybeal et al., 2019b), we modified a series of virtual scenarios where participants had to search for, detect, and acquire a target. These interactive scenarios were created using the C5ISR Center-owned and -developed Night Vision Image Generator (NVIG) software to simulate an infrared sensor feed (Harkrider et al., 2011; May et al., 2011). NVIG has the ability to represent a virtual world from a sensor viewpoint that can be controlled by a human operator (i.e., ability to rotate a sensor 360° in an immersive virtual world). The simulation presented the operator with a series of scenarios sequentially (i.e., when a given experimental trial was completed and the user gave a response, the next scenario loaded automatically). The simulated long-wave infrared (LWIR) sensor used in the scenarios had a resolution of 768x480 pixels, along with two horizontal field of view options: 8° for a wide field of view and 2.7° for a narrow field of view.

Each scene contained only one target, defined as a human holding a rifle, and a varying number of non-target humans. The virtual humans were arranged in a partial ring (i.e., arc) around the sensor's location, so that each virtual human was equidistant from the sensor. Likewise, the virtual humans were inserted into a flat, open terrain devoid of buildings, vegetation, and other visual clutter (see **Figure 1**). In accordance with our hypotheses, the trials varied by the range at which the ring of virtual humans appeared, the density of the virtual humans, and the type of visual cue assistance the operator received. The ring of potential targets always covered a total area of 60° (30° on either side of the sensor's initial orientation), regardless of target range and density, and targets were placed throughout the entire field of regard. Consequently, scenes with virtual humans at greater ranges or greater densities inherently contained more virtual humans to search (see **Appendix A** for the number of virtual humans present at

each range by density within the wide and narrow fields of view when the sensor is at its center origin).

Since the primary dependent variable of the study was search time, it was important to carefully consider and control the placement of targets within the ring of virtual humans. For example, even with a perfect visual cue, targets that were placed near the starting azimuth would likely be found sooner than targets placed far from the starting azimuth simply due to the time it took to rotate the virtual sensor to a further target. However, repeatedly placing targets in identical positions for different trials would result in participants learning the location and azimuth of targets. As such, the location of all targets was counterbalanced across visual cue conditions, ranges, and densities so that targets appeared in roughly equivalent, but not identical, locations. This was accomplished by creating eight target location “windows” within the rings that were 3° wide, then randomly designating one of the virtual human locations within that window to be replaced with the target. This method avoided target locations being learned by participants using the sensor’s azimuth heading. The target location windows were centered at 6°, 9°, 12°, and 15° to the left and right of the sensor’s starting origin. For example, for a target location window centered around 6°, the target was randomly placed between 4.5°-7.5°. Given unlimited time and experimental trials, a target would be placed within each of these target windows for each visual cue condition, for each range, and for each density.

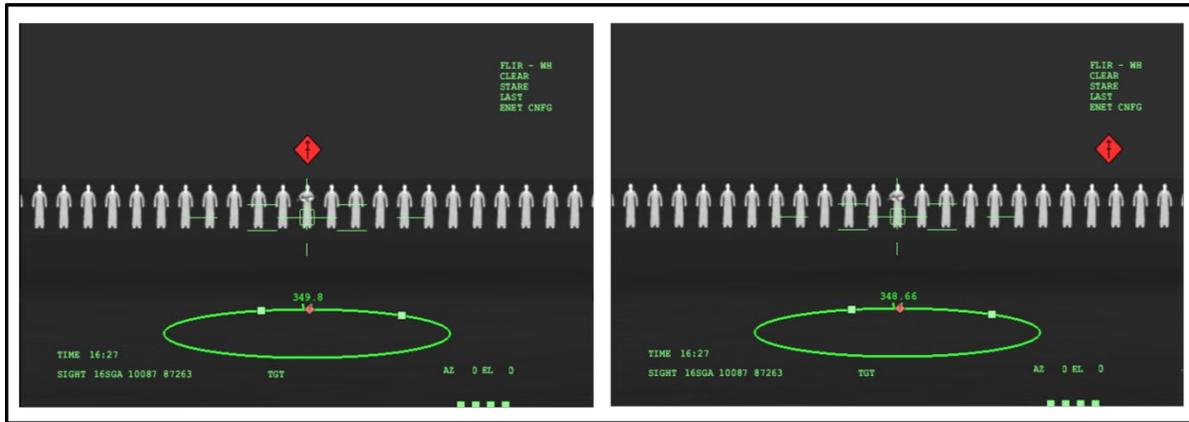
With 600 total possible experimental trials (i.e., 5 visual cue conditions × 3 ranges × 5 densities × 8 target location windows = 600 trials) there was a need to decrease the experiment length and to prevent operator fatigue. Therefore, the current experiment used 300 trials. This was accomplished by only presenting a target in either the left or right target window for a given rotational distance, rather than using both throughout the experiment. Furthermore, the 300 trials were then split in half to generate two experimental permutations, such that the 300 trials were divided across the sample of participants. For any given combination of visual cue condition, range, and density, half of the participants were presented with a target within two of the four target windows, while the other half of the participants saw the target presented within the other two target windows. As such, each participant completed a total of 150 trials.

During the experiment, each participant’s trials were subdivided into eight total blocks, with six blocks of 18 trials and two blocks of 19 trials, so that participants could periodically take breaks. Blocks were counterbalanced to contain three to four trials each of the five visual cue conditions. Each participant completed each block, and each trial within a block, in a randomized order. Participants were given a 60-second time limit to complete each trial.

It is understood that humans can react faster to visual stimuli if they are given spatial cues to inform the human of where a visual stimulus is likely to occur (Posner & Cohen, 1984). As such, our current simulation influences human target acquisition performance by both designating targets on a situational awareness ring and by placing a visual cue above the target in the operator’s field of view (see **Figure 2**) The visual cue on the situational awareness ring assists the operator in gauging which direction and how far to rotate the sensor, based on the distance between the visual cue and the sensor’s current azimuth heading. This visual cue on the situational awareness ring also acts as a pre-cue that orients a person’s attention to the visual cue that appears in the operator’s field of view once they are correctly oriented to it (Egley et al., 1994). The targeting symbol in the field of view then orients the operator’s attention to the designated target within the sensor feed. For the Imperfect Visual Cue condition, the visual cue on the situational awareness ring is presented with the same amount of fixed angular error as the visual cue designating the target within the sensor feed. While this type of visual cue symbology is not unique, limited data exists quantifying the benefits of such visual cue symbology, or how spatially accurate such symbology needs to be to improve human performance, specifically when objects are in sparse or dense scenes.

The visual cue spatial errors appeared randomly and independently of range to the target and scene density. The provided visual cue was only linked to the correct target in the sensor’s field of view, although when angular error was present, the visual cue often appeared closer to a non-target virtual human than the target. All potential targets in each scenario looked similar, and the only difference between the true target and non-targets was the presence of holding a weapon. However, researchers did not intentionally try to confuse participants by having non-targets hold objects that could be confused with weapons, nor were there additional distracting visual cues other than the single visual cue linked to the true target (i.e., confuser stimuli). In this way, researchers did not intentionally attempt to confuse participants with additional stimuli, but due to the general similarity between targets and non-targets, it was

possible for participants to incorrectly designate a non-target as a target (i.e., participant false alarm). Participants had a maximum of 60 seconds to find each target, which was used to prevent participants from spending an excessive amount of time on any given trial and to prevent operator fatigue.



**Figure 2. Sample scenes from the simulation. Each scene contained an arc of virtual humans, and participants must align the center targeting reticle with the person holding the weapon. In the left image, the visual cue correctly designates the target. In the right image, the visual cue contains angular error and is displaced slightly from the true target. Note that the visual cue appears on both the screen above the target and on the sensor’s situational awareness ring.**

### Sensor Controls and Targeting Reticle

We used highly realistic sensor grips as the human-computer interface. The controller was mounted to a stationary desktop in front of a computer monitor (see **Figure 3**). Pushing on the grips, either to the left/right or up/down, caused the sensor to rotate at a speed proportional to the strength of the push. The controller’s sensitivity was set low enough to allow operators to easily acquire the targets at the most distant range (i.e., improper sensitivity makes it difficult to acquire small, distant targets).

Participants could engage the optical zoom of the sensor, changing the sensor field of view from  $8^\circ$  to  $2.7^\circ$ , and use a “Laser Range Finder” button to designate targets. A “Menu” button was used to control a simple dialogue box that appeared after a Soldier designated a target, allowing the operator to “Confirm” or “Cancel” the designated target. A third button was used as a “speed boost” that enabled operators to rotate the sensor’s field of view faster. A different targeting reticle was displayed depending on whether the optical zoom was enabled. Each reticle included a single dot at the very center of the screen, and participants were instructed to align that targeting dot with the virtual target holding the weapon.



**Figure 3. Controller grips were mounted to the desk and positioned in front of a large, high definition 4K computer monitor.**

### Participants

Thirty U.S. Army Soldiers were recruited through Headquarters, Department of the Army. The Soldiers arrived in groups of 10 during three different sessions, participating in thermal vehicle identification training and multiple perception experiments including the current one during a week-long consolidated test event (Graybeal et al., 2018). Time spent in service of the military varied between participants ( $M = 4.00$ ,  $SD = 2.15$ ). Soldiers also reported an average of 3.43 years ( $SD = 2.37$ ) of experience with thermal imagery.

### Procedure

Participants were first given a group PowerPoint presentation explaining the simulation instructions and sensor controls. Participants were instructed to acquire the targets as quickly as possible. They were also told that a visual cue would be presented to help them during the target acquisition task, but that it would not always function perfectly.

Participants then completed training scenarios to learn the sensor controls and to practice acquiring targets. The training consisted of three trials at each of the three ranges and for three visual cue conditions: No Visual Cue, Perfect Visual Cue, and Imperfect Visual Cue with  $5^\circ$  of angular error (27 trials total). For each of the training trials, participants experienced one of three densities (0.5 m, 5 m, and 20 m). These three visual cue conditions crossed by three densities were selected because they covered the full range of visual cue performance in the experimental trials.

Once participants completed the training, they began the experiment. While participants could always take a break between any of the eight blocks of trials, they were asked to take a ten-minute break halfway through the experiment to alleviate fatigue. The instructions, training, and experiment took an approximate total of two hours.

### Data Analysis

Data analysis was similar to the previous experiment, and the parallels between analyses will be reiterated in this manuscript for the reader's convenience. The *R Project* statistical analysis software was used to conduct all data analysis. Linear mixed-effects models (Bates et al., 2015) were used to analyze human performance data (target acquisition time and target acquisition accuracy), using Satterthwaite's method of approximating degrees of freedom for the calculation of  $t$  and  $p$  values (Satterthwaite, 1946). Likewise, nested-model comparisons were used to produce interpretable main effects (due to the presence of categorical variables with more than two levels in the primary regression analyses). Two regressions were planned per dependent variable to answer our primary research questions: the first comparing all Visual Cue Conditions to the No Visual Cue Condition and the second comparing

all Imperfect Visual Cue Conditions to the Perfect Visual Cue Condition. To test main effects of range and density, only one regression was planned (each) as these variables were treated as continuous rather than categorical. To avoid the possibility of a three-way interaction, interaction effects were tested by first fitting two separate regression models: one with the range and visual cue condition variables and another with density and visual cue condition variables. The data was then segmented by range and the regression models were fitted with the angular error condition and density variables independently for each of the three ranges.

As target acquisition accuracy is a binary variable, logistic regression was used to analyze it. Target acquisition accuracy was calculated purely in terms of angular error between the true target, the sensor, and the target designation pathway through three-dimensional space indicated by the participant. Vertical accuracy was ignored. A response was scored as correct if the designated path through three-dimensional space was no more than  $1.35^\circ$  to the left or right of the target (i.e., the designated path and target were both within the sensor's  $2.7^\circ$  narrow field of view). While this is a relatively generous method of scoring accuracy, the participant's ability to finely control the grips and center the targeting reticle precisely on the target was not pertinent to our underlying research questions. As there were no confuser stimuli present in this experiment, the researchers inferred that if a designation was made while the target was within the sensor's field of view, the participant intended to designate that target.

## Outliers

Before analyzing the data, each participant's accuracy data was calculated (overall and for each visual cue condition) and checked for outliers. One participant's data was removed from all analyses because their accuracy scores fell three standard deviations below the group average. Therefore, the final sample size was 29.

## RESULTS

### Target Acquisition Time

A nested model comparison revealed a significant main effect of visual cue condition,  $\chi^2(4, N = 29) = 620.42, p < .001$ . Replicating the previous results, target acquisition times were the fastest in the Perfect Visual Cue Condition, increased with greater amounts of angular error, and were slowest in the No Visual Cue Condition (see **Table 1**). Variance in target acquisition times also increased as angular error increased. A linear mixed-effects model revealed that, compared to the No Visual Cue Condition, all visual cue information significantly improved target acquisition times (all  $p$ -values  $< .001$ ). A second linear mixed-effects model using the Perfect Visual Cue Condition as the reference group revealed that  $1^\circ$  ( $\beta = 0.08, p = .065$ ) of angular error did not significantly differ from the Perfect Visual Cue Condition, but  $3^\circ$  ( $\beta = 0.26, p < .001$ ) and  $5^\circ$  ( $\beta = 0.34, p < .001$ ) of angular error resulted in significant increases in target acquisition time.

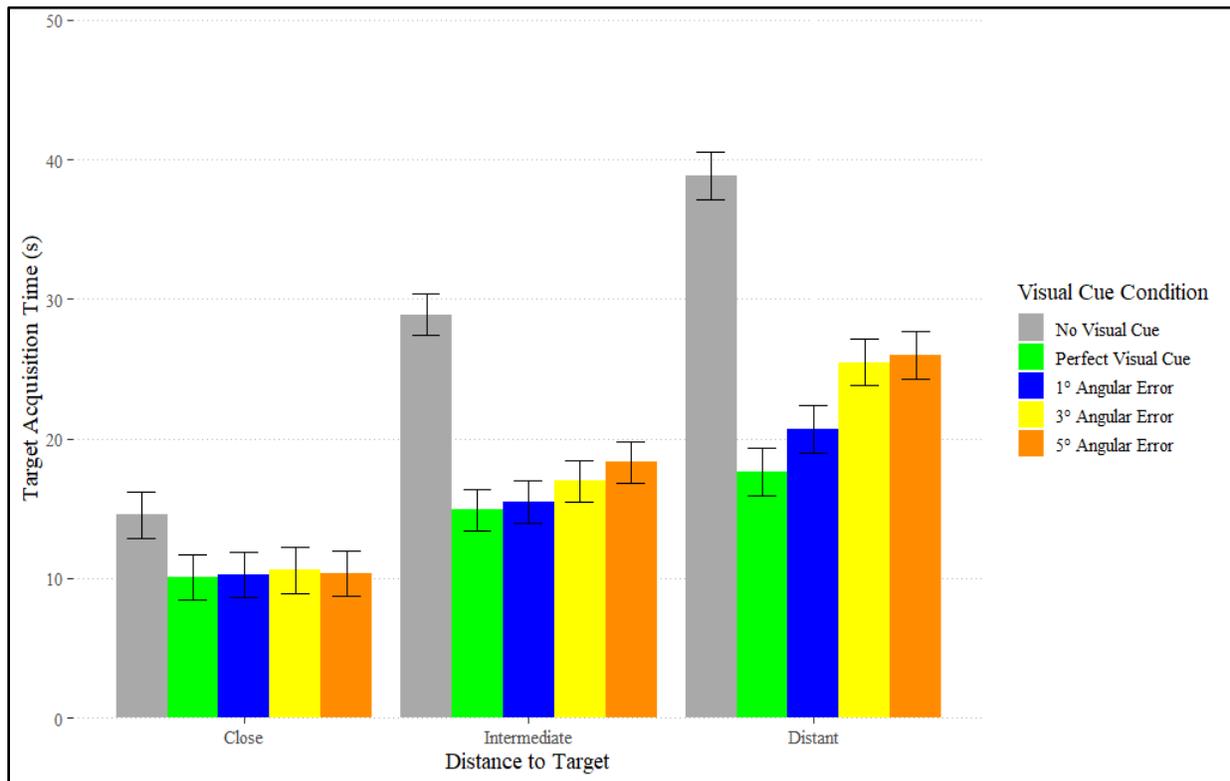
Likewise, a nested model comparison revealed a significant main effect of range on target acquisition time,  $\chi^2(1, N = 29) = 1134.8, p < .001$ , such that target acquisition time increased at longer ranges, as predicted (Close:  $M = 11.20$  s,  $SD = 5.33$  s; Intermediate:  $M = 18.86$  s,  $SD = 10.70$  s; Distant:  $M = 25.47$  s,  $SD = 13.82$  s). Additionally, there was a significant interaction effect between range and visual cue condition (see **Figure 4**),  $\chi^2(4, N=29) = 286.53, p < .001$ . Visual cues provided greater improvements to target acquisition time as the task became more difficult at more distant ranges (compared to the No Visual Cue Condition). This was consistent across all visual cue conditions (all  $p$ -values  $< .001$ ). **Table 1** shows the difference in mean target acquisition times between Imperfect Visual Cue Conditions and the No Visual Cue/Perfect Visual Cue Conditions at the three ranges, averaging across densities; it demonstrates how even imperfect visual cue assistance dramatically improved average target acquisition performance relative to the No Visual Cue Condition. Both the benefits of having assistance and the detriments of angular error became more impactful at longer ranges, such that target acquisition time was reduced by 18.14 seconds with  $1^\circ$  of angular error at the Distant range (compared to the No Visual Cue Condition), versus a smaller reduction of 12.86 seconds with  $5^\circ$  of angular error. Contrastingly, compared to the Perfect Visual Cue Condition, target acquisition time *increased* with any degree of angular error, and these increases were magnified as range to the target increased.

**Table 1. Difference in Target Acquisition Times Between Imperfect Visual Cue Information and Control Conditions**

	Compared to No Visual Cue (s)			Compared to Perfect Visual Cue (s)		
	<u>Close</u>	<u>Intermediate</u>	<u>Distant</u>	<u>Close</u>	<u>Intermediate</u>	<u>Distant</u>
1° Angular Error	-4.25***	-13.39 ***	-18.14 ***	0.20	0.58	3.09 **
3° Angular Error	-3.94***	-11.91 ***	-13.37 ***	0.51	2.06	7.86 ***
5° Angular Error	-4.17 ***	-10.55 ***	-12.86 ***	0.29	3.42***	8.36 ***

\*\* $p < .01$

\*\*\* $p < .001$



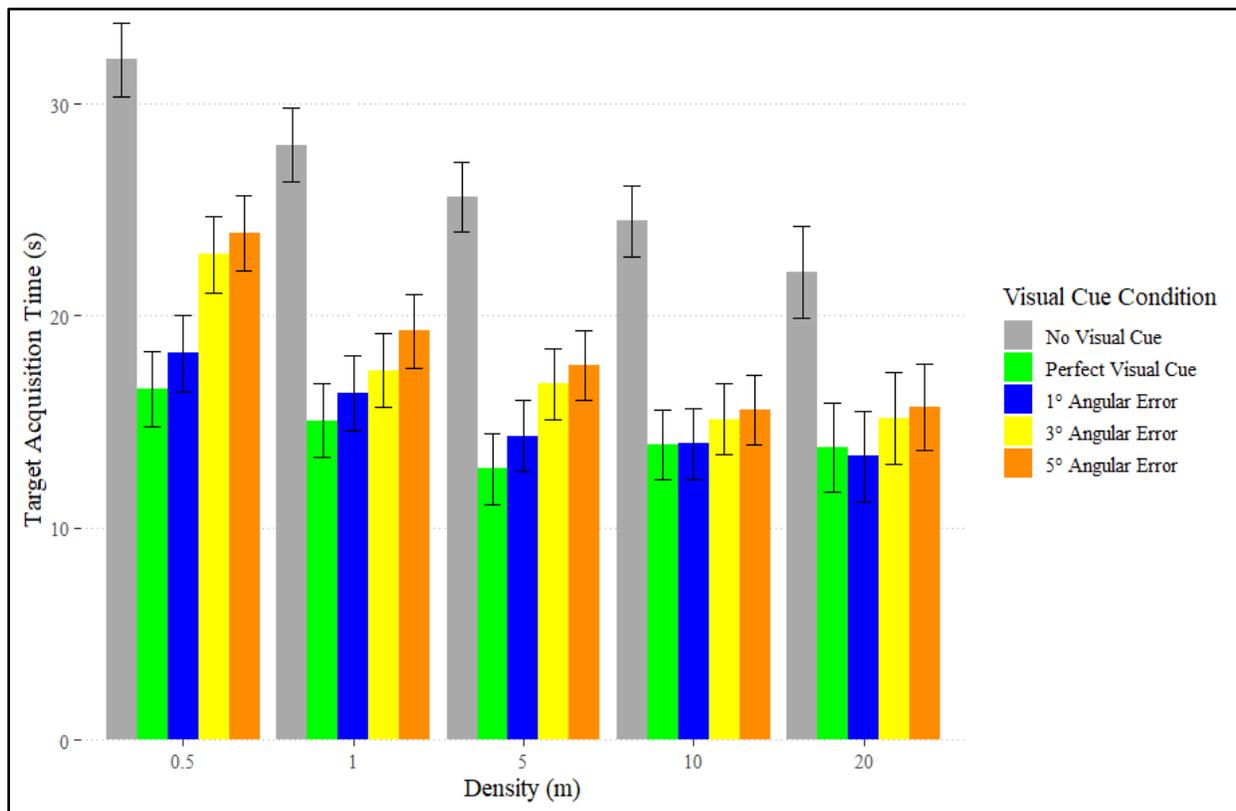
**Figure 4. Target acquisition time by range and visual cue condition. Increased range (i.e., distance to the target) and increased amounts of angular error increased target acquisition times. Error bars represent 95% confidence intervals (based on standard error estimates calculated in the linear mixed-effected models).**

Novel to the current study, a third nested model comparison testing the main effect of scene density also revealed a significant effect,  $X^2(1, N = 29) = 119.51, p < .001$ . Target acquisition times were the fastest at the lowest scene density, increased with greater amounts of scene density, and were slowest when the scene was the densest (see **Table 2**). As with angular error, variance in target acquisition times also increased as the scene became denser.

**Table 2. Average Target Acquisition Times by Scene Density Condition.**

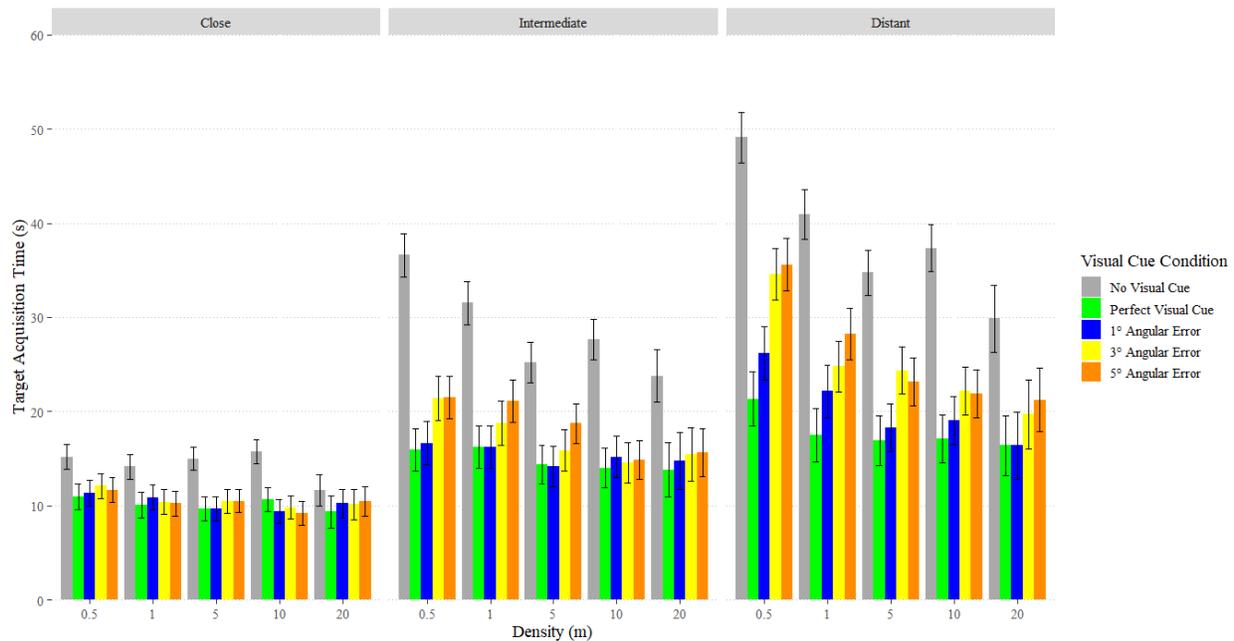
Scene Density (m)	Mean (s)	Standard Deviation (s)
0.50	23.02	15.00
1.00	19.34	12.27
5.00	17.56	11.28
10.00	16.12	9.76
20.00	15.96	9.27

There was also a significant interaction effect between density and visual cue condition (see **Figure 5**),  $X^2(4, N=29) = 19.57, p < .001$ , such that the improvements to target acquisition time that visual cue information provided, compared to the No Visual Cue Condition, were magnified as scene density increased. However, this was only true for the Perfect Visual Cue Condition and  $1^\circ$  angular error ( $p < .001$  and  $p = .008$ , respectively). These increased benefits of visual cues in denser scenes were not observed when visual cues were presented at  $3^\circ$  and  $5^\circ$  of angular error ( $p = .085$  and  $p = .182$ , respectively).



**Figure 5. Target acquisition time by density and visual cue condition. Increased density and increased amounts of angular error increased target acquisition times. Error bars represent 95% confidence intervals (based on standard error estimates calculated in the linear mixed-effects models).**

When segmenting the data by range, density had a significant effect on observed target acquisition times with visual cues (compared to the No Visual Cue Condition) at Intermediate and Distant ranges ( $p$ -values  $< .001$ ), but not at the Close range ( $p = .570$ ) (see **Figure 6**). That is, for Intermediate and Distant ranges, the magnitude of the reduction in target acquisition time between the No Visual Cue Condition and most of the angular error conditions significantly increased as the density of the scene increased. For instance, at the furthest distance, the magnitude of improvement in target acquisition time *without* visual cue assistance as scene density increased (49.11 seconds in the densest condition versus 29.87 seconds in the sparsest condition) was significantly greater than with Perfect Visual Cue assistance (21.33 seconds in the densest condition versus 16.38 seconds in the sparsest condition). See **Appendix B** for the full table of interaction effects.



**Figure 6. Target acquisition time by visual cue condition and scene density at each of the three range categories.**

### Target Acquisition Accuracy

There were two reasons why participants could fail to correctly acquire a target: they could either incorrectly acquire a target or fail to find the target within the allotted time. Considering only the trials where participants designated a target within the 60 second time limit, accuracy was extremely high for all participants at all three ranges. Participants achieved 99.85% accuracy at the Close range, 99.70% accuracy at the Intermediate range, and 98.70% accuracy at the Distant range. This indicates that the majority of inaccurate responses were due to participants failing to find a target within the allotted time, rather than the ability to accurately designate the target in general (i.e., errors in correctly aligning the target reticle or mistakenly designating incorrect targets). 105 out of 127 total trials that were scored as inaccurate occurred due to timing out, representing 82.68% of the “inaccurate” data. Consequently, accuracy in this task primarily reflects search time, rather than an ability to correctly identify the targets. The high accuracy may also partially reflect the relatively generous scoring criteria used in this experiment, (i.e., if both the target and the participant’s designation fell within the sensor’s narrow field of view, the designation was scored as correct).

A nested model comparison revealed a significant main effect of visual cue condition,  $X^2(1, N = 29) = 168.69, p < .001$ . Accuracy was lowest in the No Visual Cue Condition (88.56%), highest in the Perfect Visual Cue Condition (99.65%), and incrementally decreased as angular error increased (see **Table 3**). A second nested model comparison revealed a significant main effect of range on target accuracy,  $X^2(1, N = 29) = 145.23, p < .001$ . Accuracy was nearly perfect (99.86%) at the Close range, while progressively lower at the Intermediate range (98.46%) and the Distant range (92.23%). Finally, a third nested model comparison revealed a significant main effect of density,  $X^2(1, N = 29) = 61.91, p < .001$ , such that accuracy was lowest in the densest scenes (91.47%), and generally increased incrementally as scenes became less dense (see **Table 4**).

**Table 3. Mean Accuracy Scores by Visual Cue Condition (All Ranges Combined)**

Visual Cue Error Condition	Mean (%)	Standard Deviation (%)
Perfect Visual Cues	99.65	5.69
1° Angular Error	99.50	6.68
3° Angular Error	97.96	10.86
5° Angular Error	97.94	10.53
No Visual Cues	88.56	10.47

**Table 4. Mean Accuracy Scores by Density Condition (All Ranges Combined)**

Density Condition (m)	Mean (%)	Standard Deviation (%)
0.50	91.47	8.41
1.00	97.10	11.37
5.00	97.27	11.38
10.00	99.10	8.83
20.00	99.20	8.33

A linear mixed-effects model including visual cue condition and scene density was conducted including only the trials in the Distant range. Performance for many of the visual cue conditions was nearly perfect at both the Close and Intermediate ranges; perfect, invariant performance is problematic for regression techniques, so the Close and Intermediate ranges were excluded from further analyses of participant accuracy. With the No Visual Cue Condition as the reference group, the regression revealed significant improvement in target acquisition accuracy for all visual cue conditions (all  $p$ -values  $< .001$ ; see **Table 5**). Furthermore, density had a significant impact on participant accuracy in the No Visual Cue Condition, ( $p < .001$ ), such that participants were least accurate in the densest scenes (39.08%), most accurate in the least dense scenes (92.59%), and accuracy increased as scene density decreased (see **Table 6**). No interactions between visual cue conditions and density were significant (all  $p$ -values  $> .658$ ); performance was generally very high in all visual cue conditions regardless of scene density. A second regression, comparing performance on imperfect visual cue trials to perfect visual cue assistance was not conducted due to the low amount of variance in the Perfect Visual Cue Condition.

**Table 5. Mean Accuracy Scores by Visual Cue Condition at the Distant Range**

Visual Cue Error Condition	Mean (%)	Standard Deviation (%)
Perfect Visual Cues	99.30	2.62
1° Angular Error	98.63	4.13
3° Angular Error	94.87	7.61
5° Angular Error	94.50	7.22
No Visual Cues	70.58	19.70

**Table 6. Mean Accuracy Scores by Density Condition at the Distant Range for Targets with No Visual Cue**

Density Condition (m)	Mean (%)	Standard Deviation (%)
0.50	39.08	38.65
1.00	72.41	34.29
5.00	77.98	28.34
10.00	85.71	35.63
20.00	92.59	22.80

## DISCUSSION

### Summary

The U.S. Army DEVCOM C5ISR Center conducted an experiment that investigated the additional impact that scene density has on the effects of visual cue angular error as a function of range. The results of our previous experiment demonstrated that increasing amounts of visual cue error led to progressive detriments in target acquisition time and that visual cue accuracy requirements increase as the task becomes more difficult, such as by means of increasing range (Graybeal et al., 2019b). Thus, the current experiment expanded on our previous work by examining a second element of task difficulty: the density of objects in the scene that must be searched. We hypothesized that increasing the density of objects in the scene would increase the difficulty of the search task and subsequently alter the quality of information required to improve human task performance.

The current experiment largely replicated results from the previous experiment with respect to how range, angular error, and their interaction impacted human performance (Graybeal et al., 2019b). The current study results re-affirmed that increased range to target contributes to overall task difficulty. Additionally, perfect visual cue guidance provides more benefits to human operators as the task becomes more difficult (i.e., as range increases), but that the reduction in benefits due to angular error is more pronounced with higher task difficulty.

Likewise, the current experiment provided additional insight into a novel independent variable: the density of objects in the scene. The density of objects in the scene is a second variable that contributes to overall task difficulty, and the results of this study demonstrate it also impacted target acquisition performance and the accuracy requirements of visual cues in a similar way. Primarily, across error conditions and ranges, target acquisition time increased as density increased. We further found that scene density had a significant impact on the observed gains in human performance when completing the task with and without visual cue assistance. Specifically, target acquisition time improvements between visually cued and uncued targets were magnified as scene density increased, such that visual cue assistance, even imperfect assistance, progressively benefitted performance. Additionally, and consistent with other studies (Kruijff et al., 2010; Chen et al., 2021), increasing amounts of visual cue error led to progressive detriments in target acquisition performance at all tested scene densities, with performance generally declining as scenes got denser.

Finally, the current experiment demonstrated that visual cue accuracy requirements changed with increasing task difficulty when considering both range and density simultaneously. Although scene density did not have a large impact on performance at the Close range, as the task became more difficult, the impact of scene density became more pronounced. Target acquisition time deteriorated as scene density, angular error, and range all increased, such that performance was the worst when scenes were the densest and at the Distant range with no visual cue assistance. This finding demonstrates that, although the operator has more to gain as task difficulty also increases, the cost of system errors also increases. As such, this study joins the growing body of literature examining the impact of AR errors on human performance with similar trends (e.g., Brunyé et al., 2016; Graybeal & Du Bosq, 2018; Monfort et al., 2018).

The finding that altering two independent dimensions of task difficulty resulted in a similar pattern of results, and that

this pattern of results held true when considering both independent variables simultaneously, suggests that this pattern of results is likely to be observed with other dimensions of task difficulty. Similarly, it provides confidence that these laboratory trends will generalize to operational conditions where many factors influence the difficulty of the operator's task. Although this study's outcomes are not surprising, they do showcase the importance of sharing such findings with the engineering community, so that designers and consumers of these systems understand that the quality of visual cue information provided to Soldiers affects their task performance. This principle is important, not only because small errors can matter, but also because it demonstrates the need to measure and define visual cue accuracy requirements for specific systems and applications.

Many different environmental factors, tasks, and situations may all play a vital role in the amount of precision necessary to improve performance with visual cues. While the present task examined multiple aspects of task difficulty in a very contrived and controlled manner, the impact was clear – as the task difficulty increases, in this case by increasing scene density and range, the impacts on search time can be partially mitigated if the visual cue accuracy is sufficient. The effect that these variables may have in real-world scenarios could be even more magnified, particularly in dense, cluttered environments that are inherently more complex in comparison to the relatively simplistic scenario design employed here.

Ultimately, along with gaining specific insight into target acquisition performance with visual cue assistance at varying scene densities and ranges, this study further emphasizes the importance of understanding that, for every military task aided by visual cues, there is a quality threshold for visual cue information that must be surpassed for the visual cue information to provide quantifiable improvements to human performance. Our simulation only explores the spatial accuracy component of visual cue information; however, many other factors may affect the quality of the visual cue information and, in turn, human performance.

### **Limitations & Future Work**

All experiments have limitations, and one specific limitation of the experiment is related to its use of a highly realistic sensor controller to increase scenario realism and operational relevancy. Unfortunately, this realistic sensor controller may have led to slower and less precise target designations due to the gross motor-requirements required when using the controller grips. The use of the controller grips was not essential to the current experiment's research questions, and thus led to the researchers modifying the accuracy scoring method to account for less precise target designations.

From a broader perspective, this laboratory simulation does not encompass other influences related to Soldier performance, such as stress and fatigue present in battlefield conditions in which Soldiers would potentially use AR systems. The simulated scenarios also did not contain any visual clutter outside of the core objects that had to be searched, which is also not realistic to battlefield conditions that would be considerably more complex and dynamic.

Future work could evaluate how field of regard size that an operator is responsible for impacts the difficulty of the search task and the accuracy requirements for visual cues. Future experiments could similarly examine additional variables, such as the sensor's field of view, the amount of visual clutter in the scene, the similarity of clutter objects to the targets of interest, and other variables that may impact the difficulty of the search task. In addition to investigating other impactful variables, display modality is another avenue of research worth exploring. With an increased prevalence of head-mounted displays, replicating these experiments using virtual reality head-mounted displays could provide insight into how different devices and display modalities may alter the impact of visual cues and other significant variables on target acquisition performance. Finally, although difficult to replicate in a laboratory setting, future experiments could assess the impact of fatigue, high stress, or sleep deprivation, as many of these variables (among others) may influence how visual cues are used and may be more reflective of the extreme situations in which Soldiers may be required to use such technologies.

## CONCLUSION

This paper is the second in a series of experiments examining how visual cues, both perfect and imperfect, impact target acquisition performance. Specifically, this experiment focused on the effect scene density, degree of angular error, and range to target had on acquisition performance. Visual cues, even imperfect, aided target acquisition performance across all scene densities. Furthermore, as scene density increased, the benefits of visual cues increased, but errors in the spatial cues also had more of an impact. These findings mirror how range impacted human performance in both the present and previous study. Overall, manipulating two independent dimensions of task difficulty resulted in clear behavioral patterns, and these patterns persisted when considering both dimensions of task difficulty simultaneously. Our results demonstrate that as the task difficulty (i.e., scene density and range to target) increases, visual cue accuracy requirements change as the operator's performance can be better enhanced by visual cues, but spatial errors also become more damaging. Collectively, the consistency of the results suggests that the observed trends are robust and likely to generalize to complicated operational considerations where many factors contribute to task difficulty.

## REFERENCES

- Bates, D., Maechler, M., Bolker, B., Walker, S., Christensen, R. H. B., & Singmann, H. (2015). lme4: Linear mixed-effects models using Eigen and S4, 2014. *CRAN*, 1(4).
- Brunyé, T. T., Moran, J. M., Houck, L. A., Taylor, H. A., & Mahoney, C. R. (2016). Registration errors in beacon-based navigation guidance systems: Influences on path efficiency and user reliance. *International Journal of Human-Computer Studies*, 96, 1-11.
- Chen, L., Liu, Y., Li, Y., Yu, L., Gao, B., Caon, M., ... & Liang, H. N. (2021). Effect of visual cues on pointing tasks in co-located augmented reality collaboration. In *Proceedings of the 2021 ACM Symposium on Spatial User Interaction* (pp. 1-12).
- Dixon, B. J., Daly, M. J., Chan, H., Vescan, A. D., Witterick, I. J., & Irish, J. C. (2013). Surgeons blinded by enhanced navigation: the effect of augmented reality on attention. *Surgical endoscopy*, 27(2), 454-461.
- Dunleavy, M., & Dede, C. (2014). Augmented reality teaching and learning. In M. Spector, M. Merrill, J. Ellen & M. Bishop (Eds.), *Handbook of research on educational communications and technology* (pp. 735-745). New York, NY: Springer.
- Egley, R., Driver, J., & Rafal, R. D. (1994). Shifting visual attention between objects and locations: evidence from normal and parietal lesion subjects. *Journal of Experimental Psychology: General*, 123(2), 161.
- Graybeal, J. J., Monfort, S. S., Du Bosq, T. W., & Familoni, B. O. (2018). Human perception testing methodology for evaluating EO/IR imaging systems. In *Infrared Imaging Systems: Design, Analysis, Modeling, and Testing XXIX (Vol. 10625, p. 106250T)*. International Society for Optics and Photonics.
- Graybeal, J. J., Nguyen, R. T., & Du Bosq, T. W. (2019). Simulating human vehicle identification performance with infrared imagery and augmented reality assistance. In *Infrared Imaging Systems: Design, Analysis, Modeling, and Testing XXX (Vol. 11001, pp. 116-129)*. SPIE.
- Graybeal, J. J., Nguyen, R. T. T., Du Bosq, T. W. (2019). Simulating Augmented Reality Spatial Accuracy Requirements for Target Acquisition Tasks. In *Proc. of Interservice/Industry Training, Simulation and Education Conference*.
- Graybeal, J. J., Sharp, W. H., Lasko, E., Bagalkotkar, K., & Dang, J. (2023). Simulating the Impact of Navigation Waypoint Accuracy on Solider Performance. In *Proc. of NATO MSG-207*.
- Harkrider, S., May, C., Paul, D., Osborn, W., Tran, V., & Anderson, D. (2011). A Successful Government Furnished Solution Use Case: The Night Vision Image Generator. In *2011 Meeting of the Military Sensing Symposia (MSS) Specialty Group on Passive Sensors, Battlefield Survivability & Discrimination, Detectors and Materials*.
- Kruijff, E., Swan, J. E., & Feiner, S. (2010). Perceptual issues in augmented reality revisited. In *2010 IEEE International Symposium on Mixed and Augmented Reality* (pp. 3-12). IEEE.
- May, C., Harkrider, S., Paul, D., Osborn, W., & Tran, V. (2011) Target IR Signature Creation Used in NVIG Sensor Simulation. In *Proc. SimTect*, 31.
- Monfort, S., Graybeal, J. & de Visser, E. & Du Bosq, T. (2018). Reliability requirements for augmented reality in visual search tasks. In *Proc. of Interservice/Industry Training, Simulation and Education Conference*.
- Posner, M. I. & Cohen, Y. (1984). Components of visual orienting. *Attention and Performance*, 32, 531-556.
- Radu, I. (2014). Augmented reality in education: a meta-review and cross-media analysis. *Personal and Ubiquitous Computing*, 18(6), 1533-1543.
- Satterthwaite, F. E. (1946). An approximate distribution of estimates of variance components. *Biometrics bulletin*, 2(6), 110-114.
- Tang, A., Owen, C., Biocca, F., & Mou, W. (2003, April). Comparative effectiveness of augmented reality in object assembly. *Paper presented at the meeting of CHI*, Ft. Lauderdale, Florida.
- Wu, H. K., Lee, S. W. Y., Chang, H. Y., & Liang, J. C. (2013). Current status, opportunities and challenges of augmented reality in education. *Computers & Education*, 62, 41-49.

**APPENDIX A**

**Number of Virtual Humans in 8° Field of View at each Density Level**

Range	0.5 m	1 m	5 m	10 m	20 m
Close	42	21	4	2	1
Intermediate	140	70	14	7	3
Distant	279	140	28	14	7

**Number of Virtual Humans in 2.7° Field of View at each Density Level**

Range	0.5 m	1 m	5 m	10 m	20 m
Close	14	7	1	1	0
Intermediate	47	24	5	2	1
Distant	94	47	9	5	2

**APPENDIX B**

**Difference in Target Acquisition Time between the Densest and Sparsest Scenes at each Range for each Visual Cue Condition compared to the No Visual Cue Condition**

Range	Visual Cue Condition	Difference between Highest and Lowest Density compared to No Visual Cue Condition (s)
Close	Perfect	-1.99
	1°	-2.38
	3°	-1.60
	5°	-2.35
Intermediate	Perfect	<b>-10.70***</b>
	1°	<b>-10.96***</b>
	3°	<b>-6.87**</b>
	5°	-6.93
Distant	Perfect	<b>-14.29***</b>
	1°	<b>-9.42*</b>
	3°	-4.43
	5°	-4.89

\* $p < .05$

\*\* $p < .01$

\*\*\* $p < .001$