Drone-based Augmented Reality Platform for Bridge Inspection: Effect of AR Cue Design on Visual Search Tasks

Jared Van Dam*
Virginia Tech

Alexander Krasner
Virginia Tech

Joseph L. Gabbard
Virginia Tech

Abstract

The use of augmented reality (AR) with semi-autonomous aerial systems in civil infrastructure inspection offers an extension of human capabilities by enhancing their ability to access hard-to-reach areas, decreasing the physical requirements needed to complete the task, and augmenting their visual field of view with useful information. Still unknown though is how helpful AR visual aids may be when they are imperfect and provide the user with erroneous data. A total of 28 participants flew as an autonomous drone around a simulated bridge in a virtual reality environment and participated in a target detection task. In this study, we analyze the effect of AR cue type across discrete levels of target saliency by measuring performance in a signal detection task. Results showed significant differences in false alarm rates in the different target salience conditions but no significant differences across AR cue types (none, bounding box, corner-bound box, and outline) in terms of hits and misses.

Keywords: Augmented reality, signal detection, infrastructure inspection, workload, unmanned aerial system.

Index Terms: Multimedia information systems—artificial, augmented, and virtual realities, User interfaces—ergonomics, evaluation/ methodology, screen design, style guides

1 Introduction

Civil infrastructure (e.g., roads, dams, bridges, buildings, etc.) inspection is a labor intensive and arduous process that demands that the inspectors visually analyze and annotate each area of the structure. This process is both time consuming and sometimes dangerous, requiring heavy equipment to allow workers to visually inspect hard-to-reach areas. Given the sheer number of structures that require inspection and the requirement that these inspections occur once every two years [1], there arises a need for timely and remote inspection solutions that provide the operator with enough information to make correct decisions but allow them to conduct inspections remotely. The use of unmanned aerial systems (UASs) in this space provides a means to access, with greater ease, these hard-to-reach areas of bridges and other forms of civil infrastructure. In addition, by using stereoscopic depth cameras on the UAS paired with computer vision algorithms and video pass-through augmented reality (AR), we can detect defects on the bridge surface and relay those issues to the operator while augmenting the video stream with useful information (e.g., world-fixed rulers and annotations) that can aid the operator in their tasks. The use of these UAS-mounted camera systems, in addition to the added augmentations, would help operators beyond what a simple monoscopic camera would provide by adding the ability to display relevant information to the operator while providing them with the stereoscopic visual information needed for accurate depth perception. It is still unknown though what cognitive costs, or performance decrements, may be associated with these augmentation interface elements.

To test these possible human performance decrements and for ease of rapid prototyping and interface development, we created a virtual testbed (Figure 1) for the present study that allows us to simulate video see-through AR in virtual reality [2]. With this testbed, we can examine how different AR user interface designs impact users during bridge inspection tasks without the associated costs of real-world testing as the testing of these designs in the real-world often incurs a great deal of expense in both time and energy.

1.1 Use of unmanned aerial systems in infrastructure inspection

In general, the use of UASs in infrastructure inspections falls into the categories of construction monitoring and condition assessment, where the level of autonomy of the UAS is either autonomous or manual [3]. While bridge inspectors currently use UASs to inspect bridges, these use cases often consist of simply taking video of the bridges and analyzing the video offsite later [1]. One issue with this practice is that an expert analyzing the recorded video would be unable to take a closer look at certain sections of bridge areas, since the video has already been captured and does not afford, without significant quality loss, the ability to zoom in and traverse the bridge. These videos are also usually captured with monoscopic cameras which limits the depth perception of humans and does not easily afford complex AR overlays. In our current approach, we propose the use of stereo depth cameras that allow for video pass-through AR so that we can display useful augmentations (e.g., annotations or measuring tools) to the operator in real-time so the inspector can analyze potential problem areas and probe further if they need to see more detail.

The use of UASs in infrastructure inspections allows the inspectors to be virtually placed anywhere on the bridge. Additionally, AR interface elements can be placed that show thermal overlays [3, 4], structural stresses [5], exterior deformation or plumbness [6], and measurements of structural elements [7]. However, the use of UASs and AR in infrastructure inspection does not come without its issues. In general, the complexity of the task may impact the usefulness of the AR interface. For example, the use of AR cuing during complex tasks can be helpful, but in easy tasks, the cuing may provide no help or may even harm performance [8]. Cuing that can harm performance would be especially impactful with infrastructure inspection, where the misidentification or miss of a defect could lead to catastrophic failure if not caught in an adequate timeframe.

1.2 Signal detection with augmented reality

AR can direct a user's attention toward real-world referents (e.g., a bridge defect), even when those targets lie in the user's peripheral field of view [9]. The SEEV model, initially used in the aviation

*Jaredvd@vt.edu

978-1-7281-6532-5/20/$31.00 ©2020 IEEE
DOI 10.1109/VRW50115.2020.00043

domain, can explain in part the perceptual and cognitive elements at play when assessing whether an AR cue may or may not demand a user's attention. In this model, the scanning pattern used by individuals is determined by the factors of saliency, effort, expectancy, and value [10]. AR-enhanced manipulations like changing the shape, motion, and color of objects can increase the saliency of those objects [11]. By manipulating the saliency of objects, AR interface designers can direct user's visual attention to a specific target (e.g., a bridge defect), making detection quicker and easier. The use of subtle AR cuing can also draw users' attention without the extra clutter that comes with augmenting the environment using typical bounding boxes [12]. While these AR cues generally aid users in target detection tasks (looking for a target in an environment of noise [12]), cue reliability and target expectation can also mediate cue effectiveness [13].

2 EMPIRICAL USER STUDY

2.1 Study purpose and hypothesis

Our study aims to extend previous target detection work to the virtual reality space specific to infrastructure inspection to determine the effect of AR cue type and target saliency on operators' signal detection rates. Additionally, signal detection rates with different AR cue types will be measured to see if feature-bound AR cues afford similar detection rates as compared to fully bound and corner bound AR cues while also granting operators a better view and understanding of the target's visible features.

In this work, we posit the following hypotheses: Signal detection (percentage of hits, misses, and false alarms) in the low target saliency trials will not be significantly different amongst the three AR cue types but will differ significantly relative to the control condition (i.e., no-AR cues). When targets are highly salient, the AR cues will hurt overall signal detection performance relative to the control condition.

2.2 Experimental design

2.2.1 Procedure

We assigned participants (n = 28, μ 27 years old, 7 Females) to experimental conditions via a Latin square where participants experienced all three levels of target saliency and all four AR cue types. Specifically, our study employed a four (AR cue type: none, bounding box, corner-bound box, and outline – see Figure 1) × three (target saliency: low (0.12), medium (0.17), and high (0.22)) within-subjects design. With this design, participants experienced 12 total trials with one repetition of each target saliency and AR cue type pair. For the duration of the experiment, participants sat in an office chair that allowed them to spin around a central vertical axis, or move around the room that measured 11.6 ft. by 12 ft. Before beginning the experiment, we instructed participants on the use of the virtual reality headset and the accompanying hand controller used to select targets via a ray casting selection technique. Participants were then given brief instructions on what the targets would look like and before performing a practice trial that used the same bridge model as the experimental trials but had targets with a lower transparency value of 0.4 (making them easier to see). In all practice and experimental trials, participants were asked to visually search for bridge defects (i.e., targets) and select them using the hand controllers. After confirming that each participant understood what the targets looked like and their task for the experiment, the experimental trials began.

2.2.2 Testbed environment

We modeled and rendered the virtual bridge in the Unity game engine and presented it to participants via a wireless HTC Vive Pro headset with a 90 Hz 1440 × 1600 pixels per eye. We used an Alienware Aurora R6 running Windows 10 with an Intel Core i7-7700 CPU and a NVIDIA GeForce GTX 1080 Ti graphics card to power the HMD. Concentrated areas of dots that were a different

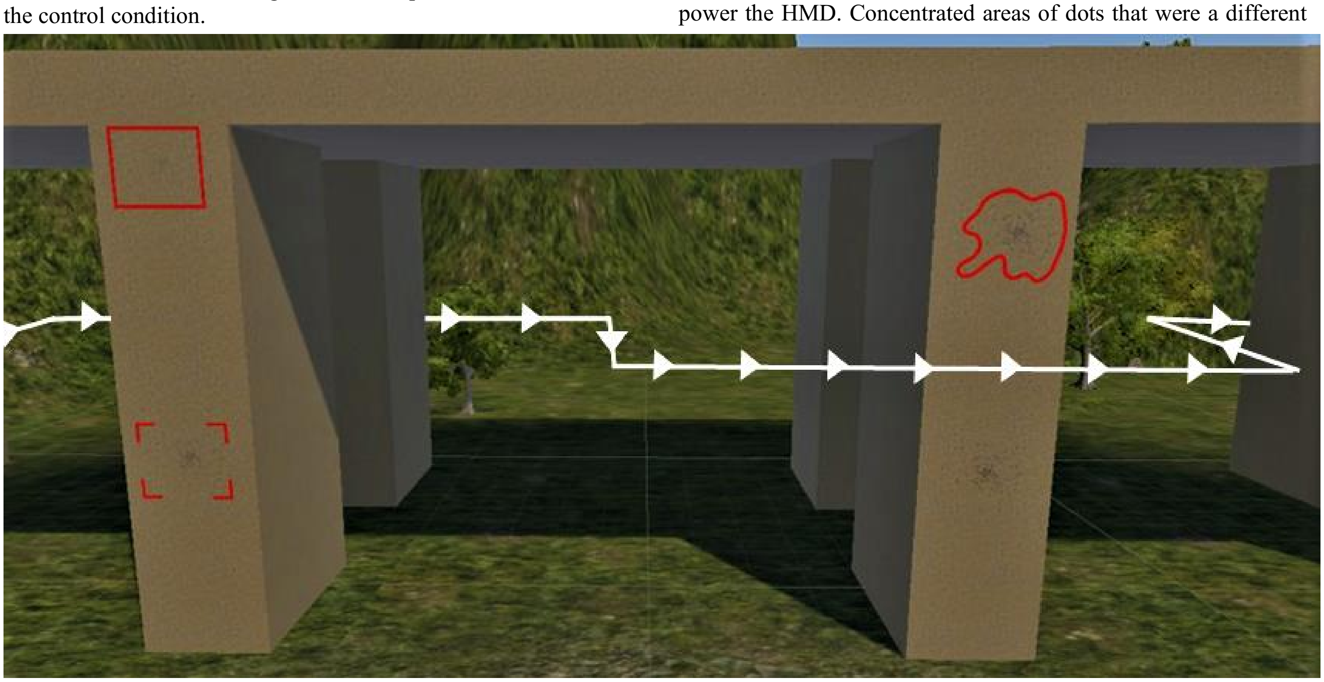

Figure 1: Simulated bridge environment with all saliency levels and augmentation types represented. The top left target is the low saliency target (0.12) with the bounding box augmentation, the bottom left target is the medium saliency target (0.17) with the corner-bound augmentation, the top right target is the high saliency target (0.22) with the outline type augmentation, and the bottom right target is the non-augmented, high saliency target. Note that this graphic is for visual representation only and that during experimental trials, only one saliency type and one augmentation type would appear for each trial.

color from the surrounding bridge area delineated the target areas (Figure 1). Each target had the same dot pattern, but flipped, or rotated randomly across the bridge so that the targets would look slightly different but still have the same global dot pattern. We adjusted saliency by adjusting the transparency values of the target area. Transparency values used were 0.12 for the low/hard condition, 0.17 for the medium condition, and 0.22 for the high/easy condition. Transparency values were determined via a pilot study in which we adjusted transparency values until participants got an average of 80% correct for the easy condition, 60% correct for the medium condition, and 30% correct for the hard condition (with no AR cues present). The program then randomly placed the targets on the bridge surface before each trial with constraints that allowed only a maximum of three targets per plane to minimize overcrowding of any individual bridge section. Additionally, governing the maximum number of targets per plane limits user misses due to a physical inability to accurately select the targets even if they could visually identify them all.

The cues used in the study were a bounding box, corner-bound box, and an outline of the target area. We used a no-augmentation condition as a control. All augmentations were red (RGB:255, 0, 0). See Figure 1 for a visual representation of the different target saliency levels and AR cue types. The virtual camera followed a fixed flightpath that simulated a typical flightpath an UAS might take as it captured video of a bridge surface. As such, participants had no control over the position of the camera, only the orientation (by pivoting in the chair and scanning with their head). The movement followed a serpentine pattern that weaved in between each bridge column and gave the participant a view of every surface plane that had targets. User inputs were classified as hits when they correctly selected a target, misses when they did not select a target area, and false alarms when they selected an area that did not contain a target. Movement speed and flightpath was the same in every condition and the total flight time of each trial was two minutes.

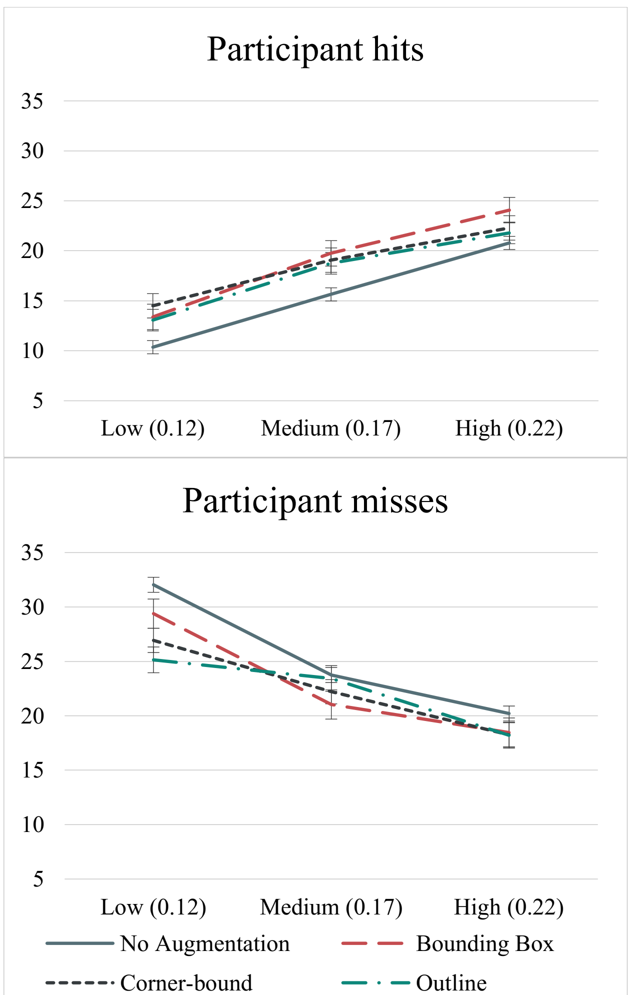

Figure 2: Group means for number of hits (top) and misses (bottom) per trial categorized by augmentation type and saliency level. Error bars represent standard error of the mean.

3 Results

We used the Winsorize technique to transform four total outliers when a data point was greater than three interquartile ranges from the upper, or lower, limit of the 25th percentile. We set the alpha at 0.05 for all statistical tests.

3.1 Hits

We used a RMANOVA with AR cue type and target saliency to determine their effects on the hit rate of subjects. Main effects were present for both AR cue type [F (3,81) = 6.94, $p < 0.001$, $\eta^2_{partial} = 0.2$] and target saliency [F (2,54) = 41.75, $p < 0.001$, $\eta^2_{partial} = 0.61$] but the interaction did not reach significance ($p = 0.49$). Significant pairwise differences were present between the control condition and bounding box cue ($p = 0.009$) as well as the control and corner bound cue ($p = 0.033$) but were not significant for the control and outline AR cue comparison ($p = 0.074$). Pairwise comparisons between the three AR cue types were not significant as all p values were greater than 0.5. Significant differences existed between the low and medium ($p < 0.001$) target saliency, low and high ($p<0.001$) target saliency, and medium and high ($p < 0.001$) target saliency. As seen in figure 2, the number of hits a participant had varied significantly on AR cue type and target saliency where participants had more hits in the higher target saliency and AR cueing conditions. These results show partial support of our first hypothesis but, did not show support for the proposed interaction wherein augmentations might hurt performance in the high target salient trials.

3.2 Misses

To determine the possible impact that augmentation type and saliency has on user misses, we used a RMANOVA. Main effects of both target saliency [F (2,54) = 35.63, $p < 0.001$, $\eta^2_{partial} = 0.57$] and AR cue type [F (2.3,62.14) = 4.20, $p < 0.05$, $\eta^2_{partial} = 0.14$] were present using the Huynh-Feldt correction as well as the interaction [F (5.53,149.32) = 2.56, $p = 0.025$, $\eta^2_{partial} = 0.09$]. While mean misses for the control conditions was higher relative to the bounding box, corner bounded box, and outline, none of the pairwise comparisons were significant ($p > 0.09$). Pairwise comparisons with the three target saliency conditions though indicated significant differences between low and medium ($p < 0.001$), low and high ($p < 0.001$), and medium and high ($p = 0.002$). As shown in figure 2, results indicate partial support for our hypothesis as performance did differ significantly based on target saliency but, in contradiction to the hypothesis. significant differences between the control condition and the AR cue conditions did not exist.

3.3 False Alarms

We used a RMANOVA to test the effect of AR cue type and target saliency on the number of false alarms. Significant main effects were present for target saliency [F (2,54) = 38.59, p = 0.02, $\eta^2_{partial}$ = 0.14], but not for AR cue type (p = 0.33) or the interaction of AR cue type and target saliency (p = 0.14) using the Huynh-Feldt correction. Pairwise comparisons within the AR cue type condition led to no significant differences with all p values greater than 0.7 for each combination. Pairwise comparisons between low and medium (p = 0.083) target saliency as well as medium and high (p = 1) levels of target saliency did not reach significance. A significant pairwise difference did exist though when comparing low to high saliency (p = 0.03). With this comparison, participants had significantly more false alarms in the high saliency (0.22) condition when compared to the low saliency (0.12) condition.

4 Discussion

Our hypothesis was partially supported in that performance on the signal detection task was generally better in the AR cue conditions relative to the control conditions, but the proposed interaction where AR cues would hurt performance in the highly salient target conditions was not supported except in the case of false alarms where participants had more false alarms in the high target salience conditions.

Given the work of Maltz and Shinar [8] who found that AR cuing was generally helpful during complex tasks, but hurtful during easy tasks, and the finding of the present study where users had more false alarms during the high target salience trials, care should be taken to only use AR cuing systems during complex tasks or those tasks where user misses carry significant weight. Future research should also examine this area further by studying AR cuing systems that adjust cue saliency to the target difficulty so that, when target saliency is low and the target is difficult to detect, the AR cue is very salient but, when target saliency is high and easy to detect, the AR cue saliency is either low and unobtrusive, or not present.

The results of this study show that the corner-bound AR cue is useful when designers need to attract the user's attention without the increased pixel density that accompanies the typical bounding box. For interface designers encountering an already cluttered visual scene, the corner-bound AR cue could be a useful addition that would still alert the operator to a target while still leaving more of the scene without augmentations. Like the corner-bound augmentation, the outline AR cue did not hinder performance beyond the other AR cue types, so these cue types could be useful in situations where the exact shape of the target area is needed, or useful for the operator to know.

5 Limitations and future work

As with any study, our experiment had limitations that limited the conclusions we could draw from the data. Thus, a possible limitation of the present study is the limited amount of time participants had with each AR cue type. With each trial only lasting two minutes, participants had little time to develop an understanding of the system. Future work analyzing the effects of different augmentation types should limit the number of analyzed AR cues to give participants ample time to become accustomed to the AR cuing aid.

Future work in this research space entails analyzing the effect that imperfect AR cuing aids have on user signal detection when the cuing aid augments objects that are near targets that the cuing aid misses. By cuing one target and not the other nearby target, the user may overlook the non-augmented target as their attention is drawn to the augmented target. With the data collected, we hope to answer this question by measuring signal detection rates and comparing performance when the aid highlights all nearby targets versus participant performance when the AR cue only highlights one of the targets when another, non-augmented target, is in close proximity. We also hope to extend this work by conducting real-world inspection with in order to determine how the present findings translate to actual infrastructure inspections.

Acknowledgments

Support for this work is provided by National Science Foundation grant #1840044.

References

[1] J. L. Wells, B. Lovelace, and T. Kalar, "Use of Unmanned Aircraft Systems for Bridge Inspections," Transp. Res. Rec. J. Transp. Res. Board, vol. 2612, no. 1, pp. 60–66, 2017.

[2] C. Merenda, C. Suga, J. Gabbard, and T. Misu, "Effects of Vehicle Simulation Visual Fidelity on Assessing Driver Performance and Behavior," *2019 IEEE Intell. Veh. Symp.*, no. Iv, pp. 1679–1686, 2019.

[3] Y. Ham, K. K. Han, J. J. Lin, and M. Golparvar-Fard, "Visual monitoring of civil infrastructure systems via camera-equipped Unmanned Aerial Vehicles (UAVs): A review of related works," Vis. Eng., vol. 4, no. 1, pp. 1–8, 2011.

[4] M. Janků, I. Březina, and J. Grošek, "Use of Infrared Thermography to Detect Defects on Concrete Bridges," Procedia Eng., vol. 190, pp. 62–69, 2017.

[5] M. Clothier and M. Bailey, "Augmented reality visualization for Kings Stormwater bridge," in IASTED International Conference on Visualization, Imaging and Image Processing (VIIP), 2004.

[6] A. H. Behzadan, S. Dong, and V. R. Kamat, "Augmented reality visualization: A review of civil infrastructure system applications," Adv. Eng. Informatics, vol. 29, no. 2, pp. 252–267, Apr. 2015.

[7] D. H. Shin and P. S. Dunston, "Evaluation of Augmented Reality in steel column inspection," Autom. Constr., vol. 18, no. 2, pp. 118–129, Mar. 2009.

[8] M. Maltz and D. Shinar, "New Alternative Methods of Analyzing Human Behavior in Cued Target Acquisition," Hum. Factors J. Hum. Factors Ergon. Soc., 2004.

[9] N. Kishishita, K. Kiyokawa, J. Orlosky, T. Mashita, H. Takemura, and E. Kruijff, "Analysing the effects of a wide field of view augmented reality display on search performance in divided attention tasks," ISMAR 2014 - IEEE Int. Symp. Mix. Augment. Real. - Sci. Technol. 2014, Proc., pp. 177–186, 2014.

[10] W. J. Horrey, C. D. Wickens, and K. P. Consalus, "Modeling drivers' visual attention allocation while interacting with in-vehicle technologies," J. Exp. Psychol. Appl., vol. 12, no. 2, pp. 67–78, 2006.

[11] L. Itti and C. Koch, "Computational modeling of visual attention," 2001.

[12] W. Lu, B. L. H. Duh, and S. Feiner, "Subtle cueing for visual search in augmented reality," ISMAR 2012 - 11th IEEE Int. Symp. Mix. Augment. Real. 2012, Sci. Technol. Pap., pp. 161–166, 2012.

[13] M. Yeh and C. D. Wickens, "Effects of cue reliability, realism, and interactivity on biases of attention and trust in augmented reality," in Proceedings of the IEA 2000/HFES 2000 Congress, 2000, pp. 511–514.