Gaze Direction Visualization Techniques for Collaborative Wide-Area Model-Free Augmented Reality

Yuan Li
Center for HCI, Dept. of Computer Science, Virginia Tech
Blacksburg, Virginia, United States
yli92@vt.edu

Feiyu Lu
Center for HCI, Dept. of Computer Science, Virginia Tech
Blacksburg, Virginia, United States
feiyulu@vt.edu

Wallace S. Lages
Center for HCI, School of Visual Arts, Virginia Tech
Blacksburg, Virginia, United States
wlages@vt.edu

Doug A. Bowman
Center for HCI, Dept. of Computer Science, Virginia Tech
Blacksburg, Virginia, United States
dbowman@vt.edu

Figure 1: Gaze ray visualizations in AR. Top: the single ray without occlusion cues is visually ambiguous. Bottom: The double ray and parallel bars are designed to reduce visual and spatial ambiguity. We add the red highlight to indicate the target column for the readers, but note that this is not visible to the AR viewer.

ABSTRACT

In collaborative tasks, it is often important for users to understand their collaborator's gaze direction or gaze target. Using an augmented reality (AR) display, a ray representing the collaborator's gaze can be used to convey such information. In wide-area AR, however, a simplistic virtual ray may be ambiguous at large distances, due to the lack of occlusion cues when a model of the environment is unavailable. We describe two novel visualization techniques designed to improve gaze ray effectiveness by facilitating visual matching between rays and targets (Double Ray technique), and by providing spatial cues to help users understand ray orientation (Parallel Bars technique). In a controlled experiment performed in a simulated AR environment, we evaluated these gaze ray techniques on target identification tasks with varying levels of difficulty. The experiment found that, assuming reliable tracking and an accurate collaborator, the Double Ray technique is highly effective at reducing visual ambiguity, but that users found it difficult to use the spatial information provided by the Parallel Bars technique. We discuss the implications of these findings for the design of collaborative mobile AR systems for use in large outdoor areas.

SUI '19, October 19–20, 2019, New Orleans, LA, USA

ACM ISBN 978-1-4503-6975-6/19/10…$15.00
https://doi.org/10.1145/3357251.3357583

CCS CONCEPTS

• **Human-centered computing → Mixed / augmented reality**; **Computer supported cooperative work**; *Virtual reality*; *User interface design.*

KEYWORDS

Augmented Reality; Collaboration; Awareness; Gaze Visualization.

ACM Reference Format:
Yuan Li, Feiyu Lu, Wallace S. Lages, and Doug A. Bowman. 2019. Gaze Direction Visualization Techniques for Collaborative Wide-Area Model-Free Augmented Reality. In *Symposium on Spatial User Interaction (SUI '19), October 19–20, 2019, New Orleans, LA, USA.* ACM, New York, NY, USA, 11 pages. https://doi.org/10.1145/3357251.3357583

1 INTRODUCTION

Joint attention on an object of mutual interest is a common requirement in many collaborative tasks [3, 4, 6] and typically requires the communication and confirmation of an object's location [29]. For example, when surveyors want to triangulate a distant target, they need to agree on the location of the target even when they are spatially separated. Similarly, a firefighter on the ground may need to pinpoint the location where the pilot of a firefighting aircraft should drop chemicals on a wildfire.

Communicating an object's spatial location is relatively easy when two collaborators are close to each other and the target. One only needs to point to the target and confirm pointing via vocal communication. At greater distances, collaborators can use tools such as laser pointers to indicate targets. In this case, the observing collaborator needs to visually search for the laser dot that indicates the point of intersection between the pointing vector and the target object. Since laser pointers only provide a small visual marker, and since there is no visual information connecting the pointing user to the target, this approach is difficult to use at very large distances. We suggest that augmented reality (AR) displays can be used to communicate pointing direction and target information from one collaborator to another. If the relative position and orientation of the two collaborators is known, a ray indicating pointing direction can be drawn on the collaborator's display.

The same idea is widely adopted in Virtual Reality (VR). Many collaborative VR applications use pointing rays and highlight objects intersected by these rays to facilitate visual consensus [2]. The pointing ray can be specified accurately by various input devices, such as hand trackers, head trackers, and eye trackers [18, 22]. Head tracking is supported in all VR systems, and only requires the user to center the target in his view, so determining a rough gaze direction based on head orientation is a common approach. In this paper, we use the term "gaze ray" to refer to a visualization of pointing direction based on head orientation.

Unfortunately, applying the gaze ray technique in AR is not as easy as in VR. To correctly display a 3D virtual ray, the AR system needs a geometric model of the real environment so that the ray can properly occlude, be occluded by, and intersect with the real world. Environment models can be obtained in many AR scenarios, through a combination of pre-built models, on-board sensors, user interaction, and 3D reconstruction software [26]. However, environment information is not always available or reliable, especially in the large, outdoor environments that are used for wide-area AR. When no environment model is available (so-called "model-free" AR [7]), the AR system can only visualize a virtual ray that appears to be overlaid on everything along its path. Thus, the virtual ray does not appear to intersect with the target object and provides false occlusion cues, as shown in Figure 1a.

While model-free AR may not be a common use case, it is likely to occur in high-stakes scenarios under conditions of extreme uncertainty. These include scenarios in which previous surveying of the environment is not feasible, such as rescuing stranded people in the wilderness, marking points for air delivery of supplies, or obtaining coordinates of survivors after a natural disaster. Similarly, there are situations when an accurate environment model is not feasible to obtain due to the dynamic nature of the scenario, such as when police in a crowded public location need to spot potential threats among moving people.

If the target object is visually isolated in the environment, the AR gaze ray technique can still be used to convey the location of the target, since the observer can simply search for the object that the ray crosses visually. However, when the ray appears to cross multiple objects in the environment, visual ambiguity (VA) occurs. Even in the presence of VA, the observer may still be able to determine the target object if he can understand the position of the pointing collaborator and the orientation of the gaze ray, and match the object's spatial location along it. However, since occlusion is the most dominant depth cue [8] and most other depth cues are missing or ineffective with a simple ray at a large distance, correct perception of the gaze ray orientation is difficult. If the target object is spatially isolated in the environment, ray orientation perception does not need to be highly accurate. However, if the target is closely surrounded by other objects, even if the user can perceive ray orientation to some extent, it is not easy to make the spatial judgment. In this case, spatial ambiguity (SA) is added to the problem of VA.

In our research, we focus on the visual and spatial understanding of the observing collaborator (we leave the issue of accurate specification of gaze rays by the pointing collaborator for future work). We describe the design of two AR gaze ray visualization techniques that aim to improve collaborative gaze awareness by addressing these two types of ambiguity. The Double Ray technique is designed to reduce VA by requiring the pointing collaborator to indicate two geometric features of the target simultaneously. The assumption is that the chances of bracketing multiple objects with two rays is lower than crossing multiple objects with a single ray, as shown in the top image of Figure 1. The Parallel Bars technique is designed to provide extra cues about the orientation of the gaze ray to facilitate orientation perception and thereby attenuate SA, as illustrated in the bottom image of Figure 1.

To understand the performance of our techniques in terms of accuracy and speed on a target identification task, we performed a controlled user study in a simulated AR environment where gaze rays are assumed to be perfectly accurate. The results show that the Double Ray technique is more robust in the presence of VA, but that there are still a small number of visually ambiguous situations even with this technique. The Parallel Bars technique did not provide increased accuracy in the presence of SA, and users take more time to interpret the additional spatial cues. We discuss the implications of these findings for the design of collaborative wide-area AR applications.

2 RELATED WORK

Here we review prior research about awareness needs in AR/VR collaboration, gaze visualization, and perception limitations.

2.1 Awareness Needs

Research on facilitating multi-user collaboration dates back to early synchronous shared window systems. Lauwers et al. [23] highlighted the concept of "collaboration awareness" and emphasized its importance in meeting usability requirements. Gutwin et al. [14] adopted the idea and developed techniques to highlight collaborator actions.

In VR and AR, collaborative awareness is needed to understand the perspectives of other users sharing the same virtual or real space. Researchers [16, 17, 27, 34, 37, 38] have reported that better understanding of other users' perspectives can lead to increased performance or usability in collaborative object-focused interaction tasks. For example, Tang et al. [34] concluded that knowing where one's collaborators are looking helps support effortless joint references, leading to easier collaboration.

2.2 Gaze Visualization

To facilitate understanding distributed collaborator's perspectives, there have been studies exploring various gaze visualization techniques. A commonly adopted method for providing gaze awareness is to determine the object one user is looking at, then to place a virtual element such as a dot or a cursor on that object in the other user's view [1, 29, 35]. Then, understanding joint object references only requires a visual search for the virtual element in the scene. While this method is easy to use, it it not applicable to model-free AR due to the lack of information about the real-world environment. Yet another approach is to present an image of one user's view to another user [12, 19]. Ideally after the viewer inspects the image, she can match it with the real-world scene and understand the situation. But there are issues in practice. First, because the device used to capture the first-person view might have different optical parameters than the human eye, matching the image to the real world is not an easy task. Second, given different viewer/collaborator location setups, one person's field of view could be very different from the other. We observed these limitations during our preliminary explorations and decided not to include this approach in this study.

Some systems use a radar view to facilitate collaboration awareness [10, 31]. In these systems, a top-down 2D view containing the collaborator's location and view direction is presented to users. An immediate problem for applying this technique in model-free AR is data loss when converting 3D gaze information to 2D display. A well designed 3D radar could possibly prevent data loss, but it is unclear how to design an effective exocentric visualization of gaze direction without a model of the environment to serve as context.

We instead adopt the approach of visualizing gaze direction as a ray [32]. Many implementations of ray casting can be found in the literature [24, 32, 36]. By using an input device with six-degree-of-freedom tracking, the user is able to provide a 3D ray from the device pointing directly to a target of interest. However, the simple ray technique is ambiguous in model-free wide-area AR due to missing occlusion cues.

2.3 Depth Perception Limitations

According to the taxonomy presented by Cutting & Vishton [8], human perceptual space can be divided into three areas: personal space (under 2 meters), action space (up to 30 meters), and vista space (beyond 30 meters). Only four depth cues are reasonably effective in vista space. Occlusion is the strongest depth cue. Relative size and density are less effective, because they only reveal larger changes in depth. Aerial perspective (i.e., fog) can only differentiate between objects with very different depths. When occlusion is not available, humans must rely on these less effective cues to judge depth.

Kruijff et al. [21] discussed the lack of occlusion cues in model-free AR systems. Virtual objects in model-free AR will occlude all real objects in the scene, making the virtual objects appear to be closer to the user. One of the main methods used to address the mis-occluding problem is wireframe visualization [11], which allows the user to see through the augmentation and perceive real-world content behind it as normal. But this method does not help the user to accurately perceive the correct depth relationship between the virtual and real content. Wireframe visualizations can even lead to the Necker Cube Illusion [20]. Researchers have explored other depth cues to improve perception accuracy. Diaz et al. [9] compared the effectiveness of shading, cast shadows, and texture in action space. Their study emphasized the importance of virtual-to-physical interactive cues (cast shadows) over pure virtual cues (texture) and physical-to-virtual cues (aerial perspective). To our knowledge, few studies have focused on this problem in far-field AR [33].

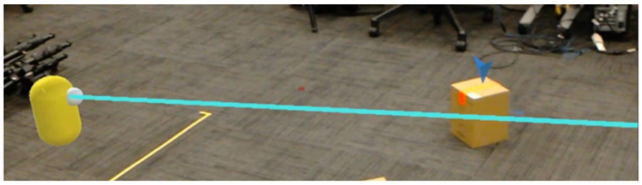
3 DESIGN OF GAZE RAY VISUALIZATION TECHNIQUES

A pointing ray is a powerful tool to reference objects in VR, because it effectively converts the spatial referencing task into a visual search task where the user is looking for the intersection between the ray and the referenced object. When an environment model is available, the pointing ray will interact with the environment by occluding and being occluded by objects in the scene, as shown in Figure 2a. However, without knowledge of the environment, which is often the case in wide-area outdoor AR, the gaze ray occludes everything in the scene, as illustrated in Figure 2b. As we discussed in the introduction, the absence of correct occlusion cues results in inaccurate perception of the ray's direction. Even when the user understands this limitation, gaze rays in model-free AR can be visually ambiguous when they cross multiple possible targets. When VA is present, users must attempt to understand the ray's orientation, but in the case where multiple objects are near the 3D ray, SA can further complicate the task of target identification. Our goal was to design enhanced gaze ray techniques that could overcome VA and SA.

3.1 Reducing Visual Ambiguity

VA occurs when a single ray visually crosses multiple objects in the scene (as seen in Figure 1a). We designed the Double Ray technique to address this issue. With Double Ray, a distant collaborator is asked to cast two rays pointing at two different geometric features

(a) Gaze ray with proper occlusion



(b) Gaze ray without proper occlusion

Figure 2: First person view of different gaze rays with or without occlusion in modle-free AR.

(typically the top and the bottom) of the target object simultaneously. The insight is that by increasing a number of rays aiming at different parts of the target, the visual matching condition is more strict than the single ray case. Instead of looking simply for objects over which a single ray passes, the user now looks for objects that are 'bracketed' by the two rays. In this way, the number of possible targets (and the VA) is significantly reduced, as shown in Figure 1b. The enhancement could be further strengthened by asking the distant collaborator to cast more rays at more features, although this would create more workload on the distant collaborator, would require more pre-agreement between the viewing user and the collaborator, and would rely on the availability of distinct geometric characteristics of the objects in the environment. Therefore, we focused on the Double Ray case to study its effectiveness while preserving maximum applicability.

3.2 Reducing Spatial Ambiguity

Even when VA is present, the viewer may still be able to understand which object is the target based on his perception of the virtual ray's orientation. But the accuracy of this perception depends on the availability of spatial cues. The strongest depth cue (occlusion) is not available in model-free settings. Size and density of the virtual ray relative to real world references are insufficient, since the ray is thin and of an unknown size. Given the thickness of the gaze ray, perspective cannot provide enough depth information either. Thus, we proposed to add artificial cues to help the viewer understand ray orientation.

While the AR system has little knowledge of the environment, the program must possess enough 3D information about the virtual ray to render it on the display. Therefore, we can exploit such information to help the user overcome SA by providing extra cues for figuring out the virtual ray's direction.

We explored various artificial cues. At first, we added virtual markers every three meters along the ray to create an artificial texture gradient effect. The markers would appear to cluster at further distances from the viewer and be sparse at closer distances. Another approach encoded the distance from the viewer with color saturation of the virtual ray to mimic aerial perspective. We also tried placing a virtual box at the point on the ray closest to the user. After we found through pilot testing that these cues were either ineffective or hard to learn, we switched to a dynamic cue in which we rendered a cube at the point on the ray where the user was looking. In this way, the user could look from left to right, and the cube would slide along the ray, providing perspective information through changes in size and apparent speed. We didn't continue with this dynamic approach because it required additional user input (looking left and right) and added visual complexity to the scene. In the end, we settled on a static technique that we named Parallel Bars.

In the Parallel Bars technique, based on the orientation of the gaze ray, we create multiple virtual bars that are all parallel to the single gaze ray (when combined with the Double Ray technique, the bars are aligned parallel to the central direction defined by both rays). The first bar is located roughly at the chest height of the observing user. We calculate the direction that is perpendicular to the gaze ray from the center of the first bar and then place the rest of the bars along that direction. Figure 4 provides a top-down view of the arrangement. In the AR version of Parallel Bars, we placed all bars at the same height as the first bar. However, in the AR simulation experiment described below, we kept the first bar at user's chest height but moved the other bars upward to avoid unintended enhancement from the experiment scene. The bars are separated by a preset fixed distance, as shown in Figure 1.

In this way, the user can both directly perceive the orientation of the gaze ray (by observing its parallel counterparts) and estimate the distance to the gaze ray (by observing the number of parallel bars). In our experiment, we tested different distance values between bars and settled on 20 meters as it provided good visual length variation under the experiment conditions without clustering.

3.3 AR Implementation

We implemented both the Double Ray and Parallel Bars techniques in a collaborative AR application using Microsoft HoloLens AR systems. Two users, each wearing a HoloLens, are able to interact with the system collaboratively. We incorporated a simple calibration procedure, where both HoloLens users place a shared virtual coordinate system at the same location via image recognition through the Vuforia API[1]. After that, users can see each other's avatar, whose position and orientation are sent through the network by UDP protocol 20 times per second and updated on both systems. Users can have two roles: the "pointer" who marks real-world objects with virtual rays, and the "observer" who observes the rays and identifies the target object.

The pointer casts the gaze ray by looking at the target and centering it in the view under an aiming reticle. With the Double Ray technique, the pointer can also use a joystick on the input device to adjust the vertical spread of the rays, which is indicated by two horizontal lines above and below the reticle. When the pointer is satisfied with both the direction and spread, he confirms the ray(s) by pressing a button on the input device (see Figure 3).

We tested the Double Ray and Parallel Bars techniques in both outdoor and indoor settings and saw their potential for improving collaborative target identification. However, we also found that current limitations of the AR display and tracking could significantly

[1] https://library.vuforia.com/articles/Training/Image-Target-Guide

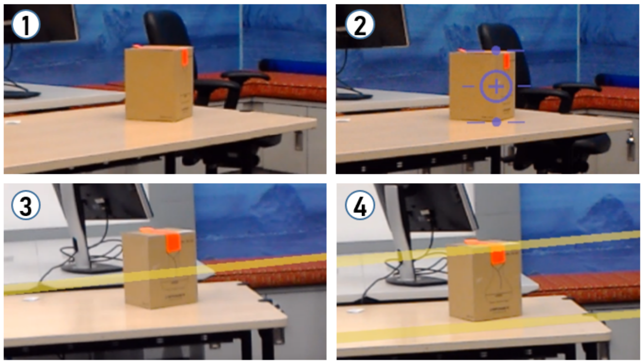


Figure 3: Defining rays in HoloLens implementation: 1) determine the target; 2) Overlay aiming reticle and top/bottom lines on the target; confirm with button press; 3) observer's view of Single Ray; 4) observer's view of Double Ray

affect the user experience and task performance. Therefore, to study our techniques in a valid and controlled fashion, we designed an experiment using simulated AR, as described in the next section.

4 EXPERIMENT

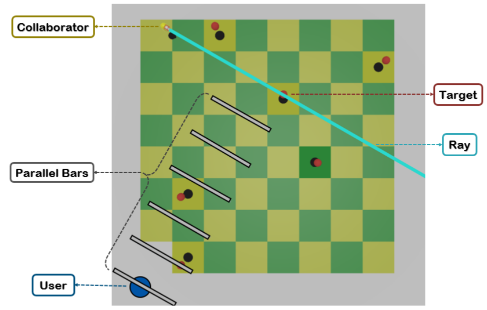
4.1 Goals

In order to evaluate the effectiveness of our gaze direction visualization techniques in conditions with varying VA and SA, we conducted a controlled experiment in which individual participants assumed the role of a passive observer who needed to identify the correct target object at which a simulated distant collaborator was currently gazing. Figure 4b shows an example of a trial in the experiment. The experiment used a simulated AR setting implemented in a VR system, both to avoid the limitations of current AR devices and to allow us to systematically control key features of the environment and task. This approach, known as Mixed Reality Simulation, has been used in a variety of prior AR experiments in which either experimental control was critical or technological limitations made the use of real AR systems impractical [5, 13, 25].

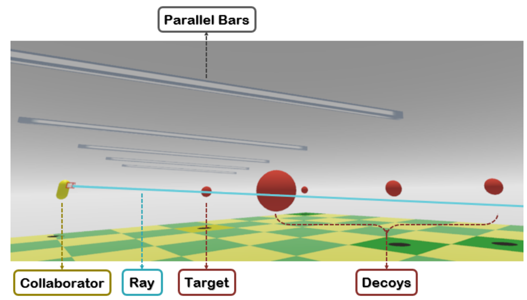
4.2 Environment and Task

The virtual environment designed for the experiment was an eight by eight square "chessboard" with a length of 138m. A virtual collaborator and the human user (participant) were located at two adjacent corners of the chessboard. The virtual collaborator was represented as a human-size yellow capsule with an orange-bordered cube on its forehead to indicate the forward direction. In each trial, there were six red spheres (possible targets) floating above six cells of the chessboard. One red sphere was chosen as the target, and the virtual collaborator turned to gaze at that sphere. The other five spheres served as decoys. The task was to correctly identify the sphere at which the virtual collaborator was looking.

There were four potential target locations, resulting in a 30-degree range of gaze directions. The decoys were carefully placed to be evenly spread across the participant's view, and the position of the decoys changed from trial to trial to avoid learning effects. Virtual spheres did not occlude one another from the participant's viewpoint. To improve the user's perception of the location of the spheres, they cast a shadow on the square over which they were



(a) Top-Down view of the Parallel Bars technique. The blue dot indicates the viewer's position.



(b) Participant's first person view.

Figure 4: Example of an experiment trial.

floating. The virtual collaborator rotated towards the target and cast gaze ray(s) from the orange box to the target. We assumed that the collaborator could cast perfectly accurate rays, so single rays pointed directly at the center of the target sphere, while double rays pointed exactly at the top and bottom of the target, resulting in a "bracketing" visual effect. A purple crosshair was fixed in the center of the participant's view and was used for selecting the sphere the participant believed to be the target in each trial.

The color patterns of the chessboard and cast shadows were designed to enhance perception of spatial position of the virtual spheres. However, the features of this virtual environment might also provide users with unrealistic advantages or unfair strategies. We took care to avoid this in our experimental design and implementation. For example, all parallel bars but the first one were displayed above the user so that the bars did not visually overlay the chessboard pattern.

Using a simulated collaborator with perfect pointing skill helped us to focus on the effectiveness of our proposed visualization techniques. If we designed the study with pairs of participants, then it would be unclear whether errors were due to the pointing skills of the users or the understandability of the visualization. Besides being able to better control the experiment, having a perfect collaborator could reveal the higher performance boundary.

4.3 Experiment Design

Our experiment followed a $2(rays) \times 2(bars) \times 3(VA) \times 2(SA)$ within-subjects design. The first two independent variables created four gaze ray conditions, while the last two created six task conditions. Participants used each of the four gaze ray techniques to complete a set of 24 trials (four trials in each of the six task conditions). We

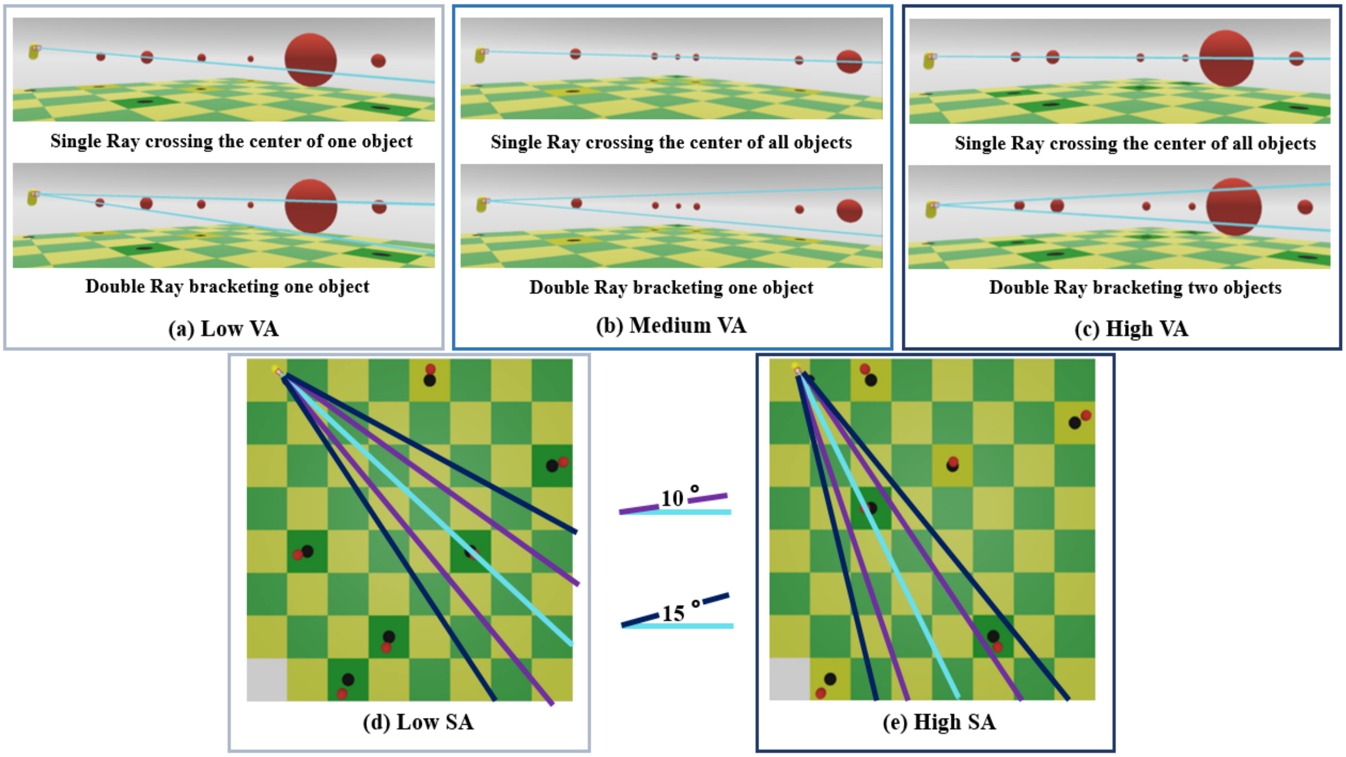


Figure 5: Levels of visual and spatial ambiguity used in the experiment.

gathered data on two objective measures: errors and task completion time on successful trials. We also measured subjective feedback through interviews. In the following subsections, we will explain the independent variables and dependent measures in detail.

4.3.1 Gaze Rays.

Based on the two enhanced visualization techniques we designed (Double Ray and Parallel Bars), we defined two independent variables with two levels each: *rays* (single or double) and *bars* (without or with). Thus, we evaluated a total of four techniques: single ray without bars (the baseline technique), single ray with bars, double ray without bars, and double ray with bars.

4.3.2 Task Conditions.

We also independently varied the levels of VA and SA, in order to evaluate the effectiveness of the techniques under different task conditions. Here we give more formal definitions of these concepts and how we controlled them in the experiment.

Visual Ambiguity (VA). In model-free AR, the observing user can use a gaze ray as a spatial referencing tool by looking for objects the ray crosses (or overlaps). This converts the spatial referencing task into a visual search task. If the target is isolated in the scene, finding the crossing is easy. However, in many real-world scenarios the gaze ray will cross multiple objects (e.g., when both collaborators are standing at ground level and targeting objects at ground level, as in Figure 1, top). To systematically study how VA affects the effectiveness of our gaze ray techniques, we defined three levels of VA as follows:

- Low VA: One and only one object is perfectly crossed by a single gaze ray (i.e., the single ray goes through the center of only one object), as shown in Figure 5a. We achieved this in the experiment by increasing the height of the virtual collaborator so that the gaze ray was not parallel to the ground, causing it to cross over the center of only one sphere.
- Medium VA: When a single ray is used, all objects are crossed perfectly (i.e., the single ray goes through the center of all objects), as shown in Figure 5b. However, when using the Double Ray, only one object is perfectly bracketed by the gaze rays.
- High VA: When a single ray is used, all objects seen by the user are crossed perfectly. When using the Double Ray, two objects are perfectly bracketed, as shown in Figure 5c.

Spatial Ambiguity (SA). Assuming that the observing user cannot easily determine the referenced object due to VA, another approach she can take is to judge the position of the collaborator and the orientation of the gaze ray in order to identify the target object by its spatial location. Assuming that the observing user can understand the ray's direction with reasonable accuracy, SA occurs when multiple targets are close to the ray. For the purposes of the experiment, we defined the region between 10 and 15 degrees away from the gaze ray as spatially ambiguous. The lower bound reasonably reduces the difficulty of the task and the upper bound ensures enough ambiguity. We defined two levels of SA:

- Low SA: No objects besides the target are within 15 degrees of the gaze ray (Figure 5d).

- High SA: One and only one object exists in the SA region (10-15 degrees away from the gaze ray; Figure 5e).

VA and SA were manipulated independently in the experiment, leading to a total of six (2×3) task conditions.

We chose four target locations on the chessboard. For each of the target locations, we created trials for all six task conditions with different sets of decoys, leading to a total of 24 (4×6) trials. Each participant repeated these trials four times (once with each of the four gaze ray techniques). We developed four randomized orderings of the 24 trials, and each participant experienced these orderings in sequence. Latin square counterbalancing was applied to the presentation of the techniques. In summary, each participant experienced 96 trials (combinations of VA, SA, and sphere layout) in the same order, but the order of the techniques varied from participant to participant.

4.3.3 Measures.

For each trial, we recorded whether or not the participant selected the correct target. For each combination of the independent variables (*rays*, *bars*, VA, and SA) we calculated the error rate in the range $[0, 4]$. Participants were not given feedback on the accuracy of their selections to avoid learning effects.

We also measured task completion time, defined as the amount of time taken to correctly identify a target. Participants were only allowed one selection per trial. We did not consider the time for failed trials, because we were interested in the amount of time needed for participants to *successfully* interpret the visual cues provided by the techniques, rather than simply the time it took to make a guess about which object was the intended target. Before each trial, only the virtual chessboard was visible. Participants started the timing manually by pressing a button to reveal the virtual collaborator's orientation, the gaze ray, and the spheres. The timer automatically stopped when a selection was made. No time limit was set for the trials.

Finally, we gathered subjective feedback through an interview. We asked participants about their perceptions of the usability of the four gaze ray techniques and their strategies for completing the task.

4.4 Hypotheses

We tested five hypotheses in the experiment:

H1. The Double Ray technique will result in fewer errors and lower task completion time than the Single Ray technique when VA is at the medium or high level. This hypothesis suggests that Double Ray will be effective at eliminating or decreasing the effects of visual ambiguity.

H2. For the Single Ray technique, both medium and high levels of VA will result in a higher error rate than low VA. But for the Double Ray technique, only the high level of VA will result in increased errors compared to the low level of VA. Given our definitions of medium and high VA, this difference should be expected, because by design Double Ray is immune to medium VA.

H3. The use of the Parallel Bars technique will decrease errors, but increase task completion time, in the presence of VA and low SA. We designed Parallel Bars to improve the user's understanding of the ray's orientation, so that even when VA exists it would be possible to guess which object is the target (note that for the Single Ray, VA exists at both the medium and high levels, while it only exists at the high level for the Double Ray technique). However, we expected that using the spatial information provided by Parallel Bars would require significant mental workload, thus increasing the task completion time. In addition, we surmised that Parallel Bars would only be effective at the low level of SA, when no spheres other than the target were within 15 degrees of the gaze ray. In the high SA condition, we expected that even with orientation information from the Parallel Bars, correctly identifying the target would still be very difficult.

H4. Double Ray will be perceived as a more usable enhancement than Parallel Bars. We expected that the visual enhancement of Double Ray could be used automatically and intuitively, since it only requires visual perception of bracketing, but that Parallel Bars would require extra mental workload to reason about the gaze ray direction.

H5. Users with higher spatial orientation ability will have more accurate task performance with the Parallel Bars technique. Spatial orientation includes the ability to imagine the appearance of objects from different locations [15]. This ability seems related to how we assume the user interprets the orientation of the 3D gaze ray through different cues including the parallel bars, in the sense that after the user observes the rough facing direction of the collaborator and the gaze ray, he may try to imagine what the scene looks like from the collaborator's point of view. We used a perspective taking test to measure participants' spatial orientation score [15]. In short, the test asks the participant to imagine facing in a certain direction and then to draw a line in the direction of a specific target. The score is obtained by averaging the angular errors for 12 trials (please refer to the original publication for details). Thus, a low score indicates high perspective taking ability on this test. We anticipated that the spatial orientation ability of participants would affect their accuracy in the experiment. More precisely, the difference should be even greater when using Parallel Bars techniques because participants with higher spatial ability could have a better understanding of the gaze ray's direction through the parallel bars and therefore have a better chance to correctly identify the target.

4.5 Apparatus

The experiment used a desktop PC running Windows 10 with 16GB RAM, 3.4GHz i7 CPU and a dedicated GTX1070 GPU. We used a consumer version HTC VIVE Pro HWD. The VIVE has two screens, each with a resolution of 1440×1600 pixels. The total horizontal field of view is 110 degrees. It was tracked with six degrees-of-freedom by the hybrid inertial-optical Lighthouse 2.0 system. We also used a wireless Xbox controller for input. We used the 'A' button to confirm a selection and the 'Y' button to start a trial. The software used in the experiment was written in Unity3D.

4.6 Participants and Procedure

We recruited 24 participants (11 females) between 20 and 41 years old (M = 26.42, SD = 1.07) from a local university. All of the participants were right-handed. Three of them did not have prior experience with VR/AR before the experiment.

The experiment was divided into six phases. In the first phase, participants were welcomed upon arrival and asked to read and

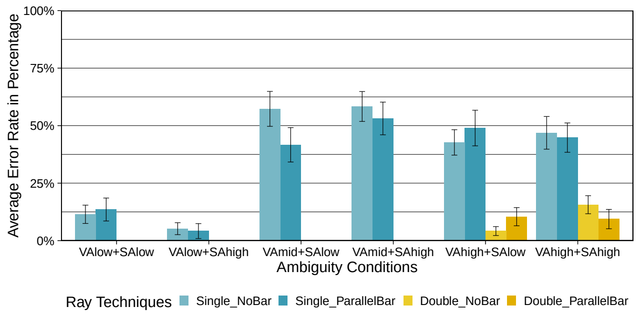


Figure 6: Average error rate for different ray techniques under different VA and SA combinations. Error bars represent standard error.

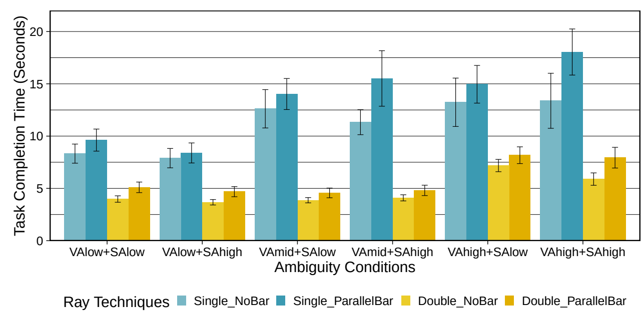


Figure 7: Successful task completion time for different ray techniques under different VA and SA combinations. Error bars represent standard error.

sign an informed consent form (the study was approved by the Institutional Review Board of the university). Second, they were asked to fill out a pre-study questionnaire to collect demographic information and prior experience with VR and AR. In the third phase, participants were asked to complete three pre-tests: (1) measurement of inter-pupillary distance in order to adjust the VR headset appropriately; (2) the perspective-taking test, which tested their spatial orientation abilities; (3) a stereoscopic vision test, which ensured their ability to perceive stereo information (all participants qualified).

After completing these initial steps, in the fourth phase, participants were given an introduction to our experiment background and the experiment setup. When participants had no further questions, we helped the participants put on the VR headset and asked them to stand on a virtual marker at a particular location. Throughout the experiment, we asked participants to stand still but did allow them to lean or crouch to use motion parallax cues. Before the testing of each technique, a training session was provided, where participants had to finish at least 12 trials to familiarize themselves with the technique and try to use it under each task condition (VA + SA). To help participants understand the techniques better, participants were allowed to see a top-down view of the current scenario at any time while in the training session. They were encouraged to ask the experimenter for clarification if they had any questions.

When participants reported they felt ready, the formal trials (fifth phase) began. The selected object and the time spent for each trial was automatically recorded by the program for later assessment. After participants finished all the trials for each technique, they were invited to rest briefly, followed by an interview asking for their subjective impressions of the technique. Finally in the sixth phase, after participants finished testing all four techniques, we asked some additional questions about the techniques and the enhancements, including how participants liked them and how they would rank them in terms of helpfulness and usefulness.

4.7 Results

Figure 6 plots the average error rate by task condition for the four ray techniques. There are 576 data points ($24(participants) \times 2(SA) \times 3(VA) \times 2(rays) \times 2(bars)$). Each data point is the number of errors made by one participant with a particular combination of *rays*, *bars*, VA, and SA. These values are in the range of $[0, 4]$ (because there were four repetitions of each combination); we converted these to percentages in Figure 6.

Figure 7 presents the average successful task completion time by task condition for the four ray techniques. At a glance, we can see that the successful task completion time shares some similarity to the error rate measurement. Since only successful trials were considered, there were some missing data points when the participant made an error in all four repetitions of a particular condition. To be more precise, we lost 36 data points for this reason, leaving 540 valid time data points.

We conducted a series of analyses to test our hypotheses. For H1, since we hypothesized benefits of the Double Ray at medium and high VA levels, we first aggregated the data for the medium and high VA conditions. Thus, in this analysis, VA became a two-level independent variable (low, medium/high). Since the error rate had only five possible integer values in the range $[0, 4]$, we used an Ordinal Logistic Regression (OLR) model and a likelihood-ratio test [30] to analyze the fixed effects of VA, *rays*, and the interaction between them. The model (Nagelkerke $R^2 = 0.503$) [28] found no significant interaction between VA and *rays* ($\chi^2(1) = 1.781, p > 0.1$). So we took out the interaction term and fit a new OLR model (Nagelkerke $R^2 = 0.501$), and found a significant fixed effect for both *rays* ($\chi^2(1) = 252.17, \beta = -2.446, p < 0.0001$) and VA ($\chi^2(1) = 139.55, \beta = 2.06, p < 0.0001$). This result indicates that Double Ray was significantly more accurate than Single Ray overall, not only at the medium/high VA levels. It also suggests that the higher levels of VA were significantly more difficult than the low VA level.

To analyze H1 for task completion time, we started with a linear model (Adjusted $R^2 = 0.368, F(3, 536) = 105.9, p < 0.0001$) that included *rays*, two-level VA, and their interaction. To meet the linearity assumption, we took the log-transformed time as the response variable. The interaction was found significant ($F(1, 536) = 3.88, p = 0.049$). We used a post-hoc estimated marginal means (least-squares means) pairwise comparison to analyze their interaction, and found that all pairs were significantly different ($p < 0.005$). The interaction is demonstrated in Figure 8.

The analysis for H2 was similar to the analysis for H1, except that we kept the three VA levels separate to understand the differential effects of VA on errors the Single Ray and Double Ray techniques. An OLR model (Nagelkerke $R^2 = 0.548$) found a significant interaction between three-level VA and *rays* ($\chi^2(2) = 42.855, p < 0.0001$).

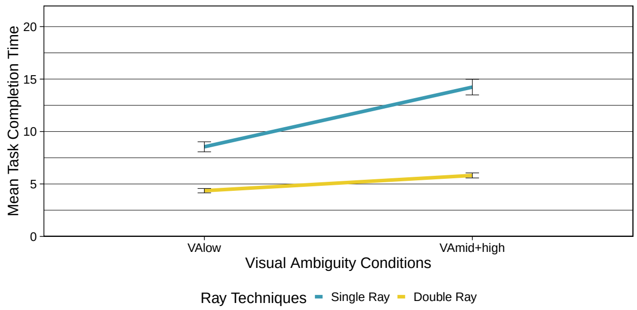


Figure 8: Interaction plot for Single and Double Ray techniques at two levels of VA. Error bars represent standard error. All pairs are significantly different.

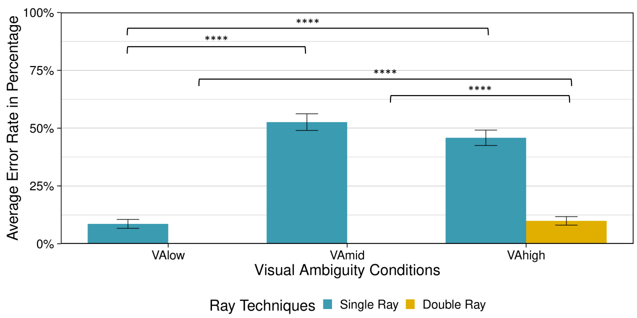


Figure 9: Average error rate for Single and Double Ray under different VA combinations. Error bars represent standard error. Brackets indicate statistical significance with p < 0.0001.

Since the response variable was categorical, instead of using post-hoc estimated marginal means, we performed three OLRs on [low, medium], [low, high] and [medium, high] VA subsets. For the Single Ray, we found significant differences between low and medium VA ($\chi^2(1) = 90.819, p < 0.0001$) and between low and high VA ($\chi^2(1) = 80.69, p < 0.0001$), but not between medium and high VA ($\chi^2(1) = 1.819, p = 0.177$). For the Double Ray, we found significant differences between low and high VA ($\chi^2(1) = 43.62, p < 0.0001$) and between medium and high VA ($\chi^2(1) = 78.61, p < 0.0001$), but not between low and medium VA ($\chi^2(1) = 1.35, p = 0.206$). Figure 9 shows the average error rates for the Single Ray and Double Ray techniques at the three different levels of VA conditions.

When testing H3, we rearranged the data based on the existence of VA. When using Single Ray techniques, any trial with medium or high VA was considered to have VA, whereas when using Double Ray techniques, only high VA trials were regarded as having VA. Following this definition, we divided the data points into two groups by a new category, namely WithVA. We first tested hypothesized improvement in accuracy by fitting an OLR for error rate to analyze the fixed effects of Parallel Bars, WithVA and SA. The model (Nagelkerke $R^2 = 0.381$) found no significant effect of Parallel Bars ($\chi^2(1) = 0.879, p = 0.348$) nor of interaction terms involving Parallel Bars. For task completion time analysis, we also fitted a linear model (Adjusted $R^2 = 0.228, F(7, 532) = 23.68, p < 0.0001$) involving Parallel Bars, WithVA, SA, and their interaction terms. No significant effect found in the interaction terms. A second model removing all interactions(Adjusted $R^2 = 0.2311, F(3, 536) = 55, p < 0.0001$) found significant effect for Parallel Bar ($F(1, 536) = 11.826, \beta = 0.876, p = 0.0006$) and WithVA ($F(1, 536) = 151.962, \beta = 1.015, p < 0.0001$). The result suggests a significant increment in task completion time when Parallel Bar was presented. Also having VA significantly slowed down participants.

To test H4, we analyzed data from the subjective interviews. We found that 23 out of 24 participants preferred the Double Ray over the Single Ray technique. They reported that Double Ray was more efficient in helping them rule out the decoys and locate the target, even in High VA trials. The Single Ray was reported to be usable in low VA cases but really hard to use in medium/high VA trials. In contrast, participants were mixed in their feedback about the Parallel Bars technique. Only a slight majority (13 out of 24 participants) felt the bars were useful in medium/high VA cases when the rays provide insufficient visual cues. Others felt the bars were less useful and a bit distracting in low VA cases specifically, when the rays are sufficient to identify the target visually. All but one of the participants said that the Double Ray was a more useful enhancement than the Parallel Bars.

Finally, we tested H5 by incorporating participants' perspective taking score as an additional co-variate in the previous models. We found a significant positive correlation between perspective taking scores and error rates ($p < 0.005$) in all cases. Since lower scores on the perspective taking test indicate better spatial orientation ability, this means that participants with higher spatial orientation ability had greater accuracy in the experimental tasks. Additionally, the interaction between Parallel Bars and perspective taking score approached the borderline of significance ($\chi^2(1) = 3.097, p = 0.078$).

5 DISCUSSION

We hypothesized that the Double Ray would effectively reduce visual ambiguity, resulting in higher accuracy and greater speed than the single ray (H1). Our results support H1 by showing that both error rate and task completion time for Double Ray were significantly lower than those of the Single Ray. In contrast to our hypothesis, however, we found that Double Ray was significantly faster and more accurate than Single Ray even when VA was low. We speculate that the the "bracketing" cue in the Double Ray technique was more visually salient and therefore easy to find than the "crossing center" cue of the Single Ray. The implication is that for the observer, the Double Ray is always desirable to improve target identification performance. However, it should be noted that we assumed perfect ray alignment and perfectly symmetrical objects in this experiment, so results in the real world are likely to be more complicated. In addition, we did not consider the extra effort required for the collaborator to specify the two rays. Even though it would not be possible to get better results than ours in real life, it would be reasonable to expect independent and random pointing errors from a real collaborator. Our findings would still be valid with real collaborator with worse average performance.

Our second hypothesis (H2) sought to confirm our intuition about the differential effects of VA on the Single and Double Rays. We reasoned that the Single Ray would have reduced accuracy in both the medium and high VA conditions, since in both cases the gaze ray would cross through the center of all the objects, while the Double Ray would only suffer reduced accuracy in the high VA condition, when multiple objects were bracketed perfectly. Our analysis clearly supported this hypothesis, as illustrated in Figure 9. Although VA does affect the Double Ray technique, this effect only occurs when multiple objects are perfectly bracketed (which is less likely), and the drop in accuracy is much smaller than that for the Single Ray technique. Again, it remains to be seen how easily users can specify and interpret Double Rays in more complicated real-world AR environments, but these results are promising.

Unlike the Double Ray, our proposed Parallel Bars enhancement was not very effective at improving accuracy in visually ambiguous conditions. We hypothesized (H3) that users would be able to use the information in the Parallel Bars visualization to understand the ray orientation, and therefore to improve their chances of selecting the correct target, at least in conditions where the target was the only object near the gaze ray. However, this claim about accuracy was not supported by the data analysis. On the other hand, Parallel Bars did increase task completion time, partially supporting H3. We speculate that users took more time to successfully choose the target because it takes mental effort to process the spatial information provided by the technique. It might also be the case that Parallel Bars contributed to visual clutter and thereby caused the participants to take more time to gather enough information for them to make a selection. In either case, Parallel Bars were not found to be helpful in our experiment, and simply using the basic Single or Double Ray techniques was a better choice, even when visual ambiguity was present.

Our expectation about the impact of SA was not supported by the results; the level of SA did not seem to affect either errors or time. We speculate that this was due to the limited number of objects in the experiment scene. With an increased number of potential targets, we might observe a more clear decrease in accuracy and/or increase in task completion time when SA is present.

The relative benefits of the Double Ray and Parallel Bars enhancements were also clear from the participant interviews. As we hypothesized in H4, participants very strongly agreed that the Double Ray was more usable and useful than the Parallel Bars. It is possible that a different visualization providing spatial information about the gaze ray might achieve better objective and/or subjective results, but given our iterative design process and comparison of multiple approaches (see the Section titled "Reducing Spatial Ambiguity"), we are inclined to believe that a good visual enhancement that can be used based on visual perception alone will always be superior to a spatial enhancement that requires cognitive processing.

In our final hypothesis (H5), we proposed that spatial orientation was an important individual characteristic that would affect users' ability to make correct selections using Parallel Bars techniques. H5 was partially supported by our correlation analysis. People with better perspective taking ability seemed to achieve higher accuracy overall, thus suggesting that the spatial orientation ability measured by the perspective taking test is related to the abilities involved in understanding gaze ray orientation of a distant collaborator. However, since the interaction between the Parallel Bars and perspective taking score was not quite significant, we speculate that the users tried to interpret spatial information from the Parallel Bars as we intended, but the technique was not as effective as we expected.

6 CONCLUSIONS AND FUTURE WORK

In wide-area collaborative tasks, AR systems can be used to provide information about the awareness of collaborators' gaze in support of shared spatial references. In this work, we designed and evaluated variants of a gaze ray visualization in a model-free AR setting. The primary obstacle to be overcome in this setting is the lack of correct occlusion cues to show the intersection of the gaze ray with objects in the real world, leading to visual ambiguity. We designed the Double Ray technique to reduce or eliminate this ambiguity, and showed experimentally that this technique is highly effective in disambiguating the target across a variety of task conditions.

When visual ambiguity cannot be eliminated, viewers must try to understand the spatial configuration of the collaborator, the ray, and the possible targets. Using an iterative design process, we designed the Parallel Bars visualization to provide ray orientation information to viewers in such cases. However, our experiment revealed that the only significant effect of using Parallel Bars was to increase the time needed to make a correct selection.

The primary contributions of our research include the analysis of the primary factors affecting performance in understanding gaze rays and their targets in a model-free AR setting, the design of multiple visualization enhancements to address these factors, and the results and implications of our controlled empirical study. Our work demonstrates that, assuming reliable tracking and an accurate collaborator, the Double Ray visualization has a significant positive effect on the effectiveness of the collaboration. The principal design recommendation from our study is to use the Double Ray technique when fast and accurate target identification by the viewer is required.

There are many possible directions of future research. Most importantly, we need to study the use of gaze ray visualization techniques in real-world wide-area environments with real AR systems. Our current experiment does not capture all the issues that would occur in a real collaborative outdoor AR system. We expect that our current findings will be confirmed at a high level, but real-world limitations (such as display field of view, tracking accuracy, and visual clutter) may have unexpected impacts on performance and usability. Similarly, we plan to study the use of these techniques by multiple human collaborators. How is performance affected by the user's ability to specify accurate rays? Although Double Ray has very positive benefits for observers, to what extent are these benefits offset by the extra effort and complexity incurred by the collaborator whose gaze is being visualized? Finally, we plan to investigate whether techniques like Parallel Bars can become more effective through training, and to explore additional methods of providing spatial information visually for use in cases where visual ambiguity cannot be eliminated completely.

ACKNOWLEDGMENTS

The authors would like to thank the anonymous reviewers for their valuable comments and helpful suggestions. This work has been supported by the Office of Naval Research.

REFERENCES

[1] Deepak Akkil, Biju Thankachan, and Poika Isokoski. 2018. I See What You See: Gaze Awareness in Mobile Video Collaboration. In *Proceedings of the 2018 ACM*

Symposium on Eye Tracking Research & Applications (ETRA '18). ACM, New York, NY, USA, Article 32, 9 pages. https://doi.org/10.1145/3204493.3204542

[2] Y. Baillot, D. Brown, and S. Julier. 2001. Authoring of physical models using mobile computers. In *Proceedings Fifth International Symposium on Wearable Computers*. 39–46. https://doi.org/10.1109/ISWC.2001.962094

[3] Dare A. Baldwin. 1995. . Lawrence Erlbaum Associates, Inc, Hillsdale, NJ, US, Chapter Understanding the link between joint attention and language., 131–158.

[4] Simon Baron-Cohen. 1995. . Lawrence Erlbaum Associates, Inc, Hillsdale, NJ, US, Chapter The eye direction detector (EDD) and the shared attention mechanism (SAM): Two cases for evolutionary psychology., 41–59.

[5] Doug A Bowman, Cheryl Stinson, Eric D. Ragan, Siroberto Scerbo, Tobias Hollerer, Cha Lee, Ryan P McMahan, and Regis Kopper. 2012. Evaluating effectiveness in virtual environments with MR simulation. In *Interservice/Industry Training, Simulation, and Education Conference (I/ITSEC)*. National Training and Simulation Association, National Training and Simulation Association, Orlando, FL. http://ntsa.metapress.com/link.asp?id=9h61028206h3503w

[6] Herbert H. Clark and Deanna Wilkes-Gibbs. 1986. Referring as a collaborative process. *Cognition* 22, 1 (1986), 1 – 39. https://doi.org/10.1016/0010-0277(86)90010-7

[7] A. I. Comport, E. Marchand, M. Pressigout, and F. Chaumette. 2006. Real-time markerless tracking for augmented reality: the virtual visual servoing framework. *IEEE Transactions on Visualization and Computer Graphics* 12, 4 (July 2006), 615–628. https://doi.org/10.1109/TVCG.2006.78

[8] James E. Cutting and Peter M. Vishton. 1995. Chapter 3 - Perceiving Layout and Knowing Distances: The Integration, Relative Potency, and Contextual Use of Different Information about Depth*. In *Perception of Space and Motion*, William Epstein and Sheena Rogers (Eds.). Academic Press, San Diego, 69 – 117. https://doi.org/10.1016/B978-012240530-3/50005-5

[9] C. Diaz, M. Walker, D. A. Szafir, and D. Szafir. 2017. Designing for Depth Perceptions in Augmented Reality. In *2017 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*. 111–122. https://doi.org/10.1109/ISMAR.2017.28

[10] Jeff Dyck and Carl Gutwin. 2002. Groupspace: A 3D Workspace Supporting User Awareness. In *CHI '02 Extended Abstracts on Human Factors in Computing Systems (CHI EA '02)*. ACM, New York, NY, USA, 502–503. https://doi.org/10.1145/506443.506450

[11] Steven Feiner, Blair Macintyre, and Dorée Seligmann. 1993. Knowledge-based Augmented Reality. *Commun. ACM* 36, 7 (July 1993), 53–62. https://doi.org/10.1145/159544.159587

[12] Susan R. Fussell, Leslie D. Setlock, and Robert E. Kraut. 2003. Effects of Head-mounted and Scene-oriented Video Systems on Remote Collaboration on Physical Tasks. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '03)*. ACM, New York, NY, USA, 513–520. https://doi.org/10.1145/642611.642701

[13] J. L. Gabbard, J. E. Swan, and D. Hix. 2006. The Effects of Text Drawing Styles, Background Textures, and Natural Lighting on Text Legibility in Outdoor Augmented Reality. *Presence* 15, 1 (Feb 2006), 16–32. https://doi.org/10.1162/pres.2006.15.1.16

[14] Carl Gutwin and Saul Greenberg. 1998. Design for Individuals, Design for Groups: Tradeoffs Between Power and Workspace Awareness. In *Proceedings of the 1998 ACM Conference on Computer Supported Cooperative Work (CSCW '98)*. ACM, New York, NY, USA, 207–216. https://doi.org/10.1145/289444.289495

[15] Mary Hegarty and David Waller. 2004. A dissociation between mental rotation and perspective-taking spatial abilities. *Intelligence* 32, 2 (2004), 175 – 191. https://doi.org/10.1016/j.intell.2003.12.001

[16] Jon Hindmarsh, Mike Fraser, Christian Heath, Steve Benford, and Chris Greenhalgh. 1998. Fragmented Interaction: Establishing Mutual Orientation in Virtual Environments. *CSCW* 1998, 217–226. https://doi.org/10.1145/289444.289496

[17] Jon Hindmarsh, Mike Fraser, Christian Heath, Steve Benford, and Chris Greenhalgh. 2000. Object-focused Interaction in Collaborative Virtual Environments. *ACM Trans. Comput.-Hum. Interact.* 7, 4 (Dec. 2000), 477–509. https://doi.org/10.1145/365058.365088

[18] T. N. Hoang and B. H. Thomas. 2010. Augmented Viewport: An action at a distance technique for outdoor AR using distant and zoom lens cameras. In *International Symposium on Wearable Computers (ISWC) 2010*. 1–4. https://doi.org/10.1109/ISWC.2010.5665865

[19] S. Kasahara, S. Nagai, and J. Rekimoto. 2017. JackIn Head: Immersive Visual Telepresence System with Omnidirectional Wearable Camera. *IEEE Transactions on Visualization and Computer Graphics* 23, 3 (March 2017), 1222–1234. https://doi.org/10.1109/TVCG.2016.2642947

[20] Lars Kjelldahl and Martin Prime. 1995. A study on how depth perception is affected by different presentation methods of 3D objects on a 2D display. *Computers & Graphics* 19, 2 (1995), 199 – 202. https://doi.org/10.1016/0097-8493(94)00143-M

[21] E. Kruijff, J. E. Swan, and S. Feiner. 2010. Perceptual issues in augmented reality revisited. In *2010 IEEE International Symposium on Mixed and Augmented Reality*. 3–12. https://doi.org/10.1109/ISMAR.2010.5643530

[22] W. S. Lages, Y. Li, L. Lisle, F. Lu, T. Höllerer, and D. A. Bowman. 2019. Enhanced Geometric Techniques for Point Marking in Model-Free Augmented Reality. In *2019 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*.

[23] J. Chris Lauwers and Keith A. Lantz. 1990. Collaboration Awareness in Support of Collaboration Transparency: Requirements for the Next Generation of Shared Window Systems. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '90)*. ACM, New York, NY, USA, 303–311. https://doi.org/10.1145/97243.97301

[24] J.J. LaViola, E. Kruijff, D.A. Bowman, R.P. McMahan, and I. Poupyrev. 2017. *3D User Interfaces: Theory and Practice*. Addison-Wesley. https://books.google.com/books?id=ilUyjwEACAAJ

[25] C. Lee, S. Bonebrake, D. A. Bowman, and T. Höllerer. 2010. The role of latency in the validity of AR simulation. In *2010 IEEE Virtual Reality Conference (VR)*. 11–18. https://doi.org/10.1109/VR.2010.5444820

[26] Taehee Lee and Tobias Hollerer. 2008. Hybrid feature tracking and user interaction for markerless augmented reality. In *2008 IEEE Virtual Reality Conference*. IEEE, 145–152.

[27] Paul K. Luff, Naomi Yamashita, Hideaki Kuzuoka, and Christian Heath. 2015. Flexible Ecologies And Incongruent Locations. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems (CHI '15)*. ACM, New York, NY, USA, 877–886. https://doi.org/10.1145/2702123.2702286

[28] N. J. D. NAGELKERKE. 1991. A note on a general definition of the coefficient of determination. *Biometrika* 78, 3 (09 1991), 691–692. https://doi.org/10.1093/biomet/78.3.691 arXiv:http://oup.prod.sis.lan/biomet/article-pdf/78/3/691/712023/78-3-691.pdf

[29] Mark B. Neider, Xin Chen, Christopher A. Dickinson, Susan E. Brennan, and Gregory J. Zelinsky. 2010. Coordinating spatial referencing using shared gaze. *Psychonomic Bulletin & Review* 17, 5 (01 Oct 2010), 718–724. https://doi.org/10.3758/PBR.17.5.718

[30] Jerzy Neyman, Egon Sharpe Pearson, and Karl Pearson. 1933. IX. On the problem of the most efficient tests of statistical hypotheses. *Philosophical Transactions of the Royal Society of London. Series A, Containing Papers of a Mathematical or Physical Character* 231, 694-706 (1933), 289–337. https://doi.org/10.1098/rsta.1933.0009 arXiv:https://royalsocietypublishing.org/doi/pdf/10.1098/rsta.1933.0009

[31] Wendy A. Schafer and Doug A. Bowman. 2005. Integrating 2D and 3D Views for Spatial Collaboration. In *Proceedings of the 2005 International ACM SIGGROUP Conference on Supporting Group Work (GROUP '05)*. ACM, New York, NY, USA, 41–50. https://doi.org/10.1145/1099203.1099210

[32] Diane H. Sonnenwald, Kelly L. Maglaughlin, and Mary C. Whitton. 2004. Designing to support situational awareness across distances : An example from a scientific collaboratory. *Information Processing & Management* 40, 6 (2004), 989–.

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

[34] Anthony Tang and Omid Fakourfar. 2017. Watching 360° Videos Together. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems (CHI '17)*. ACM, New York, NY, USA, 4501–4506. https://doi.org/10.1145/3025453.3025519

[35] Boris M. Velichkovsky. 1995. Communicating attention: Gaze position transfer in cooperative problem solving. *Pragmatics & Cognition* 3, 2 (1995), 199–223. https://doi.org/10.1075/pc.3.2.02vel

[36] J. W. Woodworth and C. W. Borst. 2017. Visual cues to aid 3D pointing in a virtual mirror. In *2017 IEEE Symposium on 3D User Interfaces (3DUI)*. 251–252. https://doi.org/10.1109/3DUI.2017.7893366

[37] Huahai Yang and Gary M. Olson. 2002. Exploring Collaborative Navigation:: The Effect of Perspectives on Group Performance. In *Proceedings of the 4th International Conference on Collaborative Virtual Environments (CVE '02)*. ACM, New York, NY, USA, 135–142. https://doi.org/10.1145/571878.571899

[38] Yanxia Zhang, Ken Pfeuffer, Ming Ki Chong, Jason Alexander, Andreas Bulling, and Hans Gellersen. 2017. Look together: using gaze for assisting co-located collaborative search. *Personal and Ubiquitous Computing* 21, 1 (01 Feb 2017), 173–186. https://doi.org/10.1007/s00779-016-0969-x