Measuring the Perceived Three-Dimensional Location of Virtual Objects in Optical See-Through Augmented Reality

Farzana Alam Khan*
Mississippi State University

Veera Venkata Ram Murali Krishna Rao Muvva†
University of Nebraska–Lincoln

Dennis Wu‡
Mississippi State University

Mohammed Safayet Arefin§
Mississippi State University

Nate Phillips¶
Mississippi State University

J. Edward Swan II‖
Mississippi State University

ABSTRACT

For optical see-through augmented reality (AR), a new method for measuring the perceived three-dimensional location of virtual objects is presented, where participants verbally report a virtual object's location relative to both a vertical and horizontal grid. The method is tested with a small ($1.95 \times 1.95 \times 1.95$ cm) virtual object at distances of 50 to 80 cm, viewed through a Microsoft HoloLens 1st generation AR display. Two experiments examine two different virtual object designs, whether turning in a circle between reported object locations disrupts HoloLens tracking, and whether accuracy errors, including a rightward bias and underestimated depth, might be due to systematic errors that are restricted to a particular display. Turning in a circle did not disrupt HoloLens tracking, and testing with a second display did not suggest systematic errors restricted to a particular display. Instead, the experiments are consistent with the hypothesis that, when looking downwards at a horizontal plane, HoloLens 1st generation displays exhibit a systematic rightward perceptual bias. Precision analysis suggests that the method could measure the perceived location of a virtual object within an accuracy of less than 1 mm.

Index Terms: Depth Perception—Augmented Reality—Optical see-through display——Perceived Location—HoloLens—Human Subject Analysis

1 INTRODUCTION

In augmented reality (AR), users observe virtual objects in real-world locations. A longstanding goal is that the locations of virtual objects appear as equally real, solid, precise, and believable as the locations of real, physical objects. This concept has been termed *locational realism* [10], and is fundamental to many compelling applications of AR technology [3]. For example, sports broadcasting with real-time AR overlays is very common, and the virtual objects have very high levels of locational realism. However, sports broadcasting uses *video see-through* AR, where image features from captured video frames allow virtual objects to be placed with a very high degree of precision [11]. This access to the underlying pixels of the video stream solves an important AR measurement problem; it answers the question "where in the real world is this virtual object located"? Once this question is answered quantitatively, the virtual object's locational error can be driven towards zero.

This work reports progress towards solving this measurement problem for *optical see-through* (OST) AR. For locational realism,

---

*e-mail: fk141@msstate.edu
†e-mail: mvvrmkr@gmail.com
‡e-mail: denniskwu@gmail.com
§e-mail: arefin@acm.org
¶e-mail: Nathaniel.C.Phillips@ieee.org
‖e-mail: swan@acm.org; corresponding author

the most important property of OST AR is that the merging of the virtual and real is a perceptual phenomenon that occurs in the observer's head: there are no image pixels to analyze. However, as reviewed below, there are two relevant threads of existing work, both of which consist of methods for measuring perceptual object location. These are (1) OST AR calibration methods, which seek to measure the position of the observer's eyes within an OST AR head-mounted display (HMD), and (2) depth perception methods, which seek to measure the egocentric distance at which a virtual object is located. This paper reports adapting the calibration method of Moser et al. [21], and applying it to the task of measuring the perceived location of virtual objects seen in a Microsoft HoloLens 1st generation HMD. Unlike depth perception methods, which seek to measure the single dimension of egocentric depth, the method provides a three-dimensional measurement of perceived object location[1].

2 BACKGROUND AND RELATED WORK

2.1 Calibration Methods

For all AR displays, to render virtual objects at intended locations, the position of the camera within the display's tracking system coordinate frame must be accurately measured. This measurement is termed *calibration*. For a video see-through AR system, camera calibration is accomplished by processing the captured camera image [1]. However, with an OST AR system, the final camera is the human eye, and thus there is no captured camera image to process. For many years, calibrating an OST AR system has required some sort of interactive user procedure, where the user aligns virtual points or shapes with real world objects. These calibration procedures have often been complex, difficult for users to physically perform, and noisy. Useful surveys and tutorial treatments of AR calibration include Axholt [1], Moser [22], and Grubert et al. [10]. The OST AR display used in this work is a HoloLens 1st generation display, which includes a built-in calibration procedure that requires users to monocularly align their finger with a series of virtual objects. Although the HoloLens documentation indicates that this procedure only measures interpupillary distance [20], the calibration procedure could also provide additional eye location information [23].

In addition to developing and refining calibration techniques, there have been a number of efforts to evaluate calibration techniques; see Grubert et al. [10] for a recent review. The 3D perceptual measurement technique described in this paper is based on a calibration evaluation study reported by Moser et al. [21]. In this work, Moser et al. used a similar 3D measurement technique to evaluate the calibration accuracy of three OST AR calibration methods. The evaluated methods include the Single Point Active Alignment Method (SPAAM) [30], which has been the most widely implemented and studied calibration technique. SPAAM was compared to two calibration methods that use an eye tracking camera to partially automate the SPAAM calculations.

---

[1]Portions of this work have been reported in a poster abstract (Khan et al. [15]) and a master's thesis (Muvva [24])
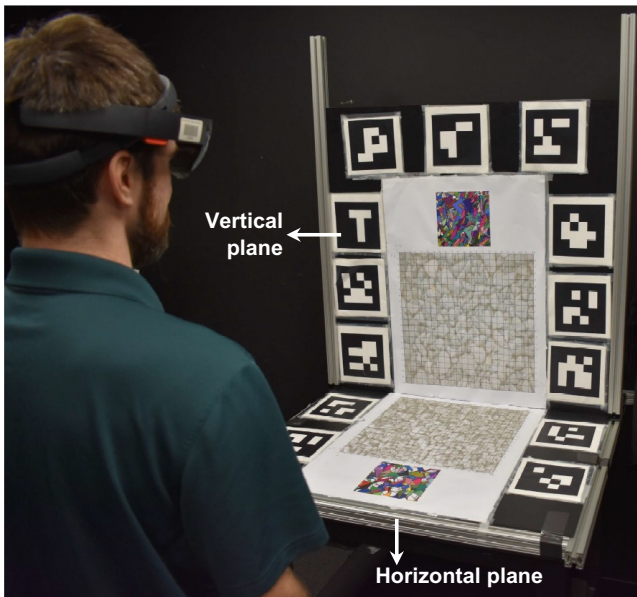
Figure 1: Participant viewing the experimental table.

2.2 Depth Perception Methods

Measuring the perceived depth of virtual objects has received significant attention in the AR research community. Much of this research has focused on adapting and validating depth judgement tasks, which allow quantitative measurement. A number of tasks have been studied, including pointer adjustment, blind walking, triangulation by walking, verbal estimation, bean bag tossing, and others [5, 8, 13–15, 28]. Useful surveys that cover AR depth judgement tasks, and report how different combinations of depth cues effect depth perception include El Jamiy and Marsh [6], Singh et al. [28], and Swan et al. [29]. Most of this existing work has investigated depth perception at *action space* distances of 2 to 15 meters, but a small number have examined *reaching space* distances of up to 1.5 meters. The method reported here operates in the range of 50 to 80 cm. Overall, measured depth perception accuracy in AR has been variable, with a generally established finding that the depth of AR virtual objects is usually underestimated [18, 26, 28, 29].

The method reported here was tested on a Microsoft HoloLens 1st generation AR display. Recent work has also reported using HoloLens 1st generation displays to estimate the perceived depth of AR objects, and again found underestimation [9, 26]. However, Fischer et al. [7] used a HoloLens 1st generation display to investigate perceptually aligning real and virtual information in a reaching space medical context, and found overestimation.

While depth perception represents a single scalar quantity, perceived 3D location is more complex, consisting not only of depth (*z*-axis), but also of abscissa (*x*-axis) and ordinate (*y*-axis) information as well. The proposed method measures all three dimensions, and thus promises to enable better ways of measuring the affects and interactions of additional cues on perceived location, such as shadows [4, 26], ground cues [4], and familiar size [19]. In addition, instead of focusing on perception, many researchers have investigated related challenges associated with object registration and tracking in AR [2, 16, 17, 21, 31, 32].

3 METHODS

The overall goal of this work was to measure the perceived three-dimensional location of a virtual object. To accomplish this, two measurement methods from Moser et al. [21] were adapted. The first method (1) evaluated the accuracy of different AR calibration techniques. A virtual cube was programmatically placed at a specific location on a grid coordinate. Both vertical and horizontal coordinate planes were evaluated, using a setup similar to Figure 1. A participant then verbalized the grid coordinate closest to where they perceived the center of the virtual cube to be located. A series of such trials served to measure the perceived accuracy of the calibration techniques. Even though participants verbalized the nearest 2×2 cm grid location, many repetitions of the task resulted in a measurement precision as low as 1 mm. In addition (2), Moser et al. [21] evaluated the *quality* of the virtual cube location. After verbalizing the nearest grid location, participants subjectively rated the placement quality on a scale of 1 (lowest quality: the cube perceptually straddled two grid locations) to 4 (highest quality: the cube was perceptually located in the center of the grid square).

Measuring Perceived 3D Location: The method proposed here adapts Moser et al.'s [21] first method, with some changes, to measure the perceived 3D location of a virtual object. The participant wears a calibrated OST AR display, and stands in front of two grids, mounted vertically and horizontally (Figure 1). A virtual object is then presented, in this case a virtual cube that is the same size as the grid spacing, against both the vertical plane (Figure 2e) and the horizontal plane (Figure 2f). For each plane, the observer then verbally reports the grid location of the cube's center.

Measuring Perceived Ground Contact: In addition, a common experience with OST AR displays, including the HoloLens 1st generation, is that virtual objects appear to float a few centimeters above the ground, or tabletop, or other flat surface. When developing the work reported here, the authors found it difficult, although not impossible, to position virtual objects directly on the tabletop (Figure 2). Therefore, participants also completed a secondary task, inspired by Moser et al.'s [21] second method, where they subjectively rated virtual objects as either being *in front of*, *on*, *penetrating*, or *behind* the grid surface.

4 EXPERIMENT 1: LOOP CLOSING AND RENDERING STYLE

The overall purpose of Experiment 1 was to determine whether perceived 3D location could be quantitatively measured. The method was applied to experimental participants wearing a Microsoft HoloLens 1st generation AR HMD. Within this overall purpose, two specific questions were examined.

The first question (1) was whether the built-in HoloLens tracking would successfully retain tracking accuracy after loop closing. The HoloLens achieves real-time tracking through a hardware implementation of the Simultaneous Localization and Mapping (SLAM) algorithm [17]. In SLAM tracking, when previously seen landmarks in the environment come back into view, the tracking system must recognize that the landmarks have been previously seen. This behavior is termed *loop closing*, and is a critical ability of a practical SLAM system [32]. This ability was tested by having observers turn in a complete circle between perceptual measurements, and examining the effect of this turn on perceived virtual object locations.

The second question (2) was whether the rendering style of the virtual cube would affect the accuracy of the perceived location. Two different rendering styles were examined (Figure 2), one in which the cube was rendered in a solid, blank style, and the other in which the cube was marked with a pair of lines that helped visually indicate the location of the cube's center. The second cube design was determined after analyzing data from the first half of the participants. Because the task requires estimating the center of the virtual cube, markings highlighting the center might make this estimation more accurate. In addition, because a green plastic cube was used to explain the task to the participants, the virtual cube color was changed to green.
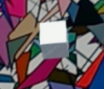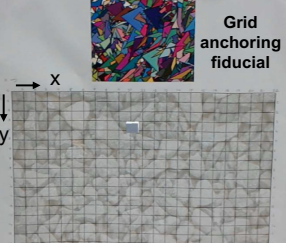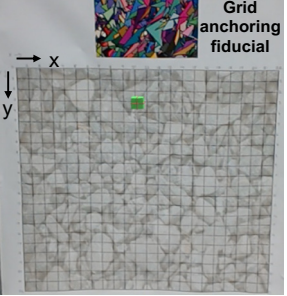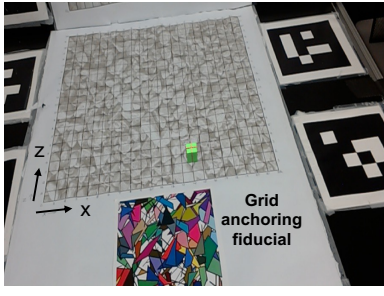
Figure 2: Experiment overview, showing the variables for both experiments (a). Two different levels of *movement* (b) were examined; when movement was present, participants turned around between trials. Two different *displays* (c) were examined, both identical HoloLens 1st generation models. Two different virtual *cube styles* (d) were also examined. All of these variables were examined against both a *vertical plane* (e), measuring the *x* (left-right) and *y* (up-down) dimensions, and a *horizontal plane* (f), measuring the *x* (left-right) and *z* (front-back) dimensions.

4.1 Method

Apparatus: As shown in Figure 1, participants wore a HoloLens 1st generation display and stood in front of a table, which supported vertical and horizontal planes made of plastic foamcore board. On each plane was mounted a 22×22 grid, where each grid cell was 1.95×1.95 cm. The vertical plane grid cells were numbered 1–22 along the *x* (left-right) and *y* (up-down) dimensions, while the horizontal grid cells were numbered 1–22 along the *x* (left-right) and *z* (front-back) dimensions. In addition to the grids, the planes showed three different styles of tracking fiducials. Above the vertical grid and in front of the horizontal grid was a colorful square Vuforia fiducial. The grids themselves overlaid a photograph of stones, which was another Vuforia fiducial. Finally, each grid was surrounded by a series of AR Toolkit tracking fiducials. The purpose of these fiducials is given below.

In addition to the HoloLens display, the experiment ran on an Alienware R17 laptop, with an Intel Core i7-670HQ CPU running at 2.60GHz, and 32 GB of memory. The experimental code was written in Unity version 5.6.4, and counterbalancing code was written in Perl.

Task: The purpose of the primary task was to quantitatively measure perceived 3D location. The participant saw a virtual cube, the same size ($1.95 \times 1.95 \times 1.95$ cm) as the grid, against either the vertical or horizontal plane (Figures 2e, f). The participant estimated the grid coordinate of the center of the cube, and verbalized these coordinates, first along *x*, and then along *y* (vertical) or *z* (horizontal). Participants estimated grid coordinates to the nearest tenth; for example, if the participant perceived the cube center to be between the *x* grid lines marked 12 and 13, slightly closer to 12, they might have verbalized "12.4 along *x*". Participants were trained to always follow the grid coordinate with the axis name, and to always state the *x* coordinate first.

In addition to the primary task, a secondary task was also performed. The purpose of the secondary task was to judge the apparent virtual cube position relative to the grid surface. After verbalizing the grid coordinates, the participant rated the quality of the cube's placement by verbalizing "1", "2", "3", or "4", indicating that the cube was perceived to be (1) floating in front of (vertical) or above (horizontal) the surface, (2) resting on the surface, (3) penetrating the surface, or (4) located behind (vertical) or below (horizontal) the surface. During this task, participants could see a sheet of paper that showed an example of each cube placement quality.

Independent Variables: Two *planes* were presented (Figure 2e, f), vertical (x, y) and horizontal (x, z), which together measured perceived location in three-dimensional (x, y, z) space. Two levels of *movement* were presented (Figure 2b). When movement was *yes*, after every trial the participant turned in a complete circle, while when movement was *no*, participants remained facing the table between trials. Movement varied within participant (Figure 2b). Movement order—*yes, no* or *no, yes*—was counterbalanced between participants, and movement direction (*left*, *right*) alternated and was counterbalanced within each participant. Finally, two different *cube styles* were shown (Figure 2d). The first 12 participants saw a *solid* white cube, and the second 12 participants saw a green cube that was *marked* with red lines. Cube style varied between participants.

Design: Within each plane, 10 *intended locations* were presented. Each location was randomly chosen without replacement from between 6.5 and 15.5 on the *x*-axis, and between 1.5 and 6.5 on the *y*- or *z*-axis. The random choice was restricted so that there was no back-to-back repetition of the same row or column. For each participant the same set of 10 intended locations was shown for each plane and movement level. Therefore, each participant completed 2 (*plane*: vertical, horizontal) $\times$ 2 (*movement*: yes, no) $\times$ 10 (*intended location*) = 40 trials. The presentation order of plane and movement was counterbalanced between participants with a 4×4 balanced Latin square [14].
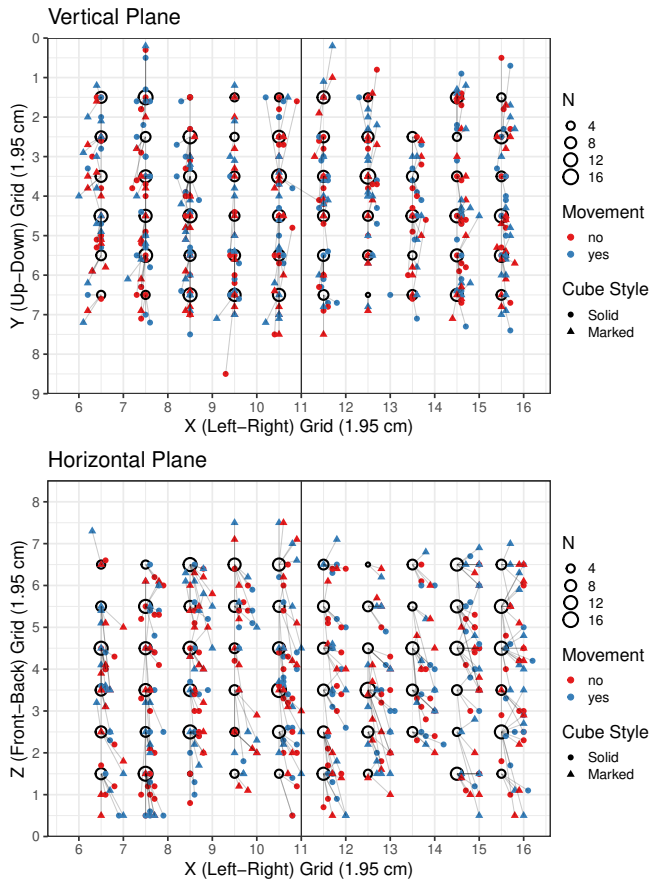
Figure 3: Experiment 1 intended vs. perceived location, shown on the experimental grid (Figure 2). The intended locations are hollow circles, sized according to the number of times the location was shown. The perceived locations are color-coded according to movement, and shape-coded according to cube style. On the vertical plane, most perceived location error occurred along the *y* (up-down) axis. On the horizontal plane, perceived locations had a notable rightward bias.

Dependent Variables: For each trial, the participant provided the perceived coordinates, x, y (vertical) or x, z (horizontal), and the quality rating 1, 2, 3, or 4.

Fiducials and Tracking: A fundamental AR requirement is aligning the display's coordinate system to a specific real-world location. This is what allows virtual objects to have specific spatial relationships with real objects. In the HoloLens development environment, this is known as *anchoring*. At the time these experiments were conducted, the HoloLens 1st generation programming environment did not have a way to set an anchor to a precisely defined real-world location. Solving this problem, the *precise anchoring problem*, was the single largest technical challenge of this experiment. Eventually, Vuforia visual tracking provided a solution.

Therefore, the experiment ran in two different tracking modes: Vuforia tracking mode, or HoloLens tracking mode. Before beginning a set of trials, the experiment was in Vuforia tracking mode. In this mode, Vuforia looked for a colorful Vuforia anchoring fiducial, and when one was found, placed the virtual cube to appear in the center of the fiducial (Figure 2d). Then, that cube position was used to set a HoloLens *anchor point*, and the experiment switched to HoloLens tracking mode. This ensured that the HoloLens coordinate system was properly positioned, rotated, and scaled with respect to the center of the anchoring fiducial. All of the intended cube positions were specified in a coordinate system centered at this anchor point.

In HoloLens tracking mode, the hardware SLAM algorithm continuously uses the tracking cameras to look for image features, such as corners, that can be reliably detected frame-to-frame [27, 32]. To provide trackable image features of the highest possible quality, AR toolkit fiducials were arranged around the horizontal and vertical planes, and the grids were overlaid on another Vuforia fiducial (Figure 1). While there was not a way to know what image features the HoloLens was tracking frame-to-frame, it was believed that these additional fiducials would provide it with an optimal visual tracking environment, and therefore produce the best possible tracking performance.

Procedure: Each participant first filled out human-subjects paperwork and several survey forms. Then the experimenter explained the task, using a 3D-printed plastic cube the same size as the virtual cube. The participant practiced the task until the experimenter was convinced that they were familiar with both the primary and secondary task. The participant also practiced turning and not turning between trials.

The participant then donned the HoloLens, and performed the HoloLens calibration procedure [20]. The HoloLens was placed in Vuforia tracking mode, and the participant then looked at either the vertical or horizontal plane, and reported on the quality of the virtual cube's placement and orientation within the Vuforia fiducial. Occasionally, the virtual cube either did not appear in the fiducial center or was strangely contorted. If the participant reported these errors, Vuforia was restarted and the procedure repeated. It was always possible to eventually overcome these errors. When the participant reported that the virtual cube was properly positioned, the HoloLens was placed in HoloLens tracking mode, and a group of 10 trials was performed. This procedure was repeated 4 times per participant, once for each level of plane and movement.

When observing the vertical plane, participants stood 60 cm from the end of the table, and when observing the horizontal plane, they stood 80 cm from the end of the table. These positions meant that the virtual object always appeared between 50 and 80 cm from the observer's eyes. The standing positions were marked and labeled on the floor with tape. Participants were not instructed to stand still; sometimes they shifted their weight or leaned while estimating object positions. Thus, participants also experienced the depth cue of motion parallax.

This experiment suffered from a longstanding problem with OST AR, which is that outside observers cannot know exactly what the participant wearing the AR device is seeing. To address this, each trial began with the HoloLens displaying explicit instructions for the participant, such as "trial 6, turn to the left, and report the *x* and *y* coordinates of the cube on the vertical plane", or "trial 4, report the *x* and *z* coordinates of the cube on the horizontal plane". The participant first read the instructions out loud, and then performed them. This procedure allowed the experimenter to know where the participant was in the experiment. The experimenter typed the participant's verbal location and quality rating into the experimental control program running on the laptop.

After the experiment, the participant was debriefed, and asked about their impressions and any strategies they may have followed.

Participants: Experiment 1 collected data from 24 participants, chosen from among the students and staff at Mississippi State University. One participant completed the experiment twice, in two separate sessions. This duplication was not discovered in time to be corrected for the reported analysis, but the duplication is not expected to change the findings. Therefore, the data was collected from 23 separate people.

4.2 Results

Figure 3 shows the intended locations as hollow circles, and the perceived locations as colored and shaped points. The hollow circles are sized according to the number of times each location was shown. Each participant saw 10 locations at random, and it is clear that some grid locations were shown more than others. For the vertical plane, there are noticeably more errors along the y (up-down) axis than along the x (left-right) axis. For the horizontal plane, there is a very interesting asymmetry in the perceived locations along the x (left-right) axis. First, there is a notable rightward bias in the perceived locations. Second, the degree of this bias changes along the x axis; there is very little bias along the left-hand side, but the bias grows towards the right, and is very prominent along the right-hand side.

Figure 4 shows the perceptual errors between intended and perceived locations. The black points represent the mean error; the error bars indicate one standard error of the mean (SEM). On the vertical plane, the x (left-right) axis judgments have a precision of 0.2 mm SEM, with no observable left-right bias ($F_{1,22} < 1$). However, consistent with Figure 3, the y (up-down) axis has a precision of 1.5 mm SEM, and an upward bias of $+3.5$ mm, which marginally differs from 0 ($F_{1,22} = 4.0, p = 0.059$). A mixed-model, repeated-measures ANOVA shows that for both the x and y dimensions there is no effect of either movement ($F_{1,22} < 1$), cube style ($F_{1,22} < 1$), or their interaction ($F_{1,22} < 1$).

For the horizontal plane, x (left-right) axis judgments indicate a precision of 0.4 mm SEM, and a rightward bias of $+4.7$ mm, which significantly differs from 0 ($F_{1,22} = 107, p < 0.001$). For the z (front-back) axis, the precision drops to 1.6 mm SEM, with a -7.4 mm frontward bias, which significantly differs from 0 ($F_{1,22} = 12.4, p < 0.01$). For both the x and z dimensions there is no effect of either movement ($F_{1,22} < 1$), cube style ($F_{1,22} < 1$), or their interaction ($F_{1,22} < 1$).

Finally, in Figure 5, the quality rating counts are displayed across the independent variables. It is clear that the *penetrating* and *below* ratings were generally rare, and movement had little effect. Participants predominately rated the cube as being *on* the grid surface, but rating the cube *above* the surface was also common. The solid cube style was more likely than the marked style to be seen as on the grid surface.

4.3 Discussion

Overall, there are no main effects of either movement or cube style, nor is their interaction significant. The lack of an effect for movement suggests that the built-in HoloLens SLAM tracking is able to successfully close loops, even when they are frequently encountered. The lack of an effect for cube style suggests that participants were able to successfully estimate the center of the solid cube.

The high degree of precision along the left-right axis, less than 0.5 mm for both planes, suggests the utility of the proposed perceptual measurement method. In particular, despite the relatively large size (1.95 cm) of the grid coordinates and virtual object, repeated measurements led to precise results. While the results are less precise in the up-down and front-back dimensions (1.5, 1.6 mm), this precision is still less than 10% of the grid and object size.

The errors show three biases: in the vertical plane an upward bias of $+3.5$ mm, and in the horizontal plane a frontward bias of -7.4 mm and a rightward bias of $+4.7$ mm. Of these, the frontward bias can be explained as a standard virtual object depth underestimation error, related to the way that HoloLens 1st generation virtual objects often appear to float above surfaces. This is likely related to the luminance of the virtual AR objects, which almost always appear brighter than background objects—if virtual objects are dimmer than background objects, then they become ghost-like and hard to see, and the display's brightness is increased. The underestimation effect is likely driven by the brightness biasing binocular vergence inwards, as reported by Singh et al. [28]. If this hypothesis is correct,
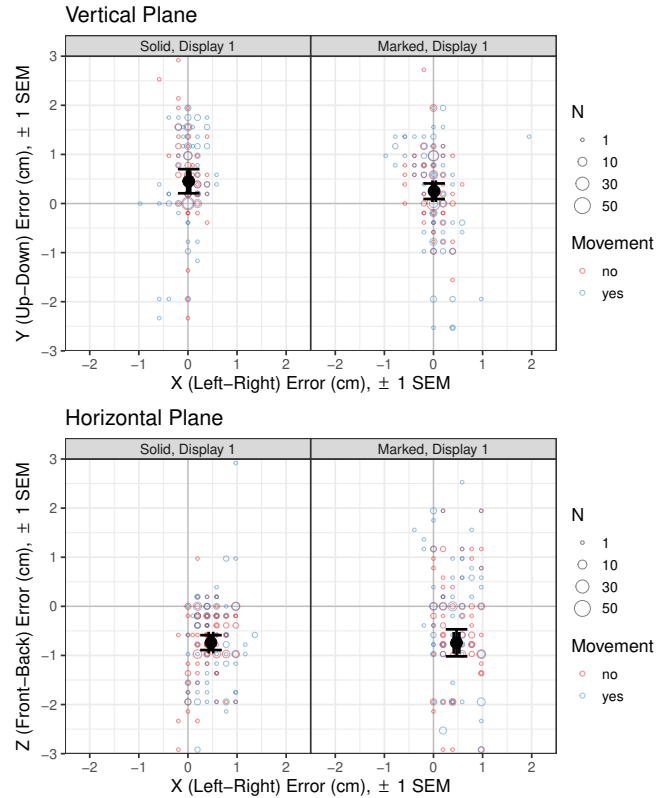

Figure 4: Experiment 1 error results. Errors are shown as hollow circles, sized according to the number of errors at that location, and colored according to movement. The black dot indicates mean error, with one standard error in each dimension. The results are separated according to cube style. On the vertical plane, there was an overall upwards bias of $+3.5$ mm, while on the horizontal plane there was an overall rightward bias of $+4.7$ mm, and an overall frontward bias of -7.4 mm.

then a repeated experiment with a monocular condition would show less of an underestimation bias. The upward bias on the vertical plane is also likely explained as an underestimation bias; all of the participants' eyes were well above the vertical grid (Figure 1), meaning that underestimated virtual objects would appear higher than intended.

If both of these biases—the upward bias on the vertical plane and the frontward bias on the horizontal plane—are explained by the virtual object appearing to float above the surface, then there should be a correlation between these biases and the probability of a quality judgment of *above* the surface, compared to *on* the surface. Figure 6 shows this correlation, in the form of error density plots separated into quality (above, on) and plane (vertical, horizontal). For the vertical plane, it is clear that the density associated with *above* ratings lies farther in the $+y$ direction than the density associated with *on* ratings. Likewise, for the horizontal plane, the density associated with *above* ratings lies farther in the $-z$ direction than the density associated with *on* ratings. The significance of these correlations was tested with logistic regressions, which predicted the quality judgement (above, on) from the error (y, z). For the vertical plane, the correlation was significant ($\chi_1^2 = 9.6, p < 0.01$); for every additional centimeter in error along $+y$, the log odds of choosing *above* over *on* increased by $\beta = 0.30$. And for the horizontal plane, the correlation was also significant ($\chi_1^2 = 63, p < 0.001$); for every additional centimeter in error along $-z$, the log odds of choosing *above* over *on* increased by $\beta = 0.79$. Therefore, both correlations
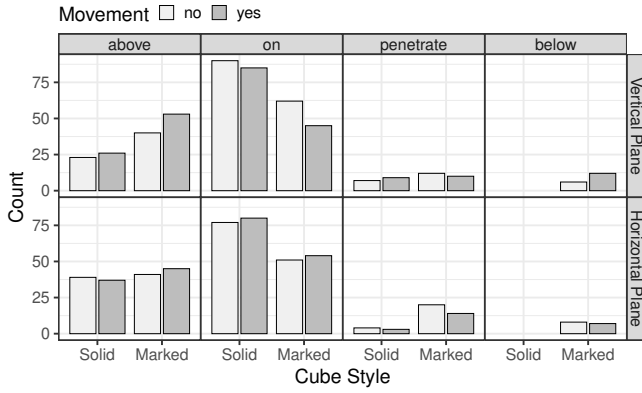
Figure 5: Experiment 1 quality rating counts for the solid and marked cube in the vertical and horizontal planes: *above*, *on*, *penetrating*, or *below* the surface.
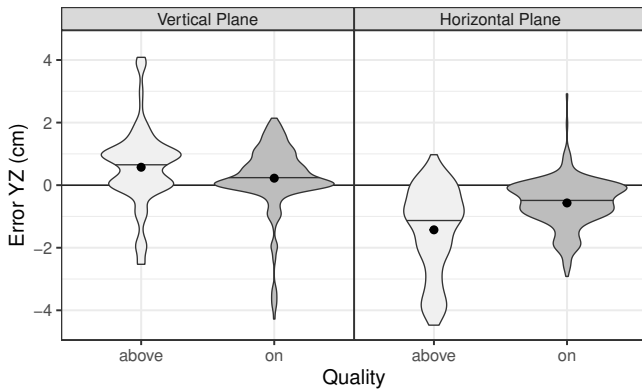

Figure 6: Experiment 1 error density for the vertical plane *y* axis and the horizontal plane *z* axis, according to the quality ratings *above* and *on*. The line shows the median value; the dot shows the mean value.

are significant in the expected directions, and consistent with the hypothesis that the upward bias in the vertical plane and the forward bias in the horizontal plane are explained by the tendency of the virtual object appearing to float above the surface.

This leaves the rightward bias on the horizontal plane. The effect is strong; in addition to the clear effect on mean errors in Figure 4, note that almost all individual error points fall to the right of the 0 line along the left-right axis. And, as illustrated by Figure 3, the effect becomes stronger in the rightward direction. Despite many discussions among themselves and with other colleagues, the authors were not able to explain this rightward bias.

5   EXPERIMENT 2: DISPLAY COMPARISON

In Experiment 1, a rightward bias was observed for the horizontal plane, which could not be adequately explained. It is possible that the bias could be due to some issue that is specific to the HoloLens display that was used, such as damage leading to an optical misalignment. The purpose of Experiment 2 was to test this hypothesis, by comparing the HoloLens from Experiment 1 (display 1) to a second, otherwise identical HoloLens 1st generation display (display 2). If the rightward bias was due to an issue that was specific to display 1, then it would no longer appear in display 2. The alternative finding would suggest that the HoloLens 1st generation display exhibits a systematic rightward bias on horizontal surfaces.

5.1   Method

Experiment 2 utilized two different HoloLens devices, display 1 and display 2, which were identical 1st generation models. As shown in Figure 2, there was no movement, and only the marked cube was used. Otherwise, the method was identical to Experiment 1. Therefore, there were two different independent variables: *plane* and *display*. Each participant completed 2 (*plane*: vertical, horizontal) × 2 (*display*: display 1, display 2) × 10 (*intended location*) = 40 trials. All variables varied within each participant. Presentation order was again counterbalanced between participants with a 4 × 4 balanced Latin square. Participants wore one display for the first half of the experiment, and the other display for the second half of the experiment.

Experiment 2 collected data from 16 participants, also chosen from among the students and staff at Mississippi State University. Among these participants, the same participant who completed Experiment 1 twice also completed Experiment 2 twice, in two separate sessions. For Experiment 2, this duplication was also not discovered in time to be corrected for the reported analysis, but the duplication is not expected to change the findings. Other than this participant, there was no overlap between the participants for Experiment 2 and Experiment 1. Therefore, data was collected from 15 separate people.

5.2   Results

Figure 7 shows intended versus perceived locations. As with Experiment 1, the results on the vertical plane show a fair amount of up-down bias, but very little left-right bias. In contrast, the rightward bias of the horizontal plane remains strikingly evident.

Figure 8 shows the perceptual errors between intended and perceived locations. On the vertical plane, for the *x* (left-right) judgements there is a marginal effect of display ($F_{1,15} = 4.1, p = 0.061$). Display 1 had a precision of 0.3 mm SEM, and no observable left-right bias ($F_{1,22} < 1$); while display 2 had a precision of 0.5 mm SEM, and a significant leftward bias of -1.5 mm ($F_{1,15} = 8.3, p < 0.05$). However, compared to the dimensions of the grid and cube, and the size of other significant biases in both experiments, this leftward bias is small and may not prove reliable. For the *y* (up-down) judgements, although there is a marginal effect of display ($F_{1,15} = 3.7, p = 0.07$), neither display differs from 0 (display 1: $F_{1,15} < 1$; display 2: $F_{1,15} = 1.5, p = 0.25$). The overall *y* precision is 1.3 mm SEM.

For the horizontal plane, the *x* (left-right) judgements have a precision of 0.7 mm SEM, with no effect of display ($F_{1,15} = 1.8, p = 0.20$). The overall accuracy still shows a rightward bias of $+3.7$ mm, which significantly differs from 0 ($F_{1,15} = 22.27, p < 0.001$). For the *z* (front-back) judgements, there is a significant effect of display ($F_{1,15} = 4.7, p < 0.05$). Display 1 has a precision of 2.5 mm SEM, and a -5.6 mm frontward bias, which significantly differs from 0 ($F_{1,15} = 4.8, p < 0.05$). However, display 2, with a similar precision of 2.9 mm SEM, does not significantly differ from 0 ($F_{1,15} < 1$).

Figure 9 gives the quality rating counts. The results generally mirror those in Figure 5: *penetrating* and *below* ratings were generally rare. For the remaining ratings, *above* and *on*, there is a clear interaction with grid plane. For the vertical plane, the cube was rated as being on the surface twice as often as floating above, but for the horizontal plane, the rating counts were very similar. The display used had little effect.

5.3   Discussion

The purpose of Experiment 2 was to determine if the rightward bias for the horizontal plane demonstrated in Experiment 1 was due to an issue that was specific to display 1. However, the results still showed a rightward bias of $+3.7$ mm. Although smaller than the previous rightward bias ($+4.7$ mm), the result remains, and was replicated across two displays and 40 participants. Therefore, the conclusion is
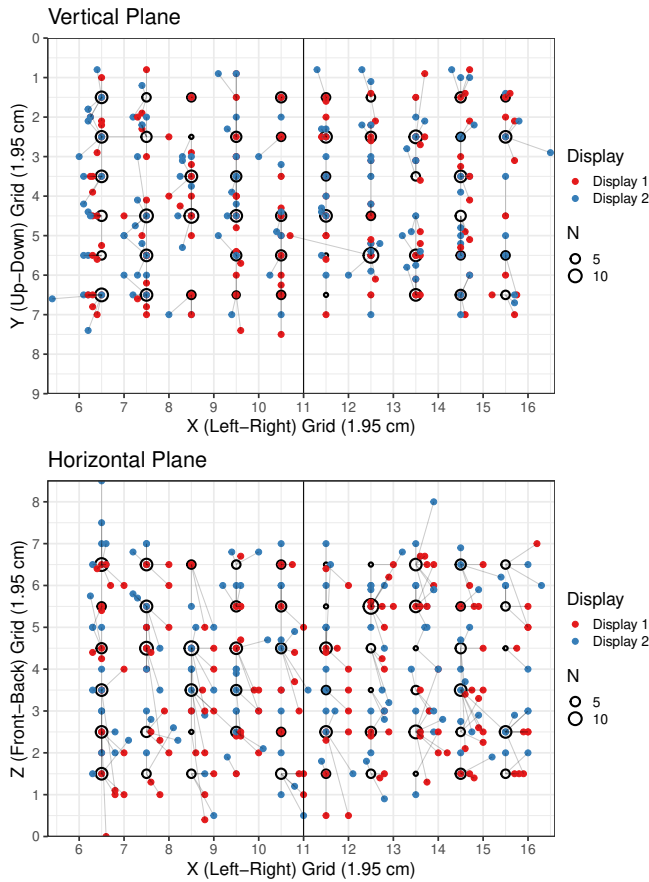
Figure 7: Experiment 2 intended vs. perceived location. See the caption for Figure 3. On the vertical plane, most perceived location errors occurred along the *y* (up-down) axis. On the horizontal plane, the rightward bias was again evident.

that HoloLens 1st generation displays have a systematic rightward bias.

The frontward underestimation bias for Experiment 1 was -7.1 mm. In Experiment 2, display 1 again demonstrated a frontward underestimation bias of -5.6 mm, but display 2 did not show any front-back bias. Therefore, in a direct comparison between the displays, display 2 exhibited less underestimation error. This, combined with the lack of an upwards bias for display 2, is consistent with the idea that display 2 is more accurate in the depth dimension.

However, display 2 still exhibits underestimation errors: note the points in Figure 8 that lie in the $+y$ and $-z$ portions of the plane. Display 1 also exhibits these errors. Therefore, the correlation between $+y$ and $-z$ and the probability of a quality judgment of *above* compared to *on* the surface was also examined. Figure 10 shows this correlation. Again, for the vertical plane the density associated with *above* ratings lies farther in the $+y$ direction than the density associated with *on* ratings, and for the horizontal plane, the density associated with *above* ratings lies farther in the $-z$ direction than the density associated with *on* ratings. The significance of these correlations was tested with logistic regressions, which again predicted the quality judgement (above, on) from the error (y, z). For the vertical plane, the correlation was significant ($\chi_1^2 = 7.0, p < 0.01$); for every additional centimeter in error along $+y$, the log odds of choosing *above* over *on* increased by $\beta = 0.41$. And for the horizontal plane, the correlation was also significant ($\chi_1^2 = 31, p < 0.001$); for every additional centimeter in error along $-z$, the log odds of choosing *above* over *on* increased by $\beta = 0.69$. Therefore,
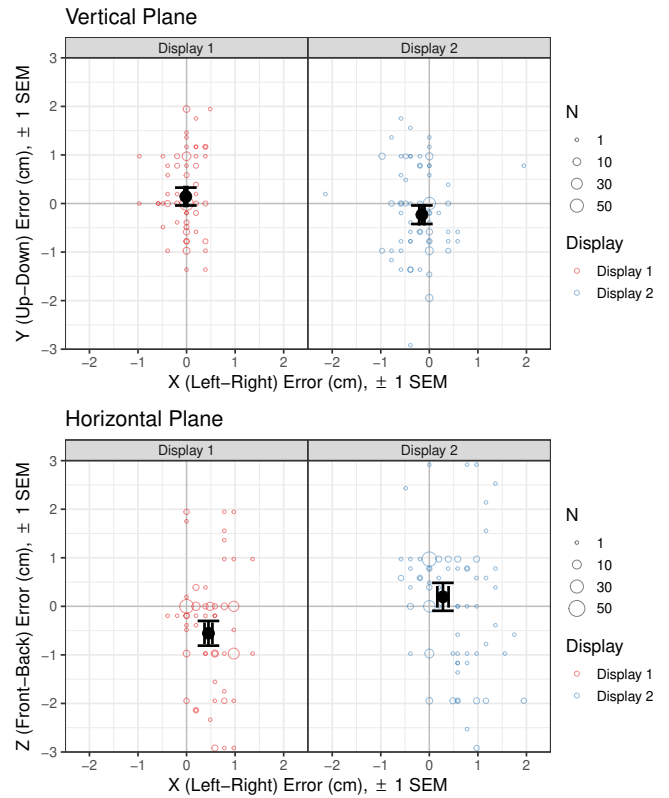

Figure 8: Experiment 2 error results. See the caption for Figure 4. On the vertical plane, overall the judgements were generally accurate. On the horizontal plane, there was an overall rightward bias of $+3.7$ mm. Only display 1 had a frontward bias of -5.6 mm.

Experiment 2 continued to support the hypothesis that the upward bias in the vertical plane and the frontward bias in the horizontal plane are explained by the tendency of the virtual object appearing to float above the surface.

Precision remained high in Experiment 2, less than 1 mm along the left-right axis, and only 1.3 mm along the up-down axis. Compared to Experiment 1, precision declined along the front-back axis (2.5, 2.9 mm), but was still less than 15% of the grid and object size.

6 GENERAL DISCUSSION

Across both of these experiments, the precision was high, ranging from 0.2 to 2.9 mm, for grid coordinates and a virtual object with a dimension of 1.95 cm. In addition, the authors found the method to be effective and flexible, and able to successfully evaluate the proposed research questions. These findings argue for the utility of the proposed perceptual measurement method. The first experiment also found no effect of cube style, which is consistent with the hypothesis that the perceptual measurement method may generalize to more visually complex virtual objects. The first experiment also found no effect of movement, suggesting robust HoloLens tracking.

Anchoring: The construction of the experimental program resulted in some knowledge about the difficulty of coding in a HoloLens environment. Object anchoring and grid alignment represented a significantly challenging technical task, despite the support of the Unity and Vuforia applications. More recently, the authors have experienced anchoring virtual objects in 2nd generation devices, the Magic Leap and the HoloLens 2nd generation. Both devices have much better anchoring support. In particular, the HoloLens 2nd
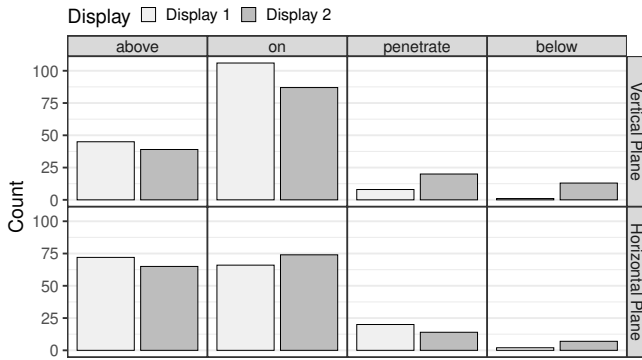
Figure 9: Experiment 2 quality rating counts for two different displays, in the vertical and horizontal planes: *above*, *on*, *penetrating*, or *below* the surface.
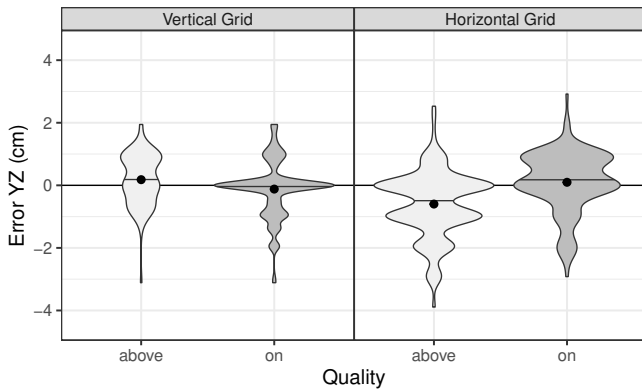

Figure 10: Experiment 2 error density for the vertical plane y axis and the horizontal plane z axis, according to the quality ratings *above* and *on*. See the caption for Figure 6.

generation allows printing an anchor point fiducial. When positioned in the real world, a HoloLens anchor point is established, with much more precision than was possible with the 1ˢᵗ generation devices used in these experiments. The functionality appears similar to what has been reported with the Vuforia and HoloLens tracking modes reported here.

Rightward Bias: Both experiments found a rightward bias on the horizontal plane, which is consistent with the hypothesis that the HoloLens 1ˢᵗ generation display has a systematic rightward bias. Although the authors cannot definitively explain this rightward bias, an anonymous reviewer proposed a possible explanation that had not been previously considered: the rightward bias could be explained by eye dominance. In a meta-analysis of eye dominance studies with large samples, Porac and Coren [25] report that 97% of observers favor an eye, with 65% sighting with the right eye, and 32% sighting with the left eye. Although eye dominance was not measured in Experiment 1, it was measured in Experiment 2. Of the 15 unique participants in Experiment 2, 14 were right-eye dominant, and only one was left-eye dominant. Therefore, the hypothesis is that when completing the task on the horizontal plane, participants primarily sight with their dominant eye, which biases the perceived location in the direction of that eye. While the eye dominance distribution in Experiment 2, which was unplanned, does not allow this hypothesis to be tested, a future experiment could test it by recruiting approximately equal numbers of right- and left-eye dominant participants.

Depth Errors: Another interesting result observed in both experiments is that the z-axis depth dimension was the most problematic for participants. This was observed both in post-experiment surveys, where participants frequently noted the difficulty of depth judgments, and in the results, where errors in depth were the largest. In addition, evidence was found that y-axis upward errors were also related to depth. These finding are consistent with the large body of previous work that has found errors in depth judgements of virtual objects.

In addition, significant correlations were found between these depth errors and an increased probability of seeing the virtual object as floating above the surface. Given that depth is sensed binocularly, it seems likely that depth errors are related to incorrect eye vergence [28]. And yet, the perceptual depth dimension remains critically important for many AR applications in reaching space, such as aligning real and virtual information in medical contexts [7]. Techniques for measuring perceived location in all three dimensions, such as the one presented here, may help in further explaining how eye movements contribute to errors in perceived depth. If properly explained, solutions leading to greater perceptual accuracy become more likely.

7 FUTURE WORK

These experiments used the built-in HoloLens 1ˢᵗ generation inter-pupillary distance calibration procedure. Recently, Hu et al. [12] introduced two calibration methods optimized for modern commercial OST AR HMDs, and tested them with a HoloLens 1ˢᵗ generation display. These methods replace the observer's eye with a camera, and compensate for parallax errors introduced by the camera's generic viewpoint. Hu et al. [12] reports better performance than both SPAAM calibration and the built-in HoloLens procedure. As Hu et al. was not available when this work was conducted, it would be worthwhile to replicate these experiments, while substituting the Hu et al. procedure for the HoloLens calibration procedure.

There are plans to replicate the presented method of measuring perceived three-dimensional location. These plans include testing the method on additional modern optical see-through reality displays, including the HoloLens 2ⁿᵈ generation and the Magic Leap One, and employing both a monocular condition and eye tracking to better understand the impact of eye movements on perceptual errors. In addition, participants will be screened by eye dominance, in order to examine an approximately equal number of right-eye and left-eye dominant people.

REFERENCES

[1] M. Axholt. "Pinhole camera calibration in the presence of human noise". PhD thesis. Linköping University, Nov. 2011.

[2] G. Ballestin, M. Chessa, and F. Solari. "A registration framework for the comparison of video and optical see-through devices in interactive augmented reality". *IEEE Access* 9 (2021), pp. 64828–64843. DOI: 10.1109/ACCESS.2021.3075780.

[3] M. Billinghurst, A. Clark, and G. Lee. "A survey of augmented reality". *Foundations and Trends in Human-Computer Interaction* 8.2-3 (Mar. 2015), pp. 73–272. DOI: 10.1561/1100000049.

[4] C. Diaz, M. Walker, D. A. Szafir, and D. Szafir. "Designing for depth perceptions in augmented reality". *IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*. 2017, pp. 111–122. DOI: 10.1109/ISMAR.2017.28.

[5] A. Dünser, R. Grasset, and M. Billinghurst. *A survey of evaluation techniques used in augmented reality studies*. TR-2008-02. Human Interface Technology Laboratory New Zealand, Sept. 1, 2008, pp. 1–27.

[6] F. El Jamiy and R. Marsh. "Survey on depth perception in head mounted displays: Distance estimation in virtual reality, augmented reality, and mixed reality". *IET Image Processing* 13.5 (2019), pp. 707–712. DOI: 10.1049/iet-ipr.2018.5920.

[7] M. Fischer, C. Leuze, S. Perkins, J. Rosenberg, B. Daniel, and A. Martin-Gomez. "Evaluation of different visualization techniques for perception-based alignment in medical AR". *IEEE International Symposium on Mixed and Augmented Reality Adjunct (ISMAR-Adjunct)*. 2020, pp. 45–50. DOI: 10.1109/ISMAR-Adjunct51615.2020.00027.

[8] S. S. Fukusima, J. M. Loomis, and J. A. Da Silva. "Visual perception of egocentric distance as assessed by triangulation." *Journal of experimental psychology: Human perception and performance* 23.1 (1997), pp. 86–100. DOI: 10.1037//0096-1523.23.1.86.

[9] H. C. Gagnon, C. S. Rosales, R. Mileris, J. K. Stefanucci, S. H. Creem-Regehr, and R. E. Bodenheimer. "Estimating distances in action space in augmented reality". *ACM Transactions on Applied Perception* 18.2 (May 2021). DOI: 10.1145/3449067.

[10] J. Grubert, Y. Itoh, K. Moser, and J. E. Swan II. "A survey of calibration methods for optical see-through head-mounted displays". *IEEE Transactions on Visualization and Computer Graphics* 24.9 (2018), pp. 2649–2662. DOI: 10.1109/TVCG.2017.2754257.

[11] J. Han, D. Farin, and P. H. N. de With. "A real-time augmented-reality system for sports broadcast video enhancement". *ACM International Conference on Multimedia (MM)*. Sept. 2007, pp. 337–340. DOI: 10.1145/1291233.1291306.

[12] X. Hu, F. R. Y. Baena, and F. Cutolo. "Alignment-free offline calibration of commercial optical see-through head-mounted displays with simplified procedures". *IEEE Access* 8 (2020), pp. 223661–223674. DOI: 10.1109/ACCESS.2020.3044184.

[13] J. A. Jones, J. E. Swan II, G. Singh, and S. R. Ellis. "Peripheral visual information and its effect on distance judgments in virtual and augmented environments". *ACM SIGGRAPH Symposium on Applied Perception in Graphics and Visualization (APGV)*. 2011, pp. 29–35. DOI: 10.1145/2077451.2077457.

[14] J. A. Jones, J. E. Swan II, G. Singh, E. Kolstad, and S. R. Ellis. "The effects of virtual reality, augmented reality, and motion parallax on egocentric depth perception". *ACM SIGGRAPH Symposium on Applied Perception in Graphics and Visualization (APGV)*. 2008, pp. 9–14. DOI: 10.1145/1394281.1394283.

[15] F. A. Khan, V. V. R. M. K. R. Muvva, D. Wu, M. S. Arefin, N. Phillips, and J. E. Swan II. "A method for measuring the perceived location of virtual content in optical see through augmented reality". *IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW)*. 2021, pp. 657–658. DOI: 10.1109/VRW52623.2021.00211.

[16] G. Klein and D. Murray. "Parallel tracking and mapping for small AR workspaces". *IEEE and ACM International Symposium on Mixed and Augmented Reality*. 2007, pp. 225–234. DOI: 10.1109/ISMAR.2007.4538852.

[17] B. C. Kress and W. J. Cummings. "Invited paper: Towards the ultimate mixed reality experience: HoloLens display architecture choices". *SID Symposium Digest of Technical Papers* 48.1 (2017), pp. 127–131. DOI: 10.1002/sdtp.11586.

[18] M. A. Livingston, Z. Ai, J. E. Swan II, and H. S. Smallman. "Indoor vs. outdoor depth perception for mobile augmented reality". *IEEE Virtual Reality Conference*. 2009, pp. 55–62. DOI: 10.1109/VR.2009.4810999.

[19] J. J. Marotta and M. A. Goodale. "The role of familiar size in the control of grasping". *Journal of Cognitive Neuroscience* 13.1 (2001), pp. 8–17. DOI: 10.1162/089892901564135.

[20] Microsoft. *Improve visual quality and comfort*. Sept. 2019. URL: https://docs.microsoft.com/en-us/hololens/hololens-calibration (visited on 05/27/2021).

[21] K. Moser, Y. Itoh, K. Oshima, J. E. Swan II, G. Klinker, and C. Sandor. "Subjective evaluation of a semi-automatic optical see-through head-mounted display calibration technique". *IEEE Transactions on Visualization and Computer Graphics* 21.4 (2015), pp. 491–500. DOI: 10.1109/TVCG.2015.2391856.

[22] K. R. Moser. "Towards system agnostic calibration of optical see-through head-mounted displays for augmented reality". PhD thesis. Mississippi State University, Aug. 2016.

[23] K. R. Moser, M. S. Arefin, and J. E. Swan II. "Impact of alignment point distance and posture on SPAAM calibration of optical see-through head-mounted displays". *IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*. 2018, pp. 21–30. DOI: 10.1109/ISMAR.2018.00025.

[24] V. V. R. M. K. R. Muvva. "Subject analysis of depth perception in augmented reality through Vuforia and HoloLens tracking". MS thesis. Mississippi State University, June 2019.

[25] C. Porac and S. Coren. "The dominant eye". *Psychological Bulletin* 83.5 (1976), pp. 880–897. DOI: 10.1037/0033-2909.83.5.880.

[26] C. S. Rosales, G. Pointon, H. Adams, J. Stefanucci, S. Creem-Regehr, W. B. Thompson, and B. Bodenheimer. "Distance judgments to on- and off-ground objects in augmented reality". *IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*. 2019, pp. 237–243. DOI: 10.1109/VR.2019.8798095.

[27] E. Rosten and T. Drummond. "Machine learning for high-speed corner detection". *Computer Vision – European Conference on Computer Vision ECCV*. Lecture Notes in Computer Science. 2006, pp. 430–443. DOI: 10.1007/11744023_34.

[28] G. Singh, S. R. Ellis, and J. E. Swan II. "The effect of focal distance, age, and brightness on near-field augmented reality depth matching". *IEEE Transactions on Visualization and Computer Graphics* 26.2 (Sept. 2018), pp. 1385–1398. DOI: https://doi.org/10.1109/TVCG.2018.2869729.

[29] J. E. Swan II, G. Singh, and S. R. Ellis. "Matching and reaching depth judgments with real and augmented reality targets". *IEEE Transactions on Visualization and Computer Graphics* 21.11 (2015), pp. 1289–1298. DOI: 10.1109/TVCG.2015.2459895.

[30] M. Tuceryan and N. Navab. "Single point active alignment method (SPAAM) for optical see-through HMD calibration for AR". *IEEE and ACM International Symposium on Augmented Reality (ISAR)*. 2000, pp. 149–158. DOI: 10.1109/ISAR.2000.880938.

[31] F. Wientapper. "Optimal spatial registration of SLAM for augmented reality". PhD thesis. Technische Universität Darmstadt, 2019.

[32] B. Williams, G. Klein, and I. Reid. "Automatic relocalization and loop closing for real-time monocular SLAM". *IEEE Transactions on Pattern Analysis and Machine Intelligence* 33.9 (Sept. 2011), pp. 1699–1712. DOI: 10.1109/TPAMI.2011.41.