International Journal of Mobile Human Computer Interaction
Volume 11 • Issue 2 • April-June 2019

Effects of Volumetric Augmented Reality Displays on Human Depth Judgments:

Implications for Heads-Up Displays in Transportation

Lee Lisle, Virginia Tech, Blacksburg, USA
Coleman Merenda, Virginia Tech, Blacksburg, USA
Kyle Tanous, Virginia Tech, Blacksburg, USA
Hyungil Kim, Virginia Tech, Blacksburg, USA
https://orcid.org/0000-0001-9014-7484

Joseph L Gabbard, Virginia Tech, Blacksburg, USA
Doug A. Bowman, Virginia Tech, Blacksburg, USA

ABSTRACT

Many driving scenarios involve correctly perceiving road elements in depth and manually responding as appropriate. Of late, augmented reality (AR) head-up displays (HUDs) have been explored to assist drivers in identifying road elements, by using a myriad of AR interface designs that include world-fixed graphics perceptually placed in the forward driving scene. Volumetric AR HUDs purportedly offer increased accuracy of distance perception through natural presentation of oculomotor cues as compared to traditional HUDs. In this article, the authors quantify participant performance matching virtual objects to real-world counterparts at egocentric distances of 7-12 meters while using both volumetric and fixed-focal plane AR HUDs. The authors found the volumetric HUD to be associated with faster and more accurate depth judgements at far distance, and that participants performed depth judgements more quickly as the experiment progressed. The authors observed no differences between the two displays in terms of reported simulator sickness or eye strain.

KEYWORDS

Augmented Reality, Head-Up Displays, Human-Computer Interaction, Perception, Volumetric Displays

INTRODUCTION

Today's modern cars collect vast amounts of data through various sources that need to be condensed and presented to drivers in a salient format. Accordingly, there are multiple ways to present the data to drivers, from console displays to dashboard displays and, recently, augmented reality (AR) head-up dis-plays (HUDs). While console displays and dashboard displays require users to look away from the road scene, HUDs display information contextually overlaid on top and into the driving scene. Since these displays do not require the user to glance away from the scene, there is opportunity for increased performance in visual and identification tasks without impacting primary task (driving) performance (Smith et al., 2017; Rusch et al., 2013; Tran et al., 2013).

DOI: 10.4018/IJMHCI.2019040101

AR displays, however, can be impacted from several factors that impede users' visual perception. Some AR HUDs, for example, can cause eye strain (Banks et al., 2013) or simulator sickness (Kennedy et al., 1993). These effects may be overcome through using binocular focal planes which further afford more accurate perception of virtual objects. Traditionally, to address these effects AR displays employ transparent or video see-through technology utilizing a fixed-focal display. Unfortunately, this solution is affected by the vergence-accommodation mismatch (Hoffman et al., 2008) which hinders depth perception. Common driving scenarios, such as collision hazards or pedestrian detection, often involve depth perception in order to comprehend and execute a response effectively. When combined with an AR HUD, these scenarios need to assist the driver with correct depth cues in order for them to identify the distance of any hazard and react accordingly. Traditional fixed-focal plane AR HUDs have to overcome their depth perception issues in order to be a more effective display for road hazards.

Volumetric AR displays purport to improve depth perception through the use of voxels. Voxels are illuminated points in three-dimensional space that can create depth cues naturally as they occupy a true depth location in variable focal planes. This approach eliminates the need for other specialized technology like stereoscopic glasses or head-tracking systems to create three-dimensional depth cues such as motion parallax or binocular disparity (Jones et al., 2008). As such, volumetric displays afford consistent oculomotor vergence and accommodation cues that help to overcome the issues with the mismatch that occurs with many other augmented reality displays (Swan et al., 2015). Moreover, volumetric displays can support natural depth cues and increase depth perception at arbitrary distances as compared to traditional fixed-focal plane display designs. Supporting natural depth cues is especially important in driving, where for example, AR graphics should guide drivers' visual attention to hazards, and presenting AR graphics at the same depth of a hazard may increase detection and subsequent reactions.

This study evaluates a swept-volume volumetric AR HUD to better understand human performance gains against a traditional fixed-focal plane AR HUD. We focused specifically on the quality of depth judgments (both time and accuracy), speed/accuracy tradeoffs and practice effects over time. Furthermore, we collected self-reported measures of eye strain and simulator sickness to see what effect these displays have on drivers.

Our study found that depth perception does improve with volumetric AR HUD technology. This can be partially attributed to the fixed-focal display anchoring users' depth perception to a particular distance which hinders users possibly more than the variable focal planes of the volumetric display helps. With further statistical analysis, we found a weak correlation between response time and judgment as well as an auto-correlation between trial number with respect to judgment and response time. There was also a multivariate negative correlation between trial number with respect to judgment and response time with the volumetric display. Since their accuracy did not improve, this shows a measure of learning where users were quicker to perceive distances. Furthermore, this significance was not present with the traditional fixed-focal display condition which shows that users were specifically more confident with the volumetric display for depth judgments.

These improvements in display technology indicate that volumetric displays can be effective for AR-based driver human machine interfaces (HMI). With the multiple focal planes afforded by volumetric displays, drivers have a means for more accurate depth perception of virtual objects. This improved accuracy becomes more prevalent as egocentric distance increases, and can be performed faster as drivers become more familiar with the display.

RELATED WORK

Depth Perception in Virtual Environments

Depth perception is the ability of the viewer to view, recognize, and estimate depth or egocentric distance (Swan et al., 2007). There are many different cues that can assist with this perception, such as occlusion, binocular disparity, motion parallax and others as discussed in Cutting and Vishton's work (Cutting and Vishton, 1995). One major hindrance with depth perception in many displays is the vergence-accommodation mismatch (Hoffman et al., 2008). Both vergence and accommodation are oculomotor cues that involve muscle groups controlling eye direction and lens flexing sending signals to the brain assisting distance and depth perception. A mismatch occurs when these two muscle groups send differing signals for the depth of objects. Volumetric displays can accurately simulate both accommodation and vergence as well as many others since they create these cues naturally through voxels.

Over the years, many experiments have explored the perceptual issues in virtual environments and display technologies. McIntire et al. (McIntire et al., 2014) synthesized the body of work evaluating 3D display technologies from 184 experiments. From their summary, they found that 3D displays outperform their 2D counterparts in 75% of experiments. For our specific use case, 3D displays outperformed 2D displays in 57% of depth perception studies. In the field of human-computer interaction, AR displays have been evaluated through discrepancies between perceived distance and depth between virtual and physical objects (Livingston et al., 2013). Kruijff et al. (Kruijff et al., 2010) performed an overview of experiments involving perception with AR displays and found that the most common issue with these displays was incorrect depth interpretation.

Unfortunately, research has found that depth perception is difficult to measure accurately. Pagano et al. (Pagano and Bingham, 1998) found participants directly estimating distances numerically was not an effective way of reporting perceived distance. Other experiments found success with blind walking (Interrante et al., 2006; Loomis and Knapp, 2003), but that will not work for our vehicular context. Distance matching, where a virtual object is matched with a real-world object at a certain egocentric distance, has also seen success in experiments (Jones et al., 2008; Swan et al., 2015; Bark et al., 2014). Since drivers are often tasked with pedestrian or collision warnings with real-world people are objects, we chose to implement this method.

Depth perception is furthermore compressed in virtual environments (Interrante et al., 2006; Dey and Sandor, 2014; Grossman and Balakrishnan, 2006; Jones et al., 2008). Research has found that users underestimate egocentric distance in virtual, mixed, and augmented realities and evaluated some root causes of the issue. Loomis et al. argue that field of view does not affect depth perception (Loomis and Knapp, 2003), though Kline and Witmer found that displays with larger field of views gave users better depth perception (Kline and Witmer, 1996). Cidota et al. (Cidota et al., 2016) explored using visual effects as a possible method to increase depth perception, but both of their proposed techniques actually decreased perception.

Volumetric Displays

Volumetric displays have continually evolved over the years to improve their definition and design. Sand et al., for example, explored using water vapor as a 'fogscreen' display (Sand and Rakkolainen, 2014), while Hirayama et al. are using 'quantum dots' to create more compact volumetric displays (Hirayama et al., 2015). Other efforts have worked to create faster image processing to create higher quality images in volumetric displays (Kim et al., 2016b).

Depth perception has been explored previously with volumetric displays in a few experiments. Grossman and Balakrishnan compared volumetric displays to head-mounted stereoscopic displays and found that, while depth perception was improved, depth perception was still underestimated by about 20% (Grossman and Balakrishnan, 2006). Bark et al. performed an experiment looking at depth perception with a swept-volume volumetric display similar to ours (Bark et al., 2014). However, in their experiment the participants had to identify which sign a paper plane was hovering over, with no granularity to the judgments. Their experiment revealed that the participants could identify the correct distance 93-97% of the time as compared to 33% of the time with a traditional 2D display.

New display technologies are also being introduced that also create depth cues accurately. Light-field displays, for example, have been researched by Huang et al., but they are computationally expensive and have issues with display brightness (Huang et al., 2015). Both of these factors limit their use in automotive contexts. Varifocal displays are another emerging competitor (Dunn et al., 2017; Liu et al., 2008). However, these displays can only use a single variable focal plane to display virtual objects while volumetric displays can use multiple focal planes simultaneously.

Automotive Head-Up Displays

As AR becomes more popular in vehicular contexts, how perceptual issues impact driver performance will become more important. Smith et al. (Smith et al., 2017) evaluated how head-up displays (HUDs) affect drivers' eye scan patterns and performance as compared to head-down displays. The authors found that HUDs increase visual task performance as compared to the head-down display and did not impact driving performance even though participants sustained longer glances at the HUD. Other experiments are working on measuring user performance with HUDs, through changes in graphics (Tonnis et al., 2008) or placement of the HUD (Haeuslschmid et al., 2015).

As improvements in HUD design continue, many driving aids are being designed to use the HUD to display contextually relevant information to the driver. Kim et al. (Kim et al., 2016a) studied how HUDs improve performance detecting pedestrians over auditory warnings. Kim et al. (Kim et al., 2018) also looked into if conformal graphics would benefit monoscopic and volumetric displays. In both cases, the conformal graphics improved both depth perception and driver performance. Rusch et al. (Rusch et al., 2013) found that graphical warnings on HUDs decreased response time in identifying road hazards on long drives. Tran et al. (Tran et al., 2013) worked on a safety aid that assisted with drivers making turns against oncoming traffic. The above drivers' aids use HUD technologies and require quick response times so that the driver can react appropriately.

The need for drivers to identify hazards necessitates for them to be able to accurately identify how far the hazards are. Several experiments have been performed to measure depth perception on HUDs. Ng-Thow-Hing et al. (Ng-Thow-Hing et al., 2013) found that when drivers were presented with focal cues in an AR display, distance perception error dropped from 22% to 9.5%. Smith et al. (Smith et al., 2015) performed an experiment in identifying depth perception of virtual pedestrian displayed on an AR HUD and physical pedestrians. They found that there was more variation in answers for the distance of the virtual pedestrian than the physical one. Broy et al. (Broy et al., 2014) studied design parameters for head-up displays for optimizing comfort and depth perception. The findings from these experiments suggest that improvements to display technology in terms of depth perception could increase performance in detection of hazards.

OBJECTIVES

Given the limited availability of literature explicitly comparing human performance using multi-focal plane AR displays to single-focal AR planes, we established several research questions to guide the work:

1. Do volumetric displays improve the accuracy of depth judgments as compared to fixed-focal plane displays?
2. Are there speed-accuracy tradeoffs in depth judgments and are those tradeoffs different with a volumetric display (as compared to single focal plane)?
3. Does the magnitude of distance between the initial position of the virtual target and the real-world target effect the time and accuracy of depth judgments?
4. Is the repeated use of a multi-focal plane display associated with less eye fatigue than a fixed-focal plane display?

METHODS

It seems clear that spatial decision-making benefits from the addition of vergence-accommodation cues allowed by volumetric displays. Past work addressing this question focused on participants' ability to identify a closest real-world reference point. To more precisely measure the effects of volumetric displays, we attempt to answer this question by allowing participants to control the distance of a virtual object in an effort to exactly match the distance of a real-world reference. We compare performance results of this task for both fixed-focal plane and volumetric displays. By doing so, we can examine volumetric displays' impact on speed and accuracy of distance perception tasks. We hypothesize that the provision of vergence-accommodation cues by volumetric displays will yield improvements in both speed and accuracy of the aforementioned task with the caveat that these benefits will experience diminishing returns up to a distance of 10 meters, at which point they cease to contribute to distinguishing between distances as outlined by Cutting and Vishton (Cutting and Vishton, 1995).

EXPERIMENTAL DESIGN

Study Design

Participants were tasked with performing 60 distance matching tasks over two display conditions: (1) volumetric presentation using multiple focal planes, and (2) fixed-focal plane presentation mimicking that of a traditional transparent display. The ordering of these conditions was alternated across participants to minimize ordering effects.

Real-world referents existed as three rectangular, single-color signs placed at 7, 9.5, and 12 meters (see Figure 1); all medium-field distances as defined by Swan et al. (Swan et al., 2007). The near limit of this range was chosen due to hardware constraints, and the far limit allowed us to examine the impact of volumetric display just past the effective range of 10 meters for vergence cues (Okoshi, 2012). By doing so, we are able to isolate other cues provided by volumetric presentation and examine their impact out and along-side vergence cues.

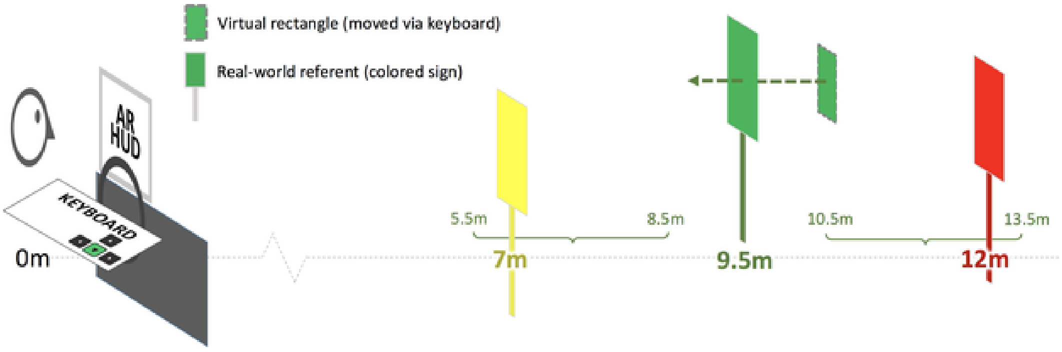
Several strategies were employed to limit the number of available distance cues. Virtual rectangles were presented with width and height ranging from 0.2 to 0.5 meters to prevent participants from matching virtual and real-world objects based on their relative size. These virtual rectangles were initially placed randomly between 1 and 3 meters away from their associated real-world referent either closer to or farther from the participant than its associated referent. Whether the virtual rectangle was closer to or further from the participant than the referent was balanced such that the amount of virtual rectangles that were initially closer was equal to the amount that were initially farther. Participants were instructed to refrain from moving their head from side to side to minimize information obtained through parallax cues. Combined, these strategies effectively constrain available distance cues to oculomotor vergence, accommodation, and binocular disparity.

Procedure

Before beginning the experiment, participants were instructed to value accuracy over speed. Additionally, participants were not made aware of which display type was actively being used to present virtual objects.

At the start of each task, the participant was presented with a virtual rectangle with the goal of matching its distance to that of its real-world referent (indicated by matching colors). The participant would manipulate the virtual rectangle's distance using the arrow keys on a provided keyboard. Once satisfied with the virtual rectangle's placement, the participant pressed the space bar to indicate completion of the task, after which the next virtual rectangle would be generated and presented to the participant. This continued until completion of the session for the relevant display condition, after which participants completed any relevant questionnaires.

Figure 1. Conceptual illustration of experimental setup depicting position of participant and three real-world referents (colored signs). In this figure, the participant is using the down key to bring the virtual green rectangle closer in an attempt to depth match the virtual green rectangle to its real-world green referent. The initial starting position of virtual rectangles was randomized between 1 and 3 meters on either side of real-world referents.



The study was a 2 × 3 repeated measures design with independent variables: display technology (fixed-focal plane and volumetric) and referent distance (7, 9.5, and 12 meters). Dependent measures consisted of distance estimation in meters and task completion time in milliseconds.

HARDWARE AND SOFTWARE

Participants sat in front of a mock vehicle dashboard which houses our volumetric display, shown in Figure 2. Voxel (red, green, blue, and depth) information is rendered by a C++/Qt5 graphical engine and transmitted over USB3 whose connection is managed by a customized version of the libusb C library. The display is oriented such that three real-world reference objects are within a participant's field of view while viewing virtual objects on the volumetric display (Figure 3). These references are placed off center to avoid confusing occlusion cues caused by intersection of the real-world and virtual objects. Virtual objects were presented in the unobstructed display volume of 7 to 13 meters, the farther end of which is physically indicated by a black sheet.

To simulate a fixed-focal plane display, the volumetric display is artificially limited to only display graphics on a single focal plane, mimicking the behavior of a traditional transparent display. During fixed-focal plane trials, graphics were displayed on the focal plane at 7 meters, the nearest available on our display.

The volumetric display used was provided by Honda Research Institute. It projects graphics at 60Hz in a swept, 17° circular volume. This display expands the range of available distance cues by rapidly switching between focal planes while displaying graphics, effectively granting users the ability to employ binocular disparity, motion parallax, and oculomotor convergence and accommodation as distance cues.

To ensure these cues can be effectively employed by participants, the display was calibrated by presenting three virtual objects of known size, shape, and location. Real-world objects mimicking the appearance of these virtual objects were placed at the same locations as the virtual objects. Field of view of the display (defined by software) was altered until the virtual objects appear to match the location of their real-world counterparts. After altering field of view, a manual focus single-lens reflex camera with a vernier focus feature was used to validate that the focal planes of the virtual and real-world objects match according to the process outlined by Kerr (Kerr, 2005).

Participants

Participants were selected based on exclusionary criteria intended to ensure their safety. Any history of epilepsy, seizures, or migraines disqualified potential participants; this is necessary to avoid adverse effects that may be caused by visual displays. Additionally, to maintain the integrity of collected data, participants were required to have working binocular vision and perfect or corrected-to-perfect vision.

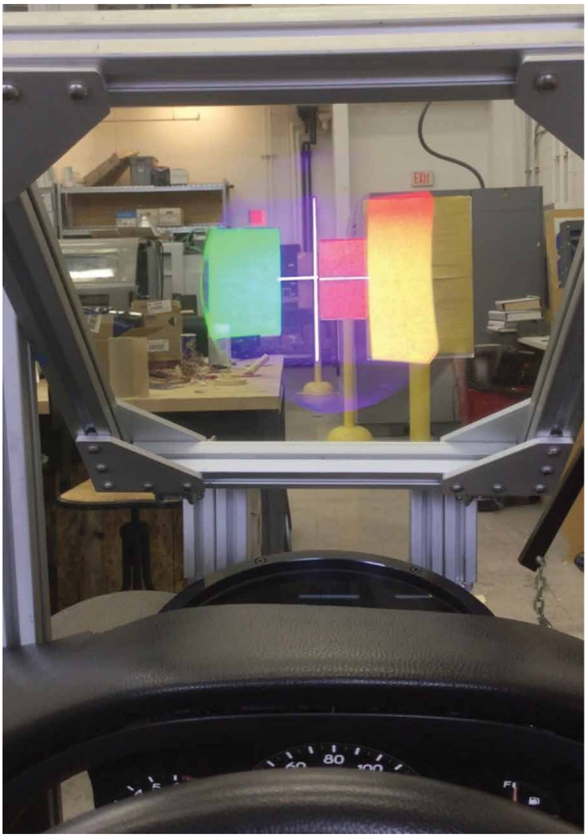
There were 11 participants with ages ranging from 20 to 33 and a mean age of 22.8. Nine of the participants were male and two were female. Five were given the volumetric display condition first, while 6 were given the fixed-focal plane condition first. Six participants had glasses or contacts that corrected their vision.

Data Collection

Participants completed five questionnaires over the course of the experiment. General background information was collected via a demographic questionnaire administered before starting the experiment. Simulator sickness questionnaires were administered before starting the experiment, after

Figure 2. The volumetric AR display used in our experiment was housed in a fixed mockup vehicle dash. Participants viewed the virtual rectangles and real-world referents through the glass pane located above the steering wheel.

Figure 3. A view of the active volumetric display with virtual rectangles overlaid on the real-world referents (colored, rectangular signs). The white center cross-hair and correctly-sized virtual rectangles show here were used for calibration purposes only; participants saw randomly sized rectangles as described in the text, presented one at a time.



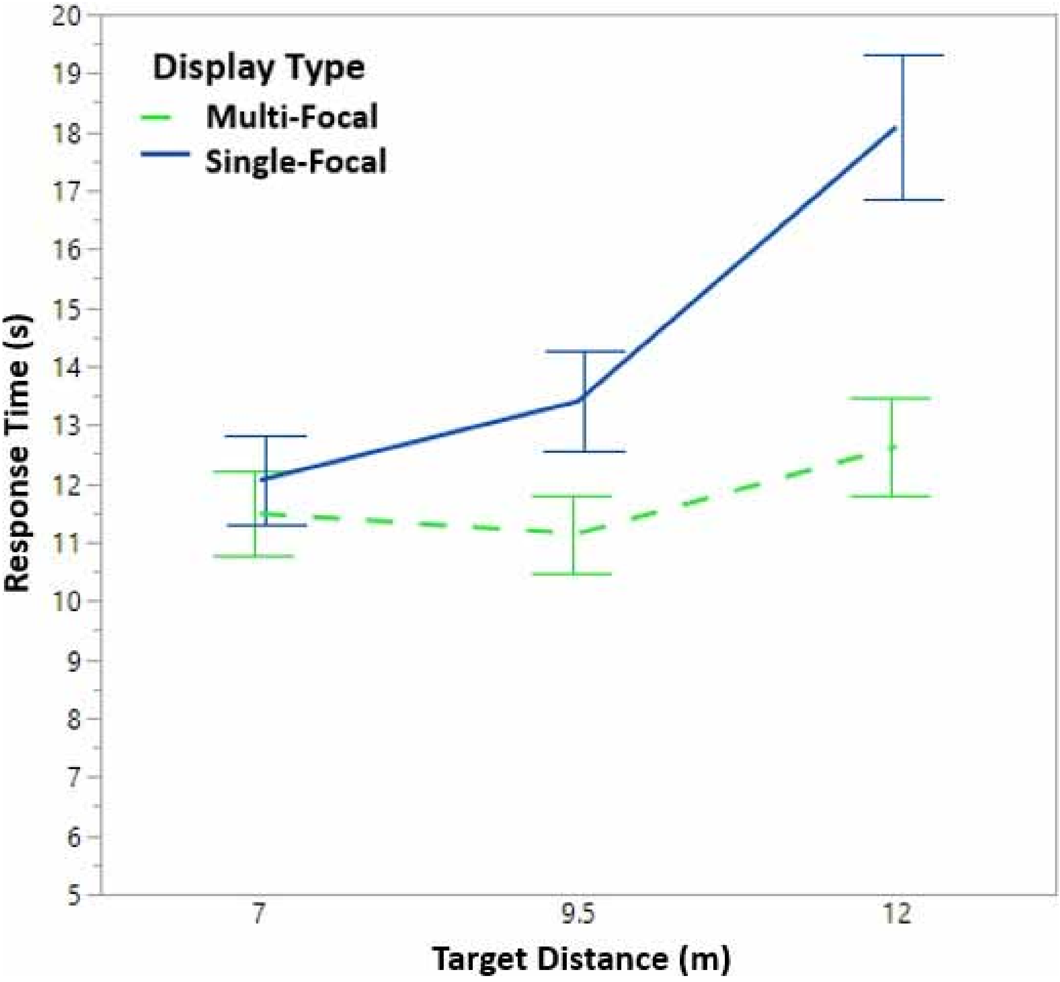
completing the first session, and at the end of the experiment to monitor participants' perceived health over the duration of the experiment. Finally, participants were asked to complete a post-experiment questionnaire containing the following open-ended questions:

1. Please describe any differences you noticed between technology setups;
2. Were there any noticeable differences in the way you had to gauge distances with the two technologies?
3. Which display technology did you prefer?
4. Were there any techniques you leveraged in order to match the distances? If yes, please describe them.

Answers to these questions informed experimenters of behavioral patterns not immediately apparent through direct observation of participants.

The software responsible for rendering the virtual objects simultaneously collected quantitative metrics of performance. These metrics included distance of the real-world referent, distance of the virtual object after manipulation by the participant, and task completion time in milliseconds.

Figure 4. Comparison of task completion time to referent target distance, with standard error of the mean bars



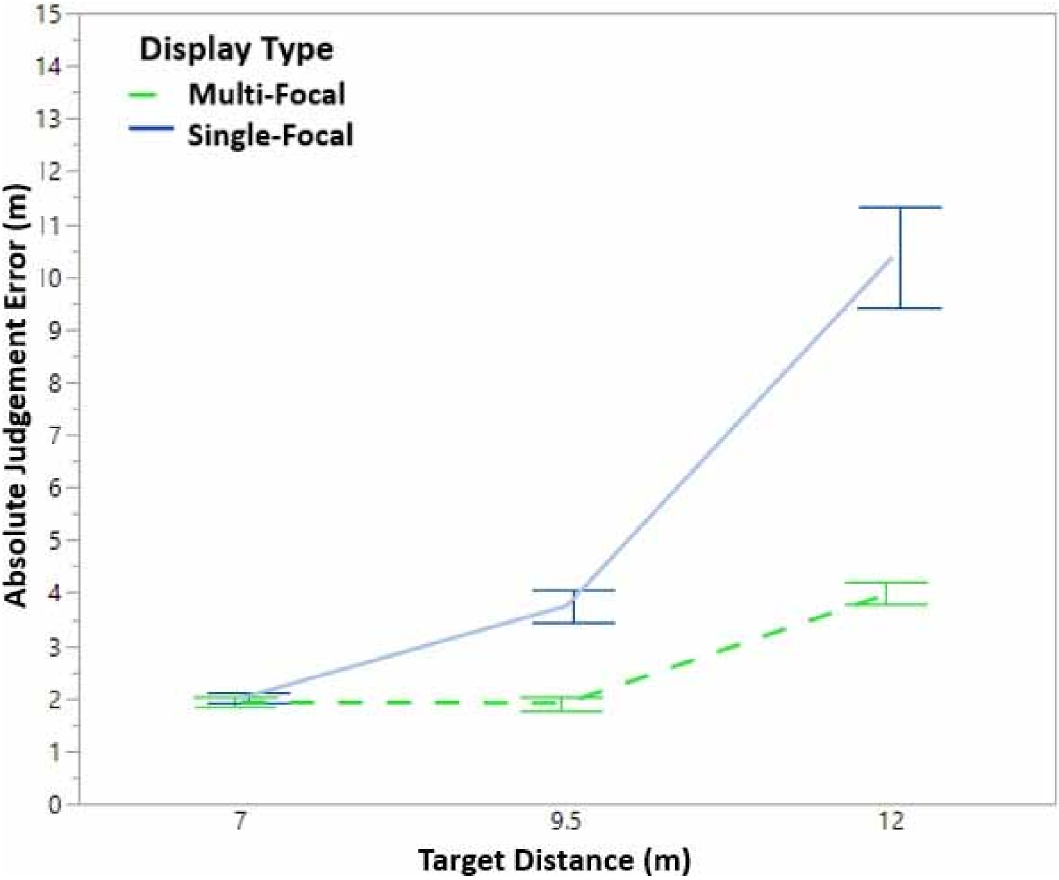
QUANTITATIVE RESULTS

Speed and Accuracy of Depth Judgments

After confirming assumptions of normality, we conducted a two-way repeated measures ANOVA to evaluate whether task completion times (ms) varied across different display types and referent distances (Figure 4). Results show significant effects of referent distance (F (2, 18) = 23.518, $p < 0.001$) and an interaction between display types and referent distance (F (2, 18) = 10.71, $p < 0.001$). Post-hoc comparisons with Tukey's adjustments for multiple comparisons revealed that task completion time is significantly different ($p < 0.01$) among display types for the referent at 12m. Participants made faster judgments in the multifocal condition over the single focal condition.

We performed two-way repeated measures of ANOVA to assess whether display type or referent distance impacted participants' absolute judgment error, defined as the absolute value error between a final judgment and actual target distance (Figure 5). We performed all testing after logarithmic transformation as data was not initially normally distributed. Tests found a main effect for display type (F (1, 613) = 74.80, $p < 0.001$) and referent distance (F (2, 613) = 35.08, $p < 0.001$). Post-hoc comparisons with Tukey's adjustment for multiple comparisons revealed that participants' absolute

Figure 5. Comparison of absolute judgment error to referent target distance, with standard error of the mean bars



judgment error was greater for the single focal display (*t* (613) = 8.65, $p < 0.001$), while post-hoc tests for target distance revealed judgment error was greater for the 12m target distance than 9.5m (*t* (613) = 10.09, $p < 0.001$) and 7m target distance (*t*(613) = 13.39, $p < 0.001$), and similarly that judgment distance was greater for the 9.5m target distance than the 7m (*t* (613) = 2.40, $p = 0.017$) Examining interactions, we saw larger judgment error for single as compared to multi-focal volumetric at 9.5m ($p = 0.002$) and 12m ($p < 0.001$) target distances with no significant difference between display types at the 7.5m target distance (Figure 3).

Speed-Accuracy Tradeoffs in Depth Judgments

To further examine the nature and relationship between response speed and judgment error, we performed multivariate analysis between the two measures using pairwise correlations. Results show that a weak to medium positive correlational relationship exists between task completion time and judgment error for both displays ($R = 0.458 \pm 0.063$). When performed specific to display type, correlational pairwise comparisons revealed a medium positive relationship between task time completion and judgment error ($R = 0.5056 \pm 0.086$) for the static display condition but no relationship between the two variables for the multi-focal display condition ($R = 0.001 \pm 0.109$).

Initial Position of Virtual Target

We were interested in evaluating whether the absolute difference in distance between target and referent location (i.e. the total distance participants were required to move the virtual object from its initial referent position towards the final correct target) impacted either response time or judgment error, and performed pairwise correlational analysis to test for effects. However, no significant correlations were found for between the required correction distance and either response time or judgment error.

Simulator Sickness, Eye Strain, and Fatigue Over Time

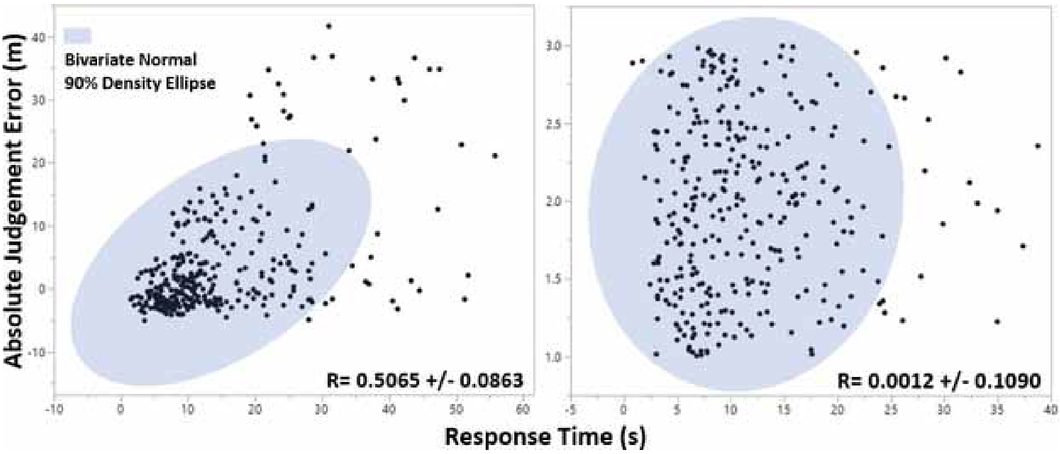
We had participants fill out the SSQ specifically to investigate eye strain common in AR displays. To address this, we analyzed the data gathered by the SSQ. Six participants reported no increase in sickness metrics at all. Four reported a slight increase in either eye strain or difficulty focusing, which makes sense given that the experiment required them to concentrate and adjust their focus between virtual objects and real-world referents. We looked into if there was any relation between those that did report eye strain and which trial they completed first (fixed focal or volumetric). However, they split evenly between the two, with two reporting eye strain after the fixed-focal display and two after the volumetric display.

We conducted Friedman's tests on the SSQ dataset, which is a non-parametric alternative to repeated measures ANOVA for ordinal data, since the SSQ data is not normally distributed. We found no differences in 15 out of 16 subscales of SSQ. Five participants felt slight (i.e. rating=1) eye-strain while using the HUD which is statistically significant (Friedman's test shows $\chi2$ (2) = 7.54, $p = 0.023$). Post-hoc comparison between experimental conditions revealed that both displays are associated with increased eye-strain ($p < 0.001$ for both). However, no differences between multi-focal volumetric and single-focal display was found.

Effects of Fatigue and Practice Over Time

Since we collected the Kennedy SSQ after every display condition, as opposed to after every trial, we are unable to examine fatigue over a series of display condition trials. Thus, we examined eye fatigue over time indirectly using measures of response time and absolute judgment error under the hypothesis that participants would get slower and/or less accurate over the course a set of display condition trials.

Figure 6. Multivariate correlation between task completion time and absolute judgment error for both fixed and multi-focal displays, with density ellipse showing where 90 percent of data is expected to lie



We first performed a Durbin-Watson test to check for effects of auto-correlation and found moderate but acceptable effects for both response time (*d* = 1.235) and absolute judgment error (*d* = 1.432) across all trials, display conditions and participants.

In order investigate whether participants experienced any learning or practice effects over time, we next performed multivariate correlational analysis to explore whether any relationship existed between either task completion time and trial number or judgment error and trial number. Pairwise correlations revealed no relationship between task time and trial number for the fixed display condition (R = −0.1381±0.106) but did reveal a weak-to-moderate negative relationship between task time and trial number for the multi-focal volumetric display condition (R = −0.4700±0.081). However, no relationship was found between judgment error and trial number for either the fixed focal plane display (R = −0.058) or the volumetric display (R = −0.099).

DISCUSSION

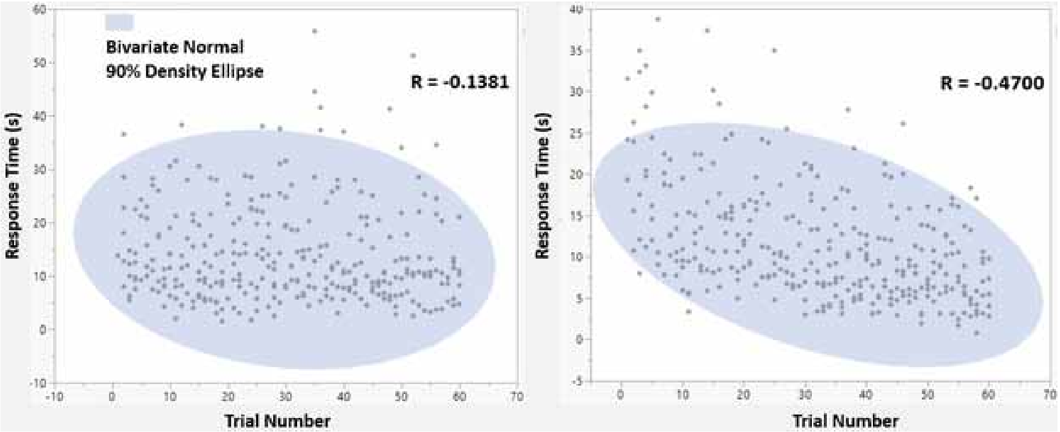
The data we gathered for our 2x3 repeated measures experiment generated results that helped us understand our research questions.

Before discussing the results, it should be noted that our distance estimation results in figure Figure 4 appear to show an overestimation of distances rather than the underestimation that other depth perception experiments find. This overestimation is actually a result of our experimental design. In most other depth perception experiments, participants control the depth of a real-world object and attempt to match its depth with a virtual object. Instead, we gave our participants control of a virtual object and asked them to match its depth with a real-world object. Therefore, our results are consistent with prior work, where participants underestimate egocentric distances when placing real-world objects at estimated virtual distances (Interrante et al., 2006).

Depth Judgments

From the collected data, we found that the judgment error for the volumetric display was not significantly better for the referent distances of seven meters. However, the participant performance was significantly better for the 9.5 and twelve-meter referent distance (see Figure 5). The disparity in significance could be explained by the fixed-focal plane being set at seven meters. In other words, since the fixed-focal plane was set to the same distance as the first (seven meter) referent target, there were no differences as compared to the multi-focal condition since the multi-focal was set to a

Figure 7. Multivariate correlation between task completion time and trial number for both fixed and multi-focal displays, with density ellipse showing where 90 percent of data is expected to lie



similar distance. For the second (9.5) and third (twelve meter) referent target, the focal cues are stark enough that significance is established.

These results could further suggest that the fixed-focal plane condition was hindering distance perception more so than the volumetric display was assisting. This hindrance would explain why participants took significantly longer in order to match the furthest (twelve meter) referent target in the fixed-focal plane condition. Since we attempted to remove the effectiveness of other focal cues like relative size (by randomizing the displayed rectangle's height and width) and removed other possible learning factors (by randomizing the initial distance to the referent target) the focal planes were the only active source providing cues to the participants. We expect that those focal planes anchored the participant's depth perception to the focal plane's distance. For the fixed-focal plane condition, this anchoring created a conflict when placing the virtual object. This effect can be seen in the further statistical analysis where participants took longer to extract cues when using the fixed-focal plane condition at twelve meters (see Figure 4). Even though the participants took more time, their judgment was still significantly worse due to the incorrect cues being provided to them. For our road scene environment, we expect this mismatch decreased confidence in a driver's perception could negatively affect safety judgments or increase reaction time when a driver is presented with a hazard. Furthermore, these results suggest that HUDs using a fixed-focal plane should avoid using conformal graphics and stick to more simple symbology when displaying hazards. In other words, if the display does not afford the visual cues a user might expect when seeing a certain type of image, it may only confuse and confound the user rather than improving their performance in part or conditionally. As a further example, if the perceptual form of a graphic is going to be placed in the real world and needs to be related to a real world referent, e.g. turn arrow with a real world street, a multi-focal display will likely be more appropriate over a traditional fixed-focal displays. These results could be investigated in future endeavors to determine the strength of this anchoring effect.

Unique Interactions With Volumetric Displays

One interesting thing to note is that participants interact with the volumetric display in a fundamentally different way than they do with the fixed-focal display. From Figure 6, we can see a clear medium strength relationship between participants' absolute judgment error and response time when using the volumetric display but no such relationship when participants interact with the fixed-focal display. We believe that participants leveraged the cues provided in the volumetric display to improve both their time and error simultaneously. We know this is due to the volumetric display since there is no correlation in the fixed-focal condition. Another finding of note is that with continued use of the volumetric display, we observed that participants tended to improve their time toward making depth judgments when using the volumetric display, but not when using the fixed-focal plane display (see Figure 7). We see this as a practice effect, in that participants gained greater confidence about where to place the virtual object more confidently. What is interesting to note is that despite this increase in judgment speed, the participants did not actually show improvement in the quality of judgments themselves over time whatsoever. It is possible that participants either incorrectly perceived their judgments as improving, or simply gained more familiarity and comfort with the visual cues afforded uniquely by the multi-focal volumetric display. The question then becomes whether this improvement in speed could truly be considered as an advantage unique to the volumetric display. However, we argue that because the display regardless allowed participants to confidently act more quickly with their depth judgments, this improvement in action time would in turn likely afford any user the ability to allocate more cognition to perform other important tasks when used in use cases where the user's actions are time dependent, such as driving. For example, while performing a common route-following task, the driver could pay more attention to pedestrian traffic or secondary tasks like changing the radio station, while still judging the distance to their next turn with greater efficiency as compared to a static display. Therefore, even though use of the volumetric display might not improve a user's judgment with practice, it still could be used more efficiently over time than a static display.

Qualitative Data

The qualitative data was analyzed to better users' perceptions of the display and any strategies they employed to match the distances. Participants noted that they often used the relative speed at which the distance changed for the virtual object between button presses to estimate its distance. This would imply that an apparent size cue may have been instrumental for participants as a method of measurement. Further experiments could have different movement intervals to hinder such strategies and thus force participants to rely more on their visual focus. Six participants further reported that the 12-meter referent distance was the hardest to place, especially in trials employing the fixed-focal plane display. That insight follows the quantitative data, where participants incorrectly estimated the 12-meter distance with the fixed-focal display technology. Only two participants stated that they concentrated on how their eyes were focusing on the rectangles as cues to the virtual object's distance. Ten of the 11 participants indicated a preference for the volumetric display technology due them feeling it was easier to work with. Further qualitative data could be gathered in future research to address why participants felt the volumetric display was easier to work with than the fixed-focal display.

LIMITATIONS

Our volumetric display was calibrated for vehicular contexts, which means that its minimum focal distance (seven meters) was higher than for many other AR displays. However, while we are focusing on distances in 'action-space,' as defined by Cutting and Vishton, we can consider the minimum to be seven meters rather than 1.5 since most driving hazards are greater than seven meters away (Cutting and Vishton, 1995). Similarly, the upper bound of our target distances for our experiment was due to the dimensions of our testing room. The maximum distance our room (an equipment bay) afforded was 13 meters. For future research we would like to find a larger testing area to experiment with further distances. Unfortunately, we also wanted to control lighting to avoid issues with display fidelity. This requires an indoor testing facility as we cannot control weather patterns. In addition, finding a larger test area would address another limitation. The equipment bay we performed our study had many visual distractors which could have impacted the qualitative and quantitative results.

Our low participant count was another limitation of our study. To account for this, we had each participant perform 30 trials for each display condition. We still would prefer to perform a larger study in the future with more participants.

We used three different colors (red, yellow, and green) to differentiate between the referent distances and indicate to the participants which target they should match depth with. However, human depth perception is affected by chromostereopsis which is the perception that certain colors appear closer than others (Allen and Rubin, 1981). However, since we color-match our targets with the virtual representation (i.e., matching green virtual rectangles with the green sign) chromostereopsis should have affected both the virtual and physical representations equally. Furthermore, since we performed the experiment indoors, the lighting was consistent such that it should not have affected the depth judgments.

CONCLUSION AND FUTURE WORK

Our results show that volumetric displays do provide better performance and a means for more accurate depth judgments of virtual objects over traditional fixed-focal plane displays. In particular, this improvement increases as the target distances are further away from the fixed-focal plane's depth. We further hypothesize that this is due to users anchoring their perception on the fixed-focal plane's depth, which hinders their ability to judge depth accurately. Since volumetric displays have multiple variable focal planes, they can overcome this issue and have a more consistent level of judgments throughout their display volume.

In the future, we would like to address some of the confounding variables in our experiment, such as limiting the visual distractors or increasing the distances and number of referent targets. We would like to repeat the experiment in a large empty space, such as an empty parking lot. Furthermore, conducting the experiment in an outdoor setting would increase external validity, as the display is intended to be used while driving, but this may create new confounding variables in terms of weather and variable lighting.

ACKNOWLEDGMENT

The authors wish to thank Honda Research Institute for their assistance in this research.

REFERENCES

Allen, R. C., & Rubin, M. L. (1981). Chromostereopsis. *Survey of Ophthalmology*, *26*(1), 22–27. doi:10.1016/0039-6257(81)90121-1 PMID:7280992

Banks, M. S., Kim, J., & Shibata, T. (2013, May). Insight into vergence/accommodation mismatch. In Head- and Helmet-Mounted Displays XVIII: Design and Applications (p. 873509). International Society for Optics and Photonics.

Bark, K., Tran, C., Fujimura, K., & Ng-Thow-Hing, V. (2014). Personal Navi: Benefits of an Augmented Reality navigational aid using a See-Thru 3D volumetric HUD. In *Proceedings of the 6th International Conference on Automotive User Interfaces and Interactive Vehicular Applications*. doi:10.1145/2667317.2667329

Broy, N., Höckh, S., Frederiksen, A., Gilowski, M., Eichhorn, J., Naser, F., & Schmid, A. et al. (2014). Exploring design parameters for a 3D head-up display. In *Proceedings of The International Symposium on Pervasive Displays*. doi:10.1145/2611009.2611011

Cidota, M. A., Clifford, R. M., Lukosch, S. G., & Billinghurst, M. (2016). Using visual effects to facilitate depth perception for spatial tasks in virtual and augmented reality. In *2016 IEEE International Symposium on Mixed and Augmented Reality (ISMAR-Adjunct)* (pp. 172-177). doi:10.1109/ISMAR-Adjunct.2016.0070

Cutting, J. E., & Vishton, P. M. (1995). Perceiving layout and knowing distances: The integration, relative potency, and contextual use of different information about depth. In *Perception of space and motion* (pp. 69–117). Academic Press.

Dey, A., & Sandor, C. (2014). Lessons learned: Evaluating visualizations for occluded objects in handheld augmented reality. *International Journal of Human-Computer Studies*, *72*(10), 704–716. doi:10.1016/j.ijhcs.2014.04.001

Dunn, D., Tippets, C., Torell, K., Kellnhofer, P., Aks¸it, K., Didyk, P., & Fuchs, H. et al. (2017). Wide Field Of View Varifocal Near-Eye Display Using See-Through Deformable Membrane Mirrors. *IEEE Transactions on Visualization and Computer Graphics*, *23*(4), 1322–1331. doi:10.1109/TVCG.2017.2657058 PMID:28129167

Grossman, T., & Balakrishnan, R. (2006). An evaluation of depth perception on volumetric displays. In *Proceedings of the working conference on Advanced visual interfaces* (pp. 193–200). doi:10.1145/1133265.1133305

Haeuslschmid, R., Schnurr, L., Wagner, J., & Butz, A. (2015). Contact-analog warnings on windshield displays promote monitoring the road scene. In *Proceedings of the 7th International Conference on Automotive User Interfaces and Interactive Vehicular Applications* (pp. 64–71). doi:10.1145/2799250.2799274

Hirayama, R., Naruse, M., Nakayama, H., Tate, N., Shiraki, A., Kakue, T., & Ito, T. et al. (2015). Design, implementation and characterization of a quantum-dot-based volumetric display. *Scientific Reports*, 5. PMID:25683656

Hoffman, D. M., Girshick, A. R., Akeley, K., & Banks, M. S. (2008). Vergence-accommodation conflicts hinder visual performance and cause visual fatigue. *Journal of Vision*, *8*(3), 33–33. doi:10.1167/8.3.33 PMID:18484839

Huang, F.-C., Chen, K., & Wetzstein, G. (2015). The light field stereoscope: Immersive computer graphics via factored near-eye light field displays with focus cues. *ACM Transactions on Graphics*, *34*(4), 60. doi:10.1145/2766922

Interrante, V., Ries, B., & Anderson, L. (2006). Distance perception in immersive virtual environments, revisited. In Virtual Reality Conference 2006 (pp. 3-10). doi:10.1109/VR.2006.52

Jones, J. A., Swan II, J. E., Singh, G., Kolstad, E., & Ellis, S. R. (2008). 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* (pp. 9–14). doi:10.1145/1394281.1394283

Kennedy, R. S., Lane, N. E., Berbaum, K. S., & Lilienthal, M. G. (1993). Simulator sickness questionnaire: An enhanced method for quantifying simulator sickness. *The International Journal of Aviation Psychology*, *3*(3), 203–220. doi:10.1207/s15327108ijap0303_3

Kerr, D. A. (2005). Principle of the split image focusing aid and the phase comparison autofocus detector in single lens reflex cameras.

Kim, H., Anon, A. M., Misu, T., Li, N., Tawari, A., & Fujimura, K. (2016a). Look at Me: Augmented Reality Pedestrian Warning System Using an In-Vehicle Volumetric Head Up Display. In *Proceedings of the 21st International Conference on Intelligent User Interfaces* (pp. 294–298). doi:10.1145/2856767.2856815

Kim, H., Gabbard, J. L., Anon, A. M., & Misu, T. (2018). Driver Behavior and Performance with Augmented Reality Pedestrian Collision Warning: An Outdoor User Study. *IEEE Transactions on Visualization and Computer Graphics*, *24*(4), 1515–1524. doi:10.1109/TVCG.2018.2793680 PMID:29543169

Kim, H.-R., Park, M.-K., Choi, J.-C., Park, J.-S., & Min, S.-W. (2016b). Volumetric 3D display with multi-layered active screens for enhanced the depth perception (Conference Presentation). In SPIE Organic Photonics+ Electronics (pp. 994005–994005).

Kline, P. B., & Witmer, B. G. (1996). Distance perception in virtual environments: Effects of field of view and surface texture at near distances. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *40*(22), 1112–1116. doi:10.1177/154193129604002201

Kruijff, E., Swan, J. E., & Feiner, S. (2010). Perceptual issues in augmented reality revisited. In *2010 9th IEEE International Symposium on Mixed and Augmented Reality (ISMAR)* (pp. 3-12). doi:10.1109/ISMAR.2010.5643530

Liu, S., Cheng, D., & Hua, H. (2008). An optical see-through head mounted display with addressable focal planes. In *7th IEEE/ACM International Symposium on Mixed and Augmented Reality ISMAR 2008* (pp. 33-42).

Livingston, M. A., Gabbard, J. L., Swan, J. E., Sibley, C. M., & Barrow, J. H. (2013). Basic perception in head-worn augmented reality displays. In *Human factors in augmented reality environments* (pp. 35–65). New York: Springer.

Loomis, J. M. and Knapp, J. M. (2003). Visual perception of egocentric distance in real and virtual environments. In *Virtual and adaptive environments* (pp. 35-60). CRC Press.

McIntire, J. P., Havig, P. R., & Geiselman, E. E. (2014). Stereoscopic 3D displays and human performance: A comprehensive review. *Displays*, *35*(1), 18–26. doi:10.1016/j.displa.2013.10.004

Ng-Thow-Hing, V., Bark, K., Beckwith, L., Tran, C., Bhandari, R., & Sridhar, S. (2013, October). User-centered perspectives for automotive augmented reality. In *IEEE International Symposium on Mixed and Augmented Reality* (pp. 13-22). IEEE.

Okoshi, T. (2012). *Three-dimensional imaging techniques*. Elsevier.

Pagano, C. C., & Bingham, G. P. (1998). Comparing measures of monocular distance perception: Verbal and reaching errors are not correlated. *Journal of Experimental Psychology. Human Perception and Performance*, *24*(4), 1037–1051. doi:10.1037/0096-1523.24.4.1037 PMID:9706709

Rusch, M. L., Schall, M. C. Jr, Gavin, P., Lee, J. D., Dawson, J. D., Vecera, S., & Rizzo, M. (2013). Directing driver attention with augmented reality cues. *Transportation Research Part F: Traffic Psychology and Behaviour*, *16*, 127–137. doi:10.1016/j.trf.2012.08.007 PMID:24436635

Sand, A., & Rakkolainen, I. (2014). A hand-held immaterial volumetric display. *Proceedings of the Society for Photo-Instrumentation Engineers*, *9011*, 90110Q. doi:10.1117/12.2035280

Smith, M., Doutcheva, N., Gabbard, J. L., & Burnett, G. (2015). Optical see-through head up displays' effect on depth judgments of real world objects. In *2015 Virtual Reality (VR)* (pp. 401–405). IEEE. doi:10.1109/VR.2015.7223465

Smith, M., Gabbard, J. L., Burnett, G., & Doutcheva, N. (2017). The effects of augmented reality head-up displays on drivers' eye scan patterns, performance, and perceptions. *International Journal of Mobile Human Computer Interaction*, *9*(2), 1–17. doi:10.4018/IJMHCI.2017040101

Swan, J. E., Jones, A., Kolstad, E., Livingston, M. A., & Smallman, H. S. (2007). Egocentric depth judgments in optical, see-through augmented reality. *IEEE Transactions on Visualization and Computer Graphics*, *13*(3), 429–442. doi:10.1109/TVCG.2007.1035 PMID:17356211

Swan, J. E., Singh, G., & Ellis, S. R. (2015). Matching and reaching depth judgments with real and augmented reality targets. *IEEE Transactions on Visualization and Computer Graphics*, *21*(11), 1289–1298. doi:10.1109/TVCG.2015.2459895 PMID:26340777

Tonnis, M., Klein, L., & Klinker, G. (2008). Perception thresholds for augmented reality navigation schemes in large distances. In *7th IEEE/ACM International Symposium on Mixed and Augmented Reality ISMAR 2008* (pp. 189-190). IEEE. doi:10.1109/ISMAR.2008.4637360

Tran, C., Bark, K., & Ng-Thow-Hing, V. (2013). A left-turn driving aid using projected oncoming vehicle paths with augmented reality. In *Proceedings of the 5th International Conference on Automotive User Interfaces and Interactive Vehicular Applications* (pp. 300-307). doi:10.1145/2516540.2516581

Lee Lisle is a PhD Candidate in the Computer Science & Applications Department of Virginia Tech studying human-computer interaction and human-centered design. He is currently working on studying and designing interaction methods in augmented and virtual reality.

Coleman Merenda is a Ph.D. candidate in Virginia Tech's Industrial and Systems engineering specializing in human factors and ergonomics. He currently researches perceptual phenomena in transportation and driver interaction with novel interfaces.

Hyungil Kim is a senior research associate at Virginia Tech Transportation Institute. He received a Ph.D. in industrial and systems engineering specializing in human factors and ergonomics in 2017 from Virginia Tech. His current research focuses on human factors in augmented reality (AR) with applications in transportation.

Joseph L. Gabbard is an Associate Professor of Human Factors at Virginia Tech. He holds Ph.D., M.S. and B.S. degrees in Computer Science (HCI) and a B.A. in Sociology from Virginia Tech. Dr. Gabbard's work centers on human-computer interaction in augmented and virtual reality. For over 20 years, Gabbard has been a pioneer in creating new methods of design and evaluation for emerging interactive systems. Currently, Dr. Gabbard directs the Cognitive Engineering for Novel Technologies (COGENT) Lab, conducting basic and applied HCI and human factors research on perception and cognition. The research emphasizes the application of principles and theories from several disciplines to the design of augmented, virtual, and mixed-reality user interfaces, information presentation and user interaction. His research has been funded by the National Science Foundation, Microsoft, Facebook, Honda Research Institute, Jaguar Land Rover, the National Institutes of Health, and the Office of Naval Research.

Doug A. Bowman is the Frank J. Maher Professor of Computer Science and Director of the Center for Human-Computer Interaction at Virginia Tech. He is the principal investigator of the 3D Interaction Group, focusing on the topics of three-dimensional user interface design and the benefits of immersion in virtual environments. Dr. Bowman is one of the co-authors of 3D User Interfaces: Theory and Practice. He has served in many roles for the IEEE Virtual Reality Conference, including program chair, general chair, and steering committee chair. He also co-founded the IEEE Symposium on 3D User Interfaces (now part of IEEE VR). He received a CAREER award from the National Science Foundation for his work on 3D Interaction, and has been named an ACM Distinguished Scientist. He received the Technical Achievement award from the IEEE Visualization and Graphics Technical Committee in 2014. His undergraduate degree in mathematics and computer science is from Emory University, and he received his M.S. and Ph.D. in computer science from the Georgia Institute of Technology.