Evaluation of Pointing Ray Techniques for Distant Object Referencing in Model-Free Outdoor Collaborative Augmented Reality

Yuan Li, Ibrahim A. Tahmid, Feiyu Lu, and Doug A. Bowman

Fig. 1. Pointing Rays in AR 4. (a). Observer's view of the single ray marked by pointer aiming at Target 2, Parallel Bars are parallel to the single ray; the white sphere denotes the closest point on the ray to the observer (highlighted by blue square); (b). Observer's view of the Double Ray technique placed by the pointer aiming at Target 4; the two rays are visually bracketing the target lamppost. Images are captured from Microsoft HoloLens 2 and carry intrinsic offsets.

Abstract— Referencing objects of interest is a common requirement in many collaborative tasks. Nonetheless, accurate object referencing at a distance can be challenging due to the reduced visibility of the objects or the collaborator and limited communication medium. Augmented Reality (AR) may help address the issues by providing virtual pointing rays to the target of common interest. However, such pointing ray techniques can face critical limitations in large outdoor spaces, especially when the environment model is unavailable. In this work, we evaluated two pointing ray techniques for distant object referencing in model-free AR from the literature: the Double Ray technique enhancing visual matching between rays and targets, and the Parallel Bars technique providing artificial orientation cues. Our experiment in outdoor AR involving participants as pointers and observers partially replicated results from a previous study that only evaluated observers in simulated AR. We found that while the effectiveness of the Double Ray technique is reduced with the additional workload for the pointer and human pointing errors, it is still beneficial for distant object referencing.

Index Terms—Augmented Reality; Collaboration; Ray Visualization; Model-Free; Outdoor

1 Introduction

Communicating the location of the object of mutual interest is essential in various collaborative tasks [3, 9]. For example, a remote expert guiding a front-line worker may frequently ask where different parts are to identify the issue. Similarly, construction workers often need to achieve consensus on different locations during onsite inspection. Such communication establishes a shared spatial frame and ensures joint attention, which is critical to the collaboration [9].

Pointing is common to communicate the spatial locations of targets of interest [14]. Typically, the pointer can either use their finger, arm, or tools like laser pointers to help indicate the targets. This type of approach is simple, fast, and accurate compared with a purely verbal description for spatial information communication. However, object referencing can be challenging when the distances between the collaborators and the target increase. At a greater distance, the collaborators have reduced visibility of each other and the target, leading to the limited efficacy of the pointing method. Augmented Reality (AR) technology can use augmentation visualization techniques to address the issue by overlaying 3D graphics in the real world. With an environment model, the AR system can simply display a virtual ray indicating the pointing direction and pointed target location to indicate the referenced target. Such an approach has been proven effective and efficient in various Virtual Reality (VR) applications [1, 2, 33].

Without an environment model, the AR system has no information about where the target is and cannot show proper depth cues like occlusion to the viewer, causing ambiguity and confusion [10]. However, the environment model vital to this type of visual enhancement is not

• *Yuan Li, Ibrahim A. Tahmid, Feiyu Lu, and Doug A. Bowman are with the Center for Human-Computer Interaction and the Department of Computer Science, Virginia Tech. E-mail: {yli92 | iatahmid | feiyulu | dbowman}@vt.edu .*

Manuscript received xx xxx. 201x; accepted xx xxx. 201x. Date of Publication xx xxx. 201x; date of current version xx xxx. 201x. For information on obtaining reprints of this article, please send e-mail to: reprints@ieee.org. Digital Object Identifier: xx.xxxx/TVCG.201x.xxxxxxx

always available or reliable, especially in large outdoor spaces. Obtaining an accurate environment model can be challenging for a number of reasons. First, it can be technically difficult to acquire environmental information. For instance, depth sensors normally have a limited sensing range and function poorly under outdoor lighting conditions. Moreover, the world is dynamic and changing. This means that the environment model may require a constant update to ensure accuracy or reliability. Under the context of using an environment model for object referencing, having an inaccurate model can be equivalent to having no model at all because the results from these two scenarios are the same: the observer faces difficulty identifying the right target easily. Hence, when looking into object referencing issues in wide-area outdoor AR, it is worth considering an extreme case where the AR system does not rely on any known geometric model in the physical environment (as in so-called *model-free* AR [10]).

In model-free AR, a virtual pointing ray will not stop or intersect with the target object. As a result, the AR display will render the virtual ray on top of all objects in the physical environment, and the ray will not appear to penetrate the target, causing visual ambiguity.

In our initial work [32], we explored this object referencing issue in a simulated AR setup by designing and evaluating visualization techniques to help disambiguate the ray's target. Our work presented: 1) the Double Ray technique, which pointed two rays at two of the target object's geometric features, enhancing the visual match condition to reduce visual ambiguity; and 2) the Parallel Bars technique, which provided extra ray orientation information to enhance orientation perception. We found the Double Ray technique improved observer performance while the Parallel Bars technique failed to help the observer identify the referenced target. However, it is unclear how well the results would translate to real AR systems. First, while AR simulation is a common approach to conducting controlled and repeatable user experiments for AR studies, this approach may fail to replicate some AR cases, especially in an outdoor environment. With the advancement in commodity technologies, there is increasing demand for more ecologically valid, in-the-wild studies [19, 36]. Moreover, we only considered the role of the observer and assumed a perfect pointer who always cast the rays precisely at the target object. As a result, the techniques' potential cost on the pointer and the effects of pointing error on the techniques' performance were never investigated.

In this paper, we implemented and modified the Double Ray technique and the Parallel Bars technique based on the findings from our previous study [32]. We also designed and implemented an interface for the pointer to specify the pointing rays. To evaluate the pointing ray techniques in a more ecologically valid setting, we designed an experiment in a large model-free outdoor environment and synchronized the collaborators at a distance of 70 meters. In our study, 32 participants took both the pointer and observer roles in a collaborative spatial referencing task. The results partly validated our previous AR simulation study, but also revealed new findings due to the more ecologically valid setting.

The primary contributions of the research are:

- The replication of a previous simulated AR study in a real wide-area outdoor AR system
- Findings that partially confirm the results from prior study of the techniques [32]
- New insight into the trade-off between the pointing technique usability and observer performance
- A synchronization method that aligns wide-area AR collaborators in a shared frame of reference
- Detailed understanding of both sides of a collaborative AR task using multiple user experience measures

2 Related Work

2.1 Spatial Referencing

Dix [14] proposed the term "Computer-Supported Cooperative Work" to describe a framework that emphasized the importance of deixis in groupware. Spatial referencing is an example of deixis that primarily focuses on locations or objects in space [38]. Whittaker [50] emphasized the importance of spatial referencing in collaboration and commented that "what is shown and how it is shown is crucial when communicating about work objects in a visual environment."

Researchers are optimistic that AR will be a promising technology for next-generation groupware [15]. There has been no lack of work studying AR-assisted collaboration. Specifically, researchers have explored various ways to enhance spatial referencing in collaborative AR. Chastine and Zhu [6] discussed spatial referencing within a shared environment as a fundamental requirement in collaboration. They identified contextual and environmental factors influential to referential awareness from several collaborative AR studies. They concluded that accurate and efficient referencing critical in collaborative tasks around shared objects to avoid additional communication costs, and well support for spatial referencing would significantly improve collaboration.

Oda and Feiner [39] studied a gesture-based 3D referencing technique in shared AR. Using a depth camera, their system was capable of capturing the user's hand pointing gesture and reconstructing a near-range depth map of the target area. Their work demonstrated that a well-designed referencing visualization technique could be significantly more accurate than a tracked controller even when the collaborators had different perspectives in the shared virtual environment. More recently, Kim et al. [25] designed two gesture-based referencing techniques for near-range virtual target referencing from different perspectives. They found that their technique utilizing the environment model performed significantly better than the other technique relying on hand tracking.

In short, while there is extensive evidence of the importance of supporting spatial referencing, many referencing techniques rely on an environment model and only work at near distances.

2.2 Pointing Rays

Pointing rays are widely used in VR for different selection tasks at a distance [24]. Ray pointing relies on intersecting a ray with a surface to determine the location, object, or menu item the user intends to interact with. This approach efficiently provides a naturalistic way to interact with virtual content in virtual environments, in a wide variety of VR systems [4, 11]. The technique is based on pointing gestures that are pervasive in the user's daily life [48]. Compared with other conventional 2D interfaces such as mouse-based pointing, ray pointing does not rely on a physical surface [46]. Ray pointing also brings convenience to access distant targets without asking the user to travel to other locations, making it ideal for large-scale working environments [42].

Ray pointing techniques have also been broadly studied in collaborative contexts. Collaborators can try to understand the pointing ray from a non-egocentric perspective to identify the referenced content, as seen in large 2D display systems [40, 41]. In 3D object referencing, pointing rays generally suffer from accuracy due to increased distance and often require enhancements [13, 17, 27, 29]. The pointing ray generally appeals to the observer as well, as researchers reported it to be simple, naturalistic, and easy to understand [8, 17, 24, 32, 45].

The benefit of using pointing rays to communicate object referencing is well demonstrated in the community. However, the technique often requires an environment model to enhance its visual appearance to the observer, especially at large distances.

2.3 Depth Perception

Cutting and Vishton [12] divided human perceptual space into three areas: the personal space (under 2 meters), the action space (up to 30 meters), and the vista space (beyond 30 meters). According to this taxonomy, only four depth cues remain reasonably effective in the vista space. Among the four available depth cues, only occlusion provides detailed depth information without significant distance changes. A recent experiment [31] studied the effect of occlusion in barehanded object referencing tasks in AR and concluded that without proper occlusion, object referencing would be severely affected, leading to decreased collaboration performance.

Unfortunately, in model-free AR, virtual content cannot be occluded by objects in the physical environment and will thus appear on top of everything. One approach to enhance depth perception is to use wireframe visualization [18]. Virtual objects are rendered in wireframe form so that the user can see through the augmentations to perceive the real world behind them. The wireframe structure also enhances the

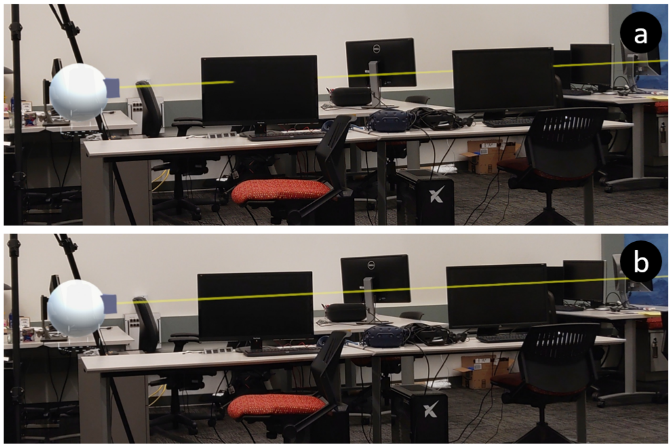

Fig. 2. First person view of different pointing rays (a) with or (b) without occlusion.

depth perception of the virtual content through relative size and density. Yet, the method fails to bridge the virtual and the real worlds and does not help the user correctly understand the relative depth between them. The visualization also raises the Necker Cube Illusion [26]. Our previous work [32] explored using an artificial orientation cue to help the user understand the direction of a pointing ray in the vista space. Our findings indicated limited effectiveness of the orientation cue on target identification but revealed potential as the participants made an effort to interpret the visualizations.

2.4 Model-Free Outdoor Collaborative AR

While AR is often thought to be the next-generation technology that can improve our daily life [15], applying the technology in an outdoor environment is still technically challenging. One of the fundamental challenges comes from tracking [7,22,44]. Tracking determines the registration of virtual content with the physical world and affects the synchronization between multiple users [47]. With continuing advancement in technology, mobile AR devices are capable of 6 degree-of-freedom (DoF) tracking in outdoor environments [20,35,43]. However, a reliable synchronization that localizes multiple outdoor users in a common frame of reference is still an open challenge.

Huo et al. [22] presented SynchronizAR, an approach to spatially register multiple SLAM devices together without sharing maps or involving external tracking infrastructures. Their approach was tested indoors at a distance of $r \approx 4$ meters with an average translational accuracy of 0.15 meters and rotational accuracy of 7.4°. McGill et al. [35] presented a two-point marker-based alignment synchronization method that observed 0.1 meters of positional error in an area of $r \approx 10$ meters. Our study adopted McGill's approach and achieved positional error smaller than 0.5 meters when the pointer and observer were separated by 70 meters.

3 Methods and Implementation

3.1 Techniques for the Observer

3.1.1 Single Pointing Ray

While a pointing ray can be a powerful referencing tool in VR, it can be severely limited in model-free AR. With an accurate and reliable environment model, the ray can interact with the real world and be occluded by physical objects in the scene, as shown in Figure 2 (a). However, in model-free AR, since the geometric information of the physical surroundings is unknown to the system, the AR display has no way of knowing where physical objects are and hence cannot render the occlusion effect when the virtual ray goes behind the objects, leading to a confusing overlaying effect as shown in Figure 2 (b). We refer to this technique as the *Single Ray* technique. With the Single Ray, the observer must execute a visual search task that looks for the targets that the ray visually crosses. Since the ray overlays the physical environment, multiple objects often intersect with a single ray.

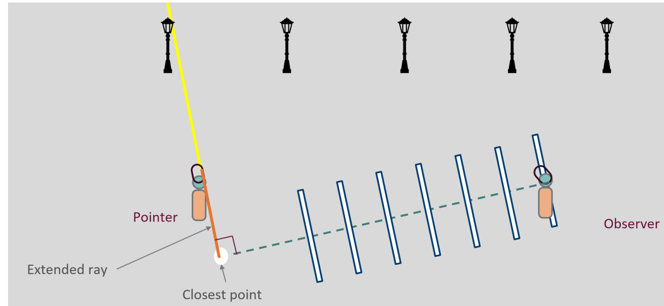

Fig. 3. A top-down view of the Parallel Bars technique. The Parallel Bars are parallel to the yellow ray and are placed along a line that is perpendicular to the yellow ray. Since the closest point (white sphere) is behind the pointer, we extend the pointing ray to highlight its direction.

3.1.2 Double Ray Technique

We designed the Double Ray technique to address the occlusion issue in model-free AR. With the Double Ray technique, instead of a single ray aiming at the center or top of the target object, the pointer needs to specify two pointing rays at two different geometric features (typically the top and the bottom) of the target object. The theoretical benefit behind is that by increasing the number of rays, the visual match condition changes from intersecting to a "bracketing" effect, reducing possible ambiguous cases, as shown in Figure 1 (b).

The obvious downside of the Double Ray technique is the added workload on the pointer. This particular negative impact was not evaluated in our prior work [32]. We discuss our pointer interface design in Section 3.2.

3.1.3 Parallel Bars Technique

A different approach aiming to solve the incorrect occlusion problem in the Single Ray technique is through enhanced orientation perception. Since the observer has typically no difficulty in understanding depth information of real-world objects, if the observer can accurately perceive the direction of the pointing ray, they may be able to identify the target even when there is visual ambiguity due to false occlusion cues. The observer is more likely to identify the target correctly if the target is separated from other objects intersecting with the virtual ray. However, the pointing ray generally has a small thickness, and its perspective cannot provide detailed depth information. Hence, an alternative approach is to provide extra spatial cues to help the observer understand the ray's direction. Even though the system does not possess environment geometric information in model-free AR, the pointing ray's origin and direction are available for the display to render in 3D space. Combined with the observer's position, multiple artificial orientation cues can be provided to the viewer. We referred to this technique as the *Single Ray with Parallel Bars* technique.

To provide a straightforward way of knowing the ray's direction, the Parallel Bars technique creates virtual segments (also called Parallel Bars) parallel to the original pointing ray. The first bar is placed at the observer's shoulder height and the rest of the bars are at a fixed interval along the direction that is perpendicular to the pointing ray, as shown in Figure 3. In addition to the original implementation, we allow the user to configure the height, length, and interval of the Parallel Bars so that the observer can adjust the parameters based on their preference. To highlight the perpendicularity relation to the observer, we further add a virtual white sphere on the ray to denote the intersection between the ray itself and the Parallel Bars placement direction, as shown in Figure 1 (a) and Figure 3. Given the geometric information, the white sphere also represents the point on the ray that is closest to the observer. Depending on the direction of the pointing ray, the closest point can be behind the pointer, as shown in Figure 3. In this case, we extend the original ray with a different color to the closest point.

While limited benefit was observed in the previous AR simulation experiment [32], we included the Parallel Bars technique for several reasons. First, we included it to provide a complete replication of the previous study. Additionally, even though the technique was not helpful to user performance in the VR study, we found that over half of the participants still preferred Parallel Bars when the ray was confusing. Moreover, the enhancements we added to the Parallel Bars technique aim to provide the users with a better understanding of the geometric relationships of the collaborators and the target. Last but not least, we anticipated that Parallel Bars might be more useful in the presence of imperfect pointing in our collaborative study.

Since the Double Ray technique and the Parallel Bars technique are designed to address the issue of distant object referencing in model-free AR via two independent approaches, we are able to combine these two techniques and define the *Double Ray with Parallel Bars* technique. The Parallel Bars are specified in the same way as the Single Ray with Parallel Bars technique, except that the bars are now parallel to the centerline of the two rays.

3.2 Techniques for the Pointer

The goal of the pointer interface design is to help the pointer to specify the target accurately and easily. Starting with the Single Ray case, the pointer only needs to define an origin and a direction to specify a pointing ray in 3D space. Since many AR head-worn displays (HWDs) are tracked in 6 DoF, a simple approach is to use the user's head position and facing direction. In other words, the pointer only needs to look in the direction of the target and use an input (such as a controller) to specify a ray from their head to the object of interest. However, as reported by Lages et al. [28], naïve use of the user's head pose can be problematic due to involuntary head tremor. To help the pointer specify precise rays, we adopt the multi-sampling idea from prior work of progressive refinement [27, 28] to achieve higher precision. We ask the pointer to specify multiple ray samples ($n = 100$), then we calculate the average to attenuate the pointing error. In practice, we display a crosshair at the center and a counter at the bottom left corner of the pointer's view. The pointer moves their head to position the target at the center of the crosshair, then, keeping their head as steady as possible, the pointer presses and holds a button on the controller to record 100 samples (which takes about three seconds in our implementation) before releasing the button. By pressing a separate confirm button, the user instructs the system to calculate the average of all the samples and produce the final ray, as illustrated in Figure 4 (a) and (b).

When using the Double Ray techniques, the pointer needs to specify two rays at the top and the bottom of the target object. Instead of asking the pointer to perform the same aiming-sampling tasks twice consecutively, which needless to say doubles the pointing time, we decided to use two crosshairs so that the pointer could aim at two geometric features simultaneously. As shown in Figure 4 (c), the pointer uses two joysticks to move the two crosshairs to align with the top and the bottom of the target object and start the sampling process.

3.3 Model-Free Outdoor Synchronization

To evaluate the pointing ray techniques in a collaborative ecologically-valid setting, we need to implement the system in a large-scale outdoor environment. A primary challenge is a reliable synchronization, such that the pointer and the observer can both be localized accurately in a common frame of reference. As the 6-DoF-tracked AR HWD maintains a local 3D coordinate system to register virtual content in space, the key is to bridge the two local coordinate systems of the two users. Inspired by works like Huo et al. [22], we initially decided to use an image-marker-based alignment approach. We first placed an image marker near each of the two collaborators, using it to place an anchor point in the users' local coordinate systems using Vuforia [1]. Then we used a laser to align the orientation of the image targets so that the offset between the image targets was limited to one axis. With the help of a rangefinder, we could thus theoretically compute the offset between the anchor points. We used Photon Engine [2] to establish a communication between the AR headsets and a server computer, which they could use to send out and receive position and rotation information relative to the anchor points at 60 frames per second. Using the relative position and rotation, the program would then theoretically be able to synchronize the pointer and the observer.

However, in practice, the alignment error was substantial with this approach. At our experiment site, the distance between the collaborators was 70 meters. At this distance, an angular error of 0.1° would result in a shift of approximately seven meters in the position of the other user. The source of such angular errors could come from image target recognition and laser alignment.

The work of McGill et al. [35] directed us to increase the number of image targets to reduce angular drift from image recognition. The insight behind this was that position tracking was much more accurate than orientation tracking. By using two image targets near the AR HWD, the device could detect two anchor points, defining a direction in space. The defined direction, the fixed world up direction, and one of the anchor positions could then define a 3D coordinate system.

Moreover, instead of using a laser to align image targets, McGill et al. [35] also introduced a synchronization strategy that relied on the AR HWD's self-tracking. This worked out much better in our testing than using two sets of image targets next to each collaborator. We eventually determined to place two image targets (with a fixed spatial relationship) midway between the two collaborators. Two experimenters would stand at the location of the image targets, where each AR device would recognize these image targets and define a coordinate system. Then, the experimenters would slowly walk to the locations of the two users, ensuring continuous and stable tracking along the way. The experimenters would also check the synchronization quality before handing the device to the pointer and observer. The result is illustrated in Figure 1.

4 EXPERIMENT

We conducted a user study with 32 participants from our university to replicate and improve upon the study in simulated AR. [32]. Unlike the prior study that was conducted in simulated AR, our experiment used outdoor AR to evaluate the ray techniques. Lee et al. [30] have found that high-dynamic range lighting coupled with an optical see-through display can have a detrimental effect on the perception of physical objects in the real world. Hence, current VR displays are unlikely to fully replicate an outdoor AR experience. Moreover, the previous study used a virtual collaborator with zero pointing error and only evaluated the techniques from the observer's perspective. By comparison, our participants took turns as pointers and observers and used our techniques to perform a collaborative spatial referencing task in a more ecologically valid setting. We designed the experiment to investigate the following research questions (RQs):

- **RQ 1**. What is the effect of the Double Ray technique, as compared to the Single Ray, on performance and subjective experience for the observer in model-free collaborative AR at a distance?
- **RQ 2**. What is the effect of Parallel Bars, as compared to techniques without Parallel Bars, on performance and subjective experience for the observer in model-free collaborative AR at a distance?
- **RQ 3**. Is there a significant decrement in performance or subjective experience for the pointer when using the Double Ray techniques, as compared to Single Ray techniques?
- **RQ 4**. What trade-off does the pointer have to make to reduce visual ambiguity for the observer?

We proposed the following hypotheses from the above RQs:

- **H1**. *The Double Ray technique will have significantly better overall performance and subjective experience than the Single Ray for the observer.* This was the major finding from the previous study. We expect to see a similar performance improvement from the Double Ray technique in the outdoor AR experiment.
- **H2**. *Techniques using Parallel Bars will be more accurate, but slower, than the techniques without Parallel Bars.* The increased observing time was also reported in the previous study. With our additions to the original technique, we hope to help the observers interpret the orientation and lead to better accuracy.
- **H3**. *Combining Double Ray and Parallel Bars will result in the best accuracy and confidence, but worse speed, than Double Ray alone.* We expect the combination of the bracketing effect of the

[1] https://www.ptc.com/en/products/vuforia

[2] https://www.photonengine.com/

Fig. 4. The pointer's view using the pointing interface to specify: A single ray (a& b), and two rays (c). The number on the top left represents the target for the pointer. The number of samples taken by the pointer (ranging from 0 to 100) is shown on the bottom left.

Double Ray and the Parallel Bars will provide observers with the most information, increasing accuracy and confidence, at the cost of increased observing time.

- **H4**. *The Double Ray technique will have significantly lower performance and subjective experience for the pointer.* Due to the need for more input and a slightly more difficult aiming task, we expect the pointer to spend a longer time and more mental effort pointing than using the Single Ray technique.

4.1 Participants

We recruited 32 participants (11 females) between 18 and 29 years old ($M = 23.62, SD = 3.58$) from a local university. All had normal vision (corrected or uncorrected), with 17 of them using glasses, and three of them using contact lenses. Eleven participants had never tried AR prior to the study. Among the others, only six participants had used AR more than twice.

4.2 Apparatus

The experiment used two HoloLens (HL) 2 HWDs with 2K resolution per eye and a diagonal field of view of 54 degrees. Despite boasting a high resolution, AR imagery loses a significant amount of contrast in well-lit environments such as outdoors in direct sunlight [16]. To compensate for this, we fashioned a custom sunscreen for the HL2 that enabled better visibility for the participants while not compromising the awareness of the physical outdoor environment around the participant. Following a HL1 template [3], we fashioned a sunscreen that fit the HL2 visor. We chose to use a tinted window film [4] because of its no-adhesive, static-cling installation method. With the sunscreen attached to the HL2 visor, the brightness of the real world seen through the HL2 visor ranged between 1500-2000 lux, as measured by the iOS application Light Meter LM-3000 [34], making the virtual content visible even in sunlight. Comparatively, a typical sunny day has a brightness of around 120K-140K lux at the experiment site . We also used a laptop with 8GB RAM, 1.8GHz i5 CPU as a control server in the study. We ran software on the laptop to change the techniques, switch the roles (pointer vs. observer) for the two HL2 HWDs, and record performance data during the experiment. Software for both the HL2 and laptop were implemented using Unity game engine v2019.4.26.

We used a wireless Xbox controller for input by both the pointer and observer. Pointers used the 'B' button to sample multiple rays and the 'A' button to confirm selection. Pointers using the Double Ray technique used the left and the right joysticks to align the two crosshairs with the top and the bottom of the target (Figure 4). Observers used the left joystick to navigate through the circular menu and the 'A' button to confirm selections (Figure 5).

4.3 Task

The study took place in a large outdoor environment with two participants positioned at a distance of 70 meters, and seven lampposts were marked as targets. The closest and the furthest target were distanced at 84.64 meters, resulting in an average of 85° range of gazing direction for each participant To enhance visibility in AR at such a distance, each participant's head was represented by a semi-transparent blue sphere

[3]https://medium.com/ocean-industries-concept-lab/how-to-create-your-own-hololens-sun-screen-68c466071a01

[4]https://www.amazon.com/gp/product/B000HE57JS/

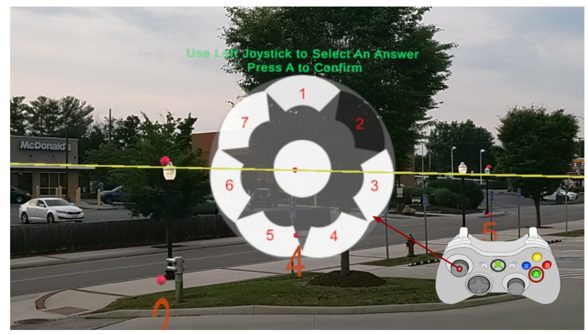

Fig. 5. Circular menu for the Observer. Left joystick is used to navigate through the options and the 'A' Button is used to confirm the selection.

and a yellow cone in front of the sphere was synchronized with user's facing direction (Figure 1 b).

In order to obtain the ground truth positions of the lampposts' top and bottom, we used the ImageRefinement marking technique described by Lages et al. [28] to mark these positions in a fixed coordinate system. We used the measured positions not only to measure pointing error, but also to place virtual markers (red spheres) at the top and bottom of the lampposts. In this way, we could ensure acceptable visibility of the targets in an outdoor environment using optical see-through headsets. Finally, having the pointers aiming at the virtual markers could minimize the impact of synchronization errors in the experiment. Since the pointers were aiming at virtual markers, in the observer's AR headset, the virtual rays were also pointing towards the virtual markers. Hence, even if the target in physical space was slightly misaligned, the alignment error would not reduce the observer's accuracy.

The participants took turns acting as pointer and observer for each technique. The participants had to complete 14 trials (2 for each target) for each technique for each role. Each trial consisted of the pointer pointing at a target, and the observer identifying the target. The sequence of the targets was randomized to reduce any learning effect. We also switched the starting role for participants at each participant location after every four pairs of participants to reduce any spatial bias.

Theoretically, the Double Ray and the Parallel Bars techniques should address distant object referencing issues independently. Therefore, we treated these two as independent variables with two levels each (namely *Number of Rays* (Single or Double) and *Orientation Cues* (absent or present)). Throughout the paper, we use these two terms to refer to the independent variables of our study to avoid confusion with our tested techniques. Since we took a within-subject design, each participant needed to complete $2(Number\ of\ Rays) \times 2(Orientation\ Cues) \times 2(roles) \times 7(targets) \times 2(repetitions) = 112$ trials.

4.4 Procedure

We conducted the experiment in four stages: pre-study, introduction, formal study, and post-study interview.

Pre-study Before the day of the study, we sent the participants an online pre-study questionnaire and collected demographic information and their prior experience with VR and AR. On the day of the study, we met the participants outside at the experiment site, and gave them an informed consent form (approved by the Institutional Review Board of the university) to read and sign.

Introduction After a brief introduction to our study, we helped the participants calibrate the HL2 headset for optimal viewing experience.

Formal Study After calibration, the experimenters escorted the participants to different designated locations , and asked them to stand on a pre-defined marker. We provided a chair for participants and encouraged them to take rest whenever needed.

For each technique, we started by demonstrating the technique on a picture of a top-down view of the study area. We proceeded to launch a training program where the participants were trained as both the pointer and the observer independently with a simulated collaborator situated at the same distance as the actual collaborator would be. The participants had to complete at least 14 trials both as pointer and as observer to successfully pass the training program. Success of failure for each trial was conveyed to the participant through auditory feedback.

For pointer training phase, the pointer could also view an additional purple ray (two rays for DoubleRay techniques) that went through exactly the center of the target while their own ray(s) was (were) displayed in yellow. The pointer could see the difference between to learn from their mistakes and update their strategy accordingly, especially in case of a failed trial (when their pointing error exceeded 0.5°). Both the auditory and the visual feedback were only available for training.

For the observer training, the observer learned to identify the target pointed at by the simulated pointer, and confirm their selection from a circular menu with the numbers one through seven in clockwise sequence. The circular design was chosen as all the numbers were equally accessible and reduced bias to any particular target.

During the observer training involving the Parallel Bars techniques, observers were able to manipulate the length, intra-bar distance, and height of the parallel bars (all measured in distance) using the controller. The observers were also instructed that the parameter adjustment was only available during the training and they should adjust these parameters to their preference.

When both of the participants were satisfied with their respective pointer and observer training, we launched the experimental trials , where the two participants were assigned pointer and observer. The pointer and observer collaborated with each other to complete a set of 14 trials using the technique they learned. Unlike the training phase, the collaborators' poses were synchronized in both headsets and followed the movements of the participants. As soon as the pointer placed the ray(s), their avatar was frozen. After finishing all trials, we gave the pointer a NASA Task Load Index (TLX) questionnaire (if it was the first time they used the pointing technique), and we gave the observer a modified System Usability Scale (SUS) questionnaire (see Section 4.5).

Next, the participants switched their roles and completed another set of 14 trials using the same technique. After completing these trials, we again collected the TLX and SUS questionnaires.

We repeated the training and collaborative trials for each technique.

Post-Study Interview Finally, after the participants used all four techniques, we had an open-ended interview with the participants. We asked how they liked the techniques and how they would rank them based on their preference. Key questions from this interview session included:

- As a pointer, what strategies did you use to try to make the rays accurate?
- As a pointer, please rank the two pointing techniques based on your preference.
- Why do you rank one technique over the other?
- As an observer, what strategies did you use to make a judgment?
- Please rank the four observing techniques based on your preference and explain your choice.

4.5 Measures

For each trial in the actual study, we calculated the task completion time for both the pointer and the observer. The *pointing completion time* is defined as the duration from the moment the target number appears on the screen to the moment the ray(s) is(are) drawn. The *observing completion time* is defined as the duration from the moment the ray(s) is(are) drawn to the moment the selection is confirmed.

The *pointing error* was defined as the angular error between the user-specified ray(s) and the ideal direction towards the target sphere(s). For Double Ray, the *pointing error* was defined as the mean of the errors for the two rays. For *observer accuracy*, we recorded whether or not the observers successfully identified the target. We did not give the observers any feedback on accuracy, to avoid learning effects.

Finally, during the experiment, we gathered subjective feedback through Raw NASA TLX [21] and a SUS questionnaire modified from the original questionnaire [5].

We used different questionnaires for the observers and pointers because we were curious about different aspects of subjective user experience for each role. For the pointers, we wanted to know what was the workload for using the pointing techniques (TLX), whereas, for the observers, we wanted to understand their perceptions of usability of the four techniques (SUS). In addition, we used an updated SUS questionnaire by dropping questions irrelevant to our study and adding relevant questions about understandability and mental effort. Following the original SUS questionnaire, the final set of questions are arranged by alternating positive and negative statements as presented below. The observers gave a rating for each question on a scale of 1-20.

- **[Learnability]** I would imagine that most people would learn to use this ray visualization technique very quickly.
- **[Prerequisite]** I needed to learn a lot of things before I could get going with this ray visualization technique.
- **[Confidence]** I felt very confident using the ray visualization technique to identify the target.
- **[Complexity]** I found the ray visualization technique unnecessarily complex.
- **[Understandability]** I thought the ray visualization technique was easy to understand.
- **[Mental Effort]** I felt that I spent a lot of mental effort to understand the ray visualization technique.

5 Results

The study initially ended with 1792 data points (DPs) $(32(participants) \times 2(repetitions) \times 2(Number\ of\ Rays) \times 2(Orientation\ Cue) \times 7(targets))$. Unfortunately, some data points contained incorrect data, such as timestamp and pointer errors, due to network issues. We excluded those erroneous data points from visualization and further analysis.

5.1 RQ 1. Effects of Double Ray Technique on Observer Performance and Subjective Experience

Figure 6 presents the observing time by pointing techniques. We started with a linear mixed model (LMM) using Welch-Satterthwaite t-Test [49] with participant pairs as random effects to test Number of Rays, Orientation Cue, and their interaction. The intercept is using Double Ray. We found the interaction significant ($t(1632.3) = -2.01$, $p = 0.044$, $\beta = -0.61$, $SE = 0.30$, $R^2 = 0.13$ [37], 1651 *DPs*). Thus, we grouped the data based on the level of Orientation Cue and investigated the effect of Number of Rays in two separate analyses. In the analysis comparing only the Double Ray technique and Single Ray technique without Parallel Bars, an LMM revealed a significant effect of the Double Ray technique on the observing time ($t(824.1) = -3.1$, $p = 0.002$, $\beta = -0.62$, $SE = 0.2$, $R^2 = 0.16$, 841 *DPs*). When Orientation Cue was present, another LMM found no significant difference between the two techniques ($t(794.1) = -0.06$, $p = 0.95$, $\beta = -0.62$, $SE = 0.2$, $R^2 = 0.091$, 810 *DPs*). The result indicates that the observers spent significantly longer using the Double Ray technique than the Single Ray technique, but only when the Orientation Cue was absent.

Similarly, Figure 7 plots the average accuracy that the observers achieved using different pointing techniques. While the average observer accuracy was calculated across different targets and observers, each trial was marked by a binary code as right or wrong. Hence, we used a binomial generalized linear mixed model (GLMM) with Laplace approximation method [23] to test Number of Rays, Orientation Cue, and their interaction. Since we found no significant effect from the interaction term, we instead used two GLMMs to analyze the effects of Number of Rays and Orientation Cue independently. We found a main

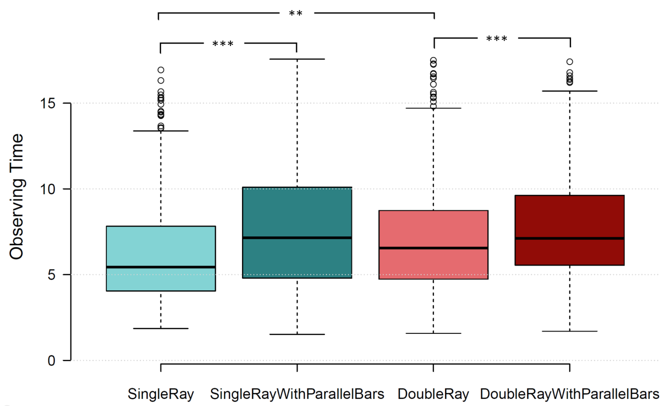

Fig. 6. The time that observers used to identify targets for different ray techniques. $*$ means $p \leq 0.05$, $**$ means $p \leq 0.01$, $***$ means $p \leq 0.001$

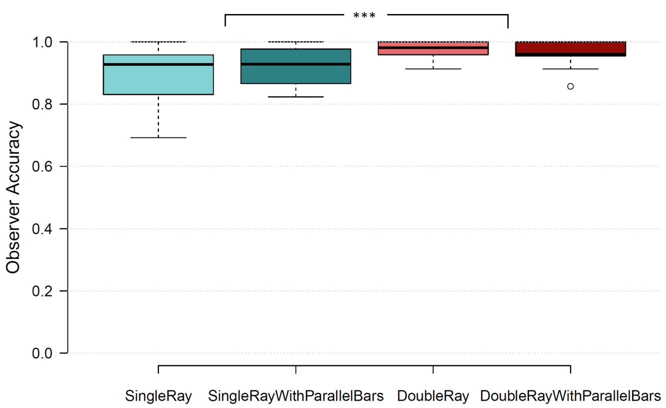

Fig. 7. Average accuracy the observers achieved using different pointing ray techniques. Accuracy was calculated across targets and observers.

effect of Number of Rays on observer accuracy ($z = -5.86$, $p < 0.0001$, $\beta = -1.2$, $SE = 0.2$, $R^2 = 0.15$, 1699 *DPs*). The result indicates that, regardless of the use of Parallel Bars, as long as the observers saw the Double Ray, their target identification was more accurate than the Single Ray technique.

Since we dropped the irrelevant questions from the original SUS questionnaire and added custom questions for our collaborative task, we could not use the same score thresholds to assess our techniques as the original SUS score. Hence, we could only compare the raw scores among the four tested techniques. A series of LMM analyses found that the Double Ray techniques had significantly higher learnability ($t(94) = -2.19$, $p = 0.031$, $\beta = -1.13$, $SE = 0.51$, $R^2 = 0.55$, 128 *DPs*), higher confidence ($t(31.0) = -4.05$, $p = 0.0003$, $\beta = -3.63$, $SE = 0.89$, $R^2 = 0.52$, 64 *DPs* with Orientation Cue absent, $t(31.0) = -2.26$, $p = 0.031$, $\beta = -1.4$, $SE = 0.62$, $R^2 = 0.66$, 64 *DPs* with Orientation Cue present), and less complexity when the Orientation Cue was present ($t(31.0) = -2.27$, $p = 0.03$, $\beta = 1.88$, $SE = 0.82$, $R^2 = 0.43$, 64 *DPs*) than the Single Ray techniques. We found no significant effect of the Number of Rays on the overall modified SUS score ($t(94.0) = -1.45$, $p = 0.149$, $\beta = -0.55$, $SE = 0.38$, $R^2 = 0.69$, 128 *DPs*). These results demonstrate that the Double Ray enhancement can benefit some aspects of observer user experience but cannot provide solid evidence to claim an overall improvement, as shown in Figure 8.

We ran a thematic analysis on the post-study user interviews. We coded the user statements and organized them as pros and cons for each technique. We found that, when the Orientation Cue is absent, 97% of the observers preferred the Double Ray over the Single Ray technique. For the techniques with Orientation Cue present, 91% of the observers preferred Double Ray over Single Ray. The observers mentioned feeling confident about using the Double Ray technique because they found it reliable and useful for a collaborative task. Twenty-eight of the 32 participants preferred the Double Ray technique over Single Ray

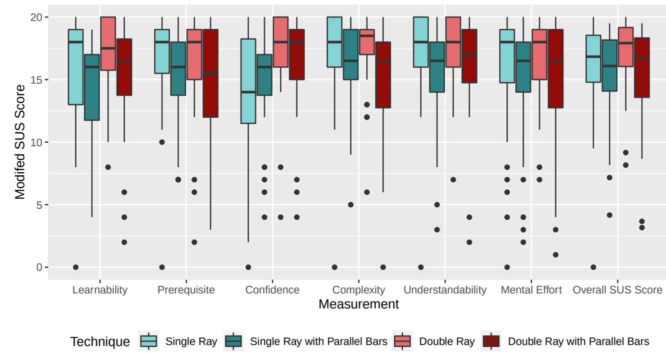

Fig. 8. Results from the modified SUS score by different ray techniques.

with Parallel Bars as observers, indicating that the enhancement of two rays was more useful for observers than the enhancement of parallel bars. These findings indicate that the preference for Double Ray relies on observers' prioritization of accuracy over speed, especially when they consider working with a collaborator.

5.2 RQ 2. Effects of Parallel Bars Technique on Observer Performance and Subjective Experience

When analyzing the effects of Orientation Cue on observer performance and subjective experience, we adopted the same methods as Section 5.1 with Parallel Bars as the intercept. As mentioned above, we detected a significant interaction between Number of Rays and Orientation Cue on the time observers spent identifying the target. After we grouped the data by Number of Rays, we used LMMs on the two subsets to investigate the effect of Orientation Cue on observing time. In the case of the Single Ray, there was a significant effect of Orientation Cue ($t(792.5) = -6.06$, $p < 0.0001$, $\beta = -1.27$, $SE = 0.21$, $R^2 = 0.26$, 809 *DPs*). With the Double Ray, there was also a significant effect of Orientation Cue ($t(825.6) = -3.27$, $p = 0.001$, $\beta = -0.69$, $SE = 0.21$, $R^2 = 0.072$, 842 *DPs*). These results suggest that the observers were more likely to spend a longer time identifying the target using the Parallel Bars techniques.

GLMM analysis found no significant effect of Orientation Cue on observer accuracy ($z = -1.16$, $p = 0.25$, $\beta = -0.21$, $SE = 0.18$, $R^2 = 0.055$, 1699 *DPs*), suggesting that the use of Parallel Bars did not help the observers make better target identification as we initially hoped.

In terms of modified SUS score, techniques with the Orientation Cue present were found to have a significantly lower overall score ($t(94.0) = 3.46$, $p = 0.0008$, $\beta = 1.31$, $SE = 0.38$, $R^2 = 0.69$, 128 *DPs*). Moreover, a closer look into the sub-scores revealed that the Parallel Bars techniques were harder to learn ($t(94.0) = 3.59$, $p = 0.0005$, $\beta = 1.84$, $SE = 0.51$, $R^2 = 0.55$, 128 *DPs*), harder to understand ($t(94.0) = 3.12$, $p = 0.002$, $\beta = 1.36$, $SE = 0.44$, $R^2 = 0.61$, 128 *DPs*), and required more learning effort ($t(94.0) = 3.61$, $p = 0.0005$, $\beta = 1.75$, $SE = 0.49$, $R^2 = 0.62$, 128 *DPs*) and more mental effort ($t(94.0) = 2.22$, $p = 0.029$, $\beta = 1.4$, $SE = 0.63$, $R^2 = 0.58$, 128 *DPs*). Overall, the use of Parallel Bars generally led to lower SUS scores in the study.

For qualitative analysis, we grouped the data into Single Ray and Double Ray groups, and compared user responses to find the effects of Orientation Cue. For Single Ray, 40% of observers perceived the Parallel Bars to provide useful orientation information to the observer by reducing directional ambiguity and adding necessary depth perception. This is reflected in their ranking, as 87.5% of observers preferred to have Single Ray with Parallel Bars over just Single Ray. For Double Ray, however, 53% of the observers found the Parallel Bars redundant and time consuming. As a result, only 59% of the observers preferred to have Double Ray with Parallel Bars over Double Ray.

5.3 RQ 3. Pointer Performance and Subjective Experience with Double Ray Technique

Figure 9 visualizes the time that pointers spent to specify the pointing rays. Note that the pointing experience only differed based on Number of Rays (i.e., the level of Orientation Cue had no effect on how the

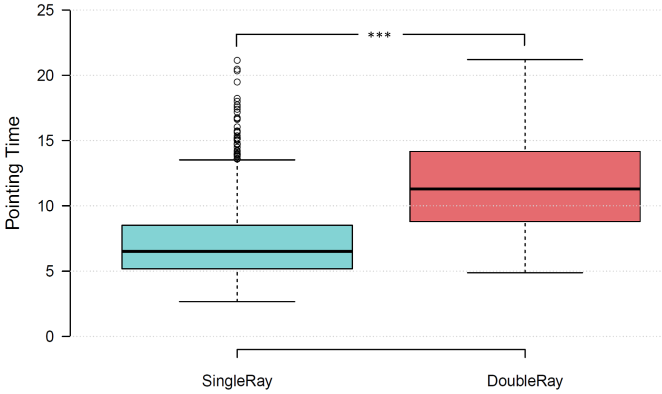

Fig. 9. Pointing time spent by the pointers using Double Ray and Single Ray techniques.

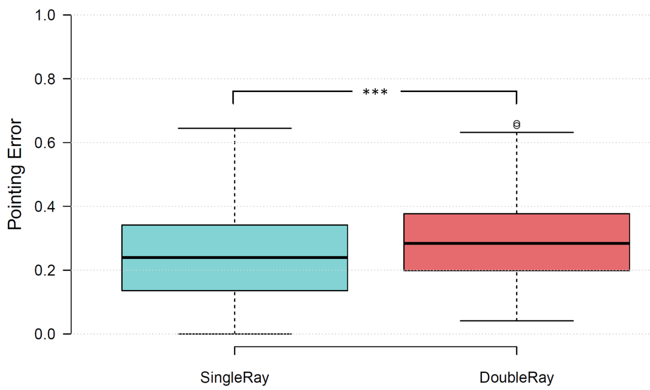

Fig. 10. Pointing error by the pointers using Double Ray and Single Ray techniques. Errors are measured in degrees.

pointing interaction occurred). An LMM indicated that Number of Rays had statistically significant effect on pointing time ($t(1682.1) = -33.83$, $p < 0.0001$, $\beta = -4.63$, $SE = 0.14$, $R^2 = 0.52$, 1699 *DPs*) with using Double Ray as the intercept. The increased pointing time using the Double Ray technique is expected, as the pointer must align two crosshairs with the controller.

Figure 10 plots the angular pointing error made by the pointers. Again, we used LMM to analyze the influence of Number of Rays on pointing error. We found that the pointers made significantly more error when using the Double Ray technique ($t(1718.3) = -7.96$, $p < 0.0001$, $\beta = -0.043$, $SE = 0.005$, $R^2 = 0.21$, 1735 *DPs*).

Regarding the NASA TLX score, the Double Ray technique was found to have higher workload ($t(31.0) = -2.07$, $p = 0.047$, $\beta = -4.83$, $SE = 0.40$, $R^2 = 0.6$, 64 *DPs*), as shown in Figure 11. A closer look into the sub-scores only revealed trending significance of the effects from Double Ray technique on mental demand ($t(31.0) = -1.99$, $p = 0.056$, $\beta = -1.43$, $SE = 0.72$, $R^2 = 0.64$, 64 *DPs*) and physical demand ($t(31.0) = -1.96$, $p = 0.059$, $\beta = -1.4$, $SE = 0.72$, $R^2 = 0.44$, 64 *DPs*).

When asked to rank the two techniques from a Pointer perspective, there was a mixed response. 53% of the pointers chose the Double Ray as their preferred technique. Overall, then, while pointing with the Double Ray is demonstrably slower, more error-prone, and more demanding, these drawbacks were not seen as fatal flaws by the pointers. We discuss this further in the next section.

5.4 RQ 4. Trade-off between Pointer and Observer

The Double Ray technique benefits observer accuracy (5.1) but reduces some aspects of pointer experience (5.3). To understand the trade-off between the pointer and observer, we needed to compare the pointer's cost and the observer's benefit. Given the nature of the collaborative target identification task, we prioritized observer accuracy over efficiency, since spending less time but getting the wrong answer meant that the overall task was a failure. Hence, we looked into the covariance between pointer performance (pointing time and pointing error) and observer accuracy. We used GLMMs to analyze the covariance and found that pointing time was not a significant covariate to observer accuracy ($z = 1.55$, $p = 0.12$, $\beta = 0.062$, $SE = 0.040$, $R^2 = 0.13$, 1566 *DPs*). However, pointing error was found to be a statistically significant covariate to the observer accuracy ($z = -3.98$, $p < 0.0001$, $\beta = -3.35$, $SE = 0.84$, $R^2 = 0.17$, 1566 *DPs*), as shown in Figure 12. These result suggest that the more erroneous the pointer is, the less likely the observer makes a correct target identification, whereas spending more pointing time does not help the observer become more accurate.

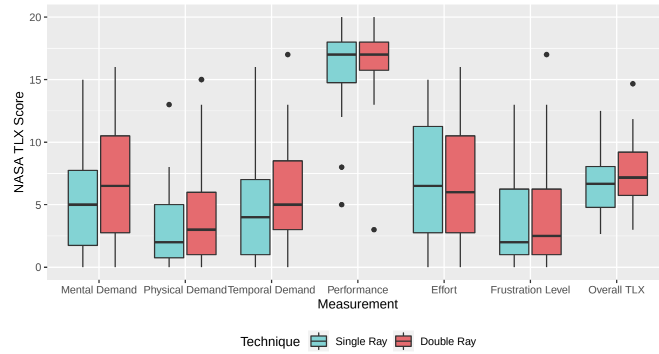

Fig. 11. Results from the raw NASA TLX score by different pointing techniques.

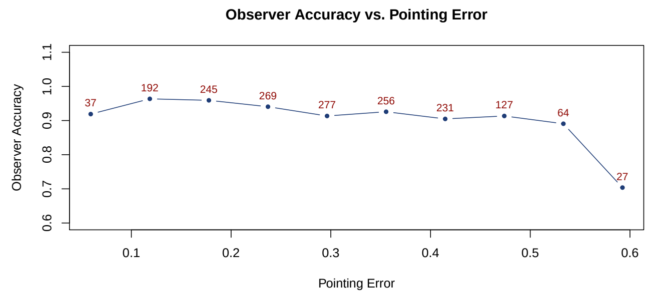

Fig. 12. The relation between pointer's pointing error and the observer's accuracy in the collaborative target identification task. The numbers denote the number of data points.

While this finding alone was not surprising, when combined with previous results, it shed light on the effect of the Double Ray technique in the pointer-observer trade-off. From Section 5.3, we learned that the Double Ray technique introduced more pointing error than the Single Ray technique. However, we also observed that the Double Ray technique led to higher observer accuracy, not lower. Our interpretation is that the benefit of the Double Ray technique trumps the negative influence of the increased pointing error. In other words, the Double Ray technique makes observers more tolerant to less-accurate rays. We discuss this further in the next section.

The results of the quantitative analysis are reflected in the user responses received from the post-study open-ended interview session. As presented in Sec 5.3, 53% of the pointers preferred Double Ray for the pointing task. We suggest that they were willing to sacrifice a bit of user experience, considering that the Double Ray helps the observer to be more accurate and confident in their selection. Since our study asked all participants to be both pointers and observers with all techniques, pointers would have understood both sides of the collaborative task when comparing Double Ray with Single Ray.

6 Discussion

We hypothesized that the Double Ray technique would significantly improve some aspects of the observer experience compared with the Single Ray technique (H1). We found good evidence to support H1 since the Double Ray technique was easier to learn, led to better observer accuracy, user confidence, and was favored by most observers in the pointing ray technique ranking. The visual "bracketing" effect created by the Double Ray technique was undoubtedly an effective enhancement to the simple "crossing" effect from the Single Ray technique. We proposed H1 based on the findings from our original study [32] where we found that the Double Ray technique was both more accurate and faster (for the observer) than the Single Ray technique. However, we did not find the same benefit in efficiency in the present study. We speculate that this was because of the human pointer. Unlike the prior work, which used a simulated pointer that cast perfectly accurate rays towards the target object, the pointers in this study were prone to make errors in pointing and led to a less ideal "bracketing" effect in the eyes of the observer, increasing the observing time. We also did not find the Double Ray technique to have a significantly higher modified SUS score. We suggest that the SUS score might not fully capture the usability for a collaborative task, since the user ranking demonstrated a dominant preference for the Double Ray technique over the Single Ray technique.

In our previous work [32], the Parallel Bars enhancement did not contribute to a higher observer accuracy, even though it increased observing time. At the time, we suspected that the visual clutter and insufficient spatial information caused the technique's ineffectiveness. Based on this, we added extra artificial spatial cues to the technique (the closest point and extended ray segment), and allowed observers to adjust the Parallel Bars parameters to avoid visual clutter. We hypothesized H2 in the hope that with our modifications, the Parallel Bars could be helpful in the collaborative target identification task. Unfortunately, our findings mirrored the prior work: increased observing time but no observer accuracy gain, even though more than 53% of the observers prefer to have the Parallel Bars). We speculate that the artificial orientation cues cannot provide detailed depth information, so that a small change in pointing direction is not noticeable from the Parallel Bars. While it still logically makes sense to exploit known spatial information to provide artificial spatial cues, a better design of such cues may be needed to prove its potential.

We hypothesized H3 as a natural extension to H1 and H2, combining the benefit of the Double Ray and the Parallel Bars techniques. While the Double Ray with Parallel Bars technique took observers a longer time, the technique was not found to improve observer accuracy over the Double Ray technique. However, even with its limited performance boost, the Double Ray with Parallel Bars technique was still preferred by most observers. The observers' comments confirmed that in the real world, the Double Ray technique might still suffer from visual ambiguity given the complexity of the environment and pointing errors made by the pointers. On the other hand, having both the Double Ray and the Parallel Bars enhancements might also cause visual clutter.

Given that the pointers needed to perform the extra task of aligning two crosshairs using the Double Ray technique, we expected a decrease in some aspects of user experience on the pointer's side (H4). H4 is well supported, as the data analyses suggested that when using the Double Ray technique, the pointers spent a longer time in pointing, yet still made more error than with the Single Ray technique. Even though we calculated the error as the arithmetic mean of the top ray and the bottom ray error, the Double Ray technique had increased pointing error, indicating that the pointing task was more challenging than the Single Ray pointing. We speculate that this was because fixing two crosshairs on two targets was more complex than fixing one. However, this increased demand was not clearly identified in the NASA TLX questionnaire as we only found trending significance in mental and physical demand, although the Double Ray technique was found to have a higher overall task load score than the Single Ray. When asked about their preference, most of the pointers chose the Double Ray technique, despite the decreased user experience. We are inclined to believe that when asked about the experience, the pointers also considered the observer benefit.

We wanted to understand the trade-off that the pointers have to make to help observers (RQ4), the results from the analysis indicate that pointer performance and subjective experience would pay off on the observer's side. The data analyses revealed that the Double Ray technique contributed to better observer accuracy, despite having higher pointing errors and workload, and despite the overall finding that a higher pointing error was likely to lead to lower observer accuracy. These seemingly contradictory findings actually demonstrate that the sacrificed performance and subjective experience from the pointer does pay off on the observer's side—the Double Ray technique helps make the observers more tolerant to pointing errors. Unlike the Single Ray, which relies on the accurate intersection between the ray and the target object's center, the Double Ray only requires the rays to visually "bracket" the target. Even though two rays may not precisely intersect the top and the bottom of the object, the two rays are also less likely to intersect with other objects in the environment simultaneously. By comparison, the Single Ray technique has lower pointing error but also leads to lower observer accuracy. This indicates that even though the Single Ray technique is simple, fast, and more precise in pointing, the visual "crossing" effect between the ray and the object's center is more sensitive to pointing error and less helpful to the observer.

6.1 Limitations

The experiment only uses lampposts as the referencing targets. Since our techniques only require objects to have top and bottom geometric features, the lampposts meet this criterion and contribute to varying spatial arrangements at the experiment site. However, the study did not consider objects of varying sizes and shapes, which may affect the techniques' performance. For example, pointing at smaller objects with the Double Ray technique may cause more ambiguity for the observer because of the smaller angle between the top and bottom rays.

Additionally, we used virtual markers at the top and bottom of the lampposts as a compromise for the execution of the user experiment. The virtual markers helped provide good visibility of the targets in an outdoor setting, ensure pointer and observer's agreement on the target position, minimize the impact of synchronization errors, and allow us to measure pointing error. We tried to preserve the model-free nature by making the ray overlay on top of the spherical markers, creating the same ambiguous visual effect as other real-world objects. The markers, by design, should not benefit the participants in a way that would hurt the validity of the study.

7 Conclusion and Future Work

Spatial referencing is important in many collaborative AR scenarios. In wide-area outdoor AR, referencing objects of interest at a distance can be highly challenging due to inaccessible physical environment geometric information. Working under model-free AR, the naïve usage of virtual pointing rays can lead to ambiguity and confusion. In this paper, we evaluated two pointing ray enhancements for remote object referencing in model-free outdoor collaborative AR.We introduced a synchronization method to align two AR users in a large-scale outdoor environment to enable an ecologically valid user study. Through this controlled study with participants playing the role of both pointer and observer, we compared four pointing ray techniques and found that: 1) the Double Ray technique was less usable for the pointer but contributed to better observer accuracy; 2) the Parallel Bars technique did not help with the user performance, but the participants preferred to have the technique. Our results partially replicated those of our previous study [32], but also added new understanding because of our use of true outdoor AR and a two-user collaborative task.

There are primarily two directions for future research. While the Parallel Bars demonstrated limited effectiveness, the user preference still suggested potential in providing artificial spatial cues to address the issues in distant object referencing in model-free AR. Hence, we plan to design and explore other ways of enhancing the user's spatial perception that remain effective in the vista space [12], such as motion parallax, relative size, or density. Additionally, a challenge in our work is to understand the trade-off between the pointer and observer user experience when they have different tasks and measurements. A method for computing a collaborative usability score that reflects such trade-offs would be useful to help designers and practitioners make critical decisions in collaborative system design, especially under an asymmetrical collaboration context.

References

[1] F. Argelaguet and C. Andujar. Efficient 3d pointing selection in cluttered virtual environments. *IEEE Computer Graphics and Applications*, 29(6):34–43, 2009. doi: 10.1109/MCG.2009.117

[2] F. Argelaguet, A. Kulik, A. Kunert, C. Andujar, and B. Froehlich. See-through techniques for referential awareness in collaborative virtual reality. *International Journal of Human-Computer Studies*, 69(6):387–400, 2011. doi: 10.1016/j.ijhcs.2011.01.003

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

[4] R. A. Bolt. "put-that-there": Voice and gesture at the graphics interface. In *Proceedings of the 7th Annual Conference on Computer Graphics and Interactive Techniques*, SIGGRAPH '80, p. 262–270. Association for Computing Machinery, New York, NY, USA, 1980. doi: 10.1145/800250.807503

[5] J. Brooke. Sus: a "quick and dirty'usability. *Usability evaluation in industry*, 189(3), 1996.

[6] J. Chastine and Y. Zhu. The cost of supporting references in collaborative augmented reality. In *Proceedings of Graphics Interface 2008*, GI '08, p. 275–282. Canadian Information Processing Society, CAN, 2008.

[7] D. Chatzopoulos, C. Bermejo, Z. Huang, and P. Hui. Mobile augmented reality survey: From where we are to where we go. *IEEE Access*, 5:6917–6950, 2017. doi: 10.1109/ACCESS.2017.2698164

[8] L. Chen, Y. Liu, Y. Li, L. Yu, B. Gao, M. Caon, Y. Yue, and H.-N. Liang. Effect of visual cues on pointing tasks in co-located augmented reality collaboration. In *Symposium on Spatial User Interaction*, SUI '21. Association for Computing Machinery, New York, NY, USA, 2021. doi: 10.1145/3485279.3485297

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

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

[11] C. Cruz-Neira, D. J. Sandin, T. A. DeFanti, R. V. Kenyon, and J. C. Hart. The cave: Audio visual experience automatic virtual environment. *Commun. ACM*, 35(6):64–72, jun 1992. doi: 10.1145/129888.129892

[12] J. E. Cutting and P. M. Vishton. Chapter 3 - perceiving layout and knowing distances: The integration, relative potency, and contextual use of different information about depth*. In W. Epstein and S. Rogers, eds., *Perception of Space and Motion*, Handbook of Perception and Cognition, pp. 69 – 117. Academic Press, San Diego, 1995. doi: 10.1016/B978-012240530-3/50005-5

[13] J. Davis and X. Chen. Lumipoint: multi-user laser-based interaction on large tiled displays. *Displays*, 23(5):205–211, 2002. doi: 10.1016/S0141-9382(02)00039-2

[14] A. Dix. *Computer Supported Cooperative Work: A Framework*, pp. 9–26. Springer London, London, 1994. doi: 10.1007/978-1-4471-2029-2_2

[15] B. Ens, J. Lanir, A. Tang, S. Bateman, G. Lee, T. Piumsomboon, and M. Billinghurst. Revisiting collaboration through mixed reality: The evolution of groupware. *International Journal of Human-Computer Studies*, 131:81–98, 2019. 50 years of the International Journal of Human-Computer Studies. Reflections on the past, present and future of human-centred technologies. doi: 10.1016/j.ijhcs.2019.05.011

[16] A. Erickson, K. Kim, G. Bruder, and G. F. Welch. Exploring the limitations of environment lighting on optical see-through head-mounted displays. In *Symposium on Spatial User Interaction*, pp. 1–8, 2020.

[17] A. Erickson, N. Norouzi, K. Kim, J. J. LaViola, G. Bruder, and G. F. Welch. Effects of depth information on visual target identification task performance in shared gaze environments. *IEEE Transactions on Visualization and Computer Graphics*, 26(5):1934–1944, 2020. doi: 10.1109/TVCG.2020.2973054

[18] S. Feiner, B. Macintyre, and D. Seligmann. Knowledge-based augmented reality. *Commun. ACM*, 36(7):53–62, jul 1993. doi: 10.1145/159544.159587

[19] B. Gordon. Keeping it real: the potential for accessible mixed-reality usability testing environments. 1 2022. doi: 10.25401/cardiffmet.19011608.v1

[20] N. Gutkowski, P. Quigley, T. Ogle, D. Hicks, J. Taylor, T. Tucker, and D. A. Bowman. Designing historical tours for head-worn ar. In *2021 IEEE International Symposium on Mixed and Augmented Reality Adjunct (ISMAR-Adjunct)*, pp. 26–33, 2021. doi: 10.1109/ISMAR-Adjunct54149.2021.00016

[21] S. G. Hart and L. E. Staveland. Development of nasa-tlx (task load index): Results of empirical and theoretical research. In *Advances in psychology*, vol. 52, pp. 139–183. Elsevier, 1988.

[22] K. Huo, T. Wang, L. Paredes, A. M. Villanueva, Y. Cao, and K. Ramani. Synchronizar: Instant synchronization for spontaneous and spatial collaborations in augmented reality. In *Proceedings of the 31st Annual ACM Symposium on User Interface Software and Technology*, UIST '18, p. 19–30. Association for Computing Machinery, New York, NY, USA, 2018. doi: 10.1145/3242587.3242595

[23] H. Joe. Accuracy of laplace approximation for discrete response mixed models. *Computational Statistics & Data Analysis*, 52(12):5066–5074, 2008. doi: 10.1016/j.csda.2008.05.002

[24] R. Jota, M. Nacenta, J. Jorge, S. Carpendale, and S. Greenberg. A comparison of ray pointing techniques for very large displays. Technical report, University of Calgary, 2009.

[25] S. Kim, A. Jing, H. Park, G. A. Lee, W. Huang, and M. Billinghurst. Hand-in-air (hia) and hand-on-target (hot) style gesture cues for mixed reality collaboration. *IEEE Access*, 8:224145–224161, 2020. doi: 10.1109/ACCESS.2020.3043783

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

[27] R. Kopper, F. Bacim, and D. A. Bowman. Rapid and accurate 3d selection by progressive refinement. In *2011 IEEE Symposium on 3D User Interfaces (3DUI)*, pp. 67–74, 2011. doi: 10.1109/3DUI.2011.5759219

[28] W. Lages, Y. Li, L. Lisle, T. Höllerer, and D. Bowman. Enhanced geometric techniques for point marking in model-free augmented reality. In *2019 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*, pp. 301–309, 2019. doi: 10.1109/ISMAR.2019.00028

[29] J. J. LaViola Jr, E. Kruijff, R. P. McMahan, D. Bowman, and I. P. Poupyrev. *3D user interfaces: theory and practice*. Addison-Wesley Professional, 2017.

[30] C. Lee, G. A. Rincon, G. Meyer, T. Höllerer, and D. A. Bowman. The effects of visual realism on search tasks in mixed reality simulation. *IEEE Transactions on Visualization and Computer Graphics*, 19(4):547–556, 2013. doi: 10.1109/TVCG.2013.41

[31] Y. Li, D. Hu, B. Wang, D. A. Bowman, and S. W. Lee. The effects of incorrect occlusion cues on the understanding of barehanded referencing in collaborative augmented reality. *Frontiers in Virtual Reality*, 2, 2021. doi: 10.3389/frvir.2021.681585

[32] Y. Li, F. Lu, W. S. Lages, and D. Bowman. Gaze direction visualization techniques for collaborative wide-area model-free augmented reality. In *Symposium on Spatial User Interaction*, SUI '19. Association for Computing Machinery, New York, NY, USA, 2019. doi: 10.1145/3357251.3357583

[33] Y. Lu, C. Yu, and Y. Shi. Investigating bubble mechanism for ray-casting to improve 3d target acquisition in virtual reality. In *2020 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*, pp. 35–43, 2020. doi: 10.1109/VR46266.2020.00021

[34] D. Maglia. The best lux meter app on ios in 2021. *https://dominikmaglia.medium.com/the-best-lux-meter-app-on-ios-bb14177d0060*, 2021.

[35] M. McGill, J. Gugenheimer, and E. Freeman. A quest for co-located mixed reality: Aligning and assessing slam tracking for same-space multi-user experiences. In *26th ACM Symposium on Virtual Reality Software and Technology*, VRST '20. Association for Computing Machinery, New York, NY, USA, 2020. doi: 10.1145/3385956.3418968

[36] C. Merenda, C. Suga, J. L. Gabbard, and T. Misu. Effects of "real-world" visual fidelity on ar interface assessment: A case study using ar head-up display graphics in driving. In *2019 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*, pp. 145–156, 2019. doi: 10.1109/ISMAR.2019.00-10

[37] S. Nakagawa and H. Schielzeth. A general and simple method for obtaining r2 from generalized linear mixed-effects models. *Methods in Ecology and Evolution*, 4(2):133–142, 2013. doi: 10.1111/j.2041-210x.2012.00261.x

[38] M. B. Neider, X. Chen, C. A. Dickinson, S. E. Brennan, and G. J. Zelinsky. Coordinating spatial referencing using shared gaze. *Psychonomic Bulletin & Review*, 17(5):718–724, Oct 2010. doi: 10.3758/PBR.17.5.718

[39] O. Oda and S. Feiner. 3d referencing techniques for physical objects in shared augmented reality. In *2012 IEEE International Symposium*

on Mixed and Augmented Reality (ISMAR), pp. 207–215, 2012. doi: 10.1109/ISMAR.2012.6402558

[40] J.-Y. Oh and W. Stuerzlinger. Laser pointers as collaborative pointing devices. In *Proceedings of the Graphics Interface 2002 Conference, May 27-29, 2002, Calgary, Alberta, Canada*, pp. 141–150, May 2002.

[41] D. R. Olsen and T. Nielsen. Laser pointer interaction. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '01, p. 17–22. Association for Computing Machinery, New York, NY, USA, 2001. doi: 10.1145/365024.365030

[42] J. K. Parker, R. L. Mandryk, and K. M. Inkpen. Tractorbeam: Seamless integration of local and remote pointing for tabletop displays. In *Proceedings of Graphics Interface 2005*, GI '05, p. 33–40. Canadian Human-Computer Communications Society, Waterloo, CAN, 2005.

[43] R. Pascoal, B. Alturas, A. de Almeida, and R. Sofia. A survey of augmented reality: Making technology acceptable in outdoor environments. In *2018 13th Iberian Conference on Information Systems and Technologies (CISTI)*, pp. 1–6, 2018. doi: 10.23919/CISTI.2018.8399155

[44] I. Rabbi and S. Ullah. A survey on augmented reality challenges and tracking. *Acta graphica: znanstveni časopis za tiskarstvo i grafičke komunikacije*, 24(1-2):29–46, 2013.

[45] K. Riege, T. Holtkamper, G. Wesche, and B. Frohlich. The bent pick ray: An extended pointing technique for multi-user interaction. In *3D User Interfaces (3DUI'06)*, pp. 62–65, 2006. doi: 10.1109/VR.2006.127

[46] R. J. Teather and W. Stuerzlinger. Assessing the effects of orientation and device on (constrained) 3d movement techniques. In *2008 IEEE Symposium on 3D User Interfaces*, pp. 43–50, 2008. doi: 10.1109/3DUI.2008.4476590

[47] B. H. Thomas and W. Piekarski. Glove based user interaction techniques for augmented reality in an outdoor environment. *Virtual Reality*, 6(3):167–180, Oct 2002. doi: 10.1007/s100550200017

[48] S. Voida, M. Podlaseck, R. Kjeldsen, and C. Pinhanez. A study on the manipulation of 2d objects in a projector/camera-based augmented reality environment. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '05, p. 611–620. Association for Computing Machinery, New York, NY, USA, 2005. doi: 10.1145/1054972.1055056

[49] B. L. WELCH. THE GENERALIZATION OF 'STUDENT'S' PROBLEM WHEN SEVERAL DIFFERENT POPULATION VARLANCES ARE INVOLVED. *Biometrika*, 34(1-2):28–35, 01 1947. doi: 10.1093/biomet/34.1-2.28

[50] S. Whittaker. Things to talk about when talking about things. *Human–Computer Interaction*, 18(1-2):149–170, 2003. doi: 10.1207/S15327051HCI18126