Assisting Spatial Referencing for Collaborative Augmented Reality

Yuan Li

Dissertation submitted to the Faculty of the
Virginia Polytechnic Institute and State University
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy
in
Computer Science & Applications

Doug A. Bowman, Chair
Joseph L. Gabbard
Wallace S. Lages
Sang Won Lee
Tobias Höllerer

April 13, 2022
Blacksburg, Virginia

Keywords: Augmented Reality, Collaboration, Spatial Referencing

Copyright 2022, Yuan Li
Assisting Spatial Referencing for Collaborative Augmented Reality

Yuan Li

(ABSTRACT)

Spatial referencing denotes the act of referring to a location or an object in space. Since it is often essential in different collaborative activities, good support for spatial referencing could lead to exceptional collaborative experience and performance. Augmented Reality (AR) aims to enhance daily activities and tasks in the real world, including various collaborations and social interactions. Good support for accurate and rapid spatial referencing in collaborative AR often requires detailed environment 3D information, which can be critical for the system to acquire as constrained by current technology. This dissertation seeks to address the issues related to spatial referencing in collaborative AR through 3D user interface design and different experiments. Specifically, we start with investigating the impact of poor spatial referencing on close-range, co-located AR collaborations. Next, we propose and evaluate different pointing ray techniques for object reference at a distance without knowledge from the physical environment. We further introduce marking techniques aiming to accurately acquire the position of an arbitrary point in 3D space that can be used for spatial referencing. Last, we provide a systematic assessment of an AR collaborative application that supports efficient spatial referencing in remote learning to demonstrate its benefit. Overall, the dissertation provides empirical evidence of spatial referencing challenges and benefits to collaborative AR and solutions to support adequate spatial referencing when model information from the environment is missing.
Assisting Spatial Referencing for Collaborative Augmented Reality

Yuan Li

(GENERAL AUDIENCE ABSTRACT)

People often exchange spatial information about objects when they work together. Example phrases include: “put that there”, or “pick the third object from left”. On the other hand, Augmented Reality (AR) is the technology that displays 3D information into the real world to enhance or augment reality. Scientists and technology practitioners think that AR can help people collaborate in a better way. The AR system needs to have a good understanding of the physical environment to support exchanging spatial information in the first place. However, limited by current technology, acquiring spatial information from the real world is not always possible or reliable. In this dissertation, we first illustrate the severity of insufficient environmental knowledge when collaborators sit next to each other in AR. Then we present pointing ray techniques to help AR collaborators refer to distant objects without knowing where those objects are. We further explore different marking techniques that can help the AR system calculate the position of a point in space without scanning the area. Last, we provide an AR application that supports efficient spatial information communication in remote discussion around physical objects.
Dedication

To my parents, my fiancée, and my family.
Acknowledgments

0.1 Personal Acknowledgments

The journey towards a Ph.D. is arduous. I would not be able to accomplish mine without the help and support from the kindness of people around me in the past five years.

First and foremost, I want to express my immense gratitude to my advisor and my role model, Dr. Doug A. Bowman. November 30, 2016, when I had my first meeting with Dr. Bowman, my life was changed when he offered me a chance to join the 3DI group at Virginia Tech. Through more than 150 enjoyable discussions, Dr. Bowman has taught me more than conducting research in academia. His intellectual charisma and nobleness have inspired me to a path of becoming a righteous, conscientious person in the future. For that, I cannot be more grateful.

I want to thank my parents, Yueke Li and Li Yuan, for their love and support along the way. Indisputably, my fiancée, Changqing Shi, has stood by me through all my travails and excitements with her tremendous understanding and encouragement. Had they not been there for me, I would not be able to keep pursuing my dream.

I would also like to acknowledge my committee members, Joseph L. Gabbard, Tobias Höllerer, Wallace S. Lages, and Sang Won Lee, for their supportive and encouraging guidance. Their insightful feedback has greatly cultivated this dissertation.
I want to thank several people from Virginia Tech that I had the privilege to work with. Their personal and professional support and guidance are what make this dissertation possible. They are my co-authors (all listed in alphabetical order): Shakiba Davari, Dr. David Hicks, Donghan Hu, Lee Lisle, Feiyu Lu, Prof. Akshay Sharma, Ibrahim Tahmid, Boyuan Wang, Run Yu, and Lei Zhang; other 3DI group members: Kylie Davidson, Cory I. Ilo, Leonardo Pavanatto; and administrative and departmental support: Mr. N. Dwight Barnette, Melanie Darden, Prof. Margaret Ellis, Teresa Hall, Andrea Kavanaugh, Sharon Kinder-Potter, Holly Williams and Melissa Wyers.

Finally, I would like to offer my sincere gratitude for the generous funding from the Immersive Sciences program in the Office of Naval Research and the rapid response program from the Institute for Creativity, Arts, and Technology.

0.2 Attribution

Dr. Doug A. Bowman is the Frank J. Maher Professor of Computer Science and the Director for the Center for Human-Computer Interaction (CHCI) at Virginia Tech. He is my Ph.D. advisor and the co-author of all the publications forming this dissertation.

Chapter 3 was in collaboration with Donghan Hu, Boyuan Wang, and Dr. Sang Won Lee. Donghan Hu is a Ph.D. student from Virginia Tech and contributed to the data analysis of the study. Boyuan Wang is a Master’s student at Virginia Tech and helped conduct the user study. Dr. Sang Won Lee is an Assistant Professor from the Computer Science department at Virginia Tech and helped co-guide the project with Dr. Doug A. Bowman.
Section 4.3 was in collaboration with Feiyu Lu and Dr. Wallace S. Lages. Feiyu Lu is a Ph.D. student from Virginia Tech and conducted the user experiment of the project. Dr. Wallace S. Lages is an Assistant Professor from the School of Visual Arts and helped design the techniques.

Section 4.4 was in collaboration with Ibrahim Tahmid and Feiyu Lu. Ibrahim Tahmid is also a Ph.D. student from Virginia Tech and helped conduct the study. Feiyu Lu contributed to designing the outdoor experiment.

Chapter 5 was in collaboration with Dr. Wallace S. Lages, Lee Lisle, and Dr. Tobias Höllerer. Dr. Wallace S. Lages designed the VectorCloud technique. Lee Lisle is a Ph.D. student from Virginia Tech and helped design the ImageRefinement technique and directly interacted with the participants. Dr. Tobias Höllerer is a Professor of Computer Science at the University of California, Santa Barbara. He contributed to the writing of the paper.

Chapter 6 was guided by Dr. David Hicks, Dr. Wallace S. Lages, Dr. Sang Won Lee, and Professor Akshay Sharma. Dr. David Hicks is a professor of the School of Education, and Professor Akshay Sharma is the chair of the Industrial Design at Iowa State University. The professors guided the design and implementation of ARCritique.

I am the primary contributor to all the work in this dissertation. No one else is claiming the work for their thesis or dissertation.
Contents

List of Figures xvi

List of Tables xxiii

1 Introduction 1

1.1 Background & Motivation ........................................ 1

1.1.1 The Potential of AR Collaboration ......................... 1

1.1.2 Spatial Referencing .......................................... 4

1.1.3 AR for Spatial Referencing .................................. 6

1.2 Challenges ....................................................... 8

1.3 Research Overview ............................................. 11

1.3.1 Research Questions .......................................... 12

1.3.2 Organization .................................................. 16

2 Related Work 18

2.1 Augmented Reality .............................................. 18

2.1.1 Displays ...................................................... 19

2.1.2 Tracking ...................................................... 23

2.1.3 Environment Reconstruction ............................... 24
2.2 Collaboration ............................................................. 26
  2.2.1 Awareness ............................................................ 26
  2.2.2 Referencing ........................................................... 27
  2.2.3 Communication Space ............................................. 30

2.3 Spatial Referencing in Collaborative AR ................................ 32
  2.3.1 Existing Approaches ............................................... 32
  2.3.2 Influence of Spatial Cues ........................................... 35

3 RQ1: Effect of Occlusion on Barehanded Referencing ................. 37
  3.1 Chapter Introduction .................................................. 37
  3.2 Experiment Design ..................................................... 38
    3.2.1 Experiment Task .................................................. 39
    3.2.2 Experiment Design ............................................... 41
    3.2.3 Participants ....................................................... 44
    3.2.4 Procedure .......................................................... 44
    3.2.5 Apparatus .......................................................... 45
  3.3 Results ................................................................ 48
    3.3.1 RQ1-A: Performance Decreases with Incorrect Occlusion .... 48
    3.3.2 RQ1-B: Collaborators Develop Compensatory Strategies in the AR Condition .............................................. 53
  3.4 Discussion ................................................................ 59
4.4 Pointing Ray Techniques in Model-Free Outdoor AR .......................... 92

4.4.1 Model-Free Outdoor Synchronization ................................. 92

4.4.2 Experiment Design ..................................................... 94

4.4.3 Study Results ............................................................. 104

4.4.4 RQ2-C. Pointer Performance and Subjective Experience with Double Ray Technique ................................................................. 108

4.4.5 RQ2-D. Trade-off between Pointer and Observer ......................... 110

4.4.6 Discussion .................................................................. 112

4.4.7 Summary of Outdoor AR Pointing Ray Study ............................. 115

4.5 Summary ....................................................................... 116

5 RQ3: Point Marking Techniques .................................................. 117

5.1 Chapter Introduction ......................................................... 117

5.2 Interaction Methods ......................................................... 118

5.2.1 VectorCloud ................................................................. 118

5.2.2 ImageRefinement .......................................................... 121

5.3 Simulated AR Study .......................................................... 123

5.3.1 Experiment Setup ......................................................... 124

5.3.2 Apparatus ................................................................. 126

5.3.3 Participants and Procedure ............................................. 126

5.3.4 Results ................................................................. 127
5.3.5 Discussion .................................................. 130
5.4 Outdoor AR Study ............................................. 131
  5.4.1 Apparatus ................................................. 131
  5.4.2 Participants and Procedure ............................ 132
  5.4.3 Results .................................................... 134
  5.4.4 Discussion ............................................... 136
5.5 Summary .......................................................... 137

6 RQ4. Supporting Spatial Referencing in Remote Critique 139
  6.1 Chapter Introduction ........................................ 139
  6.2 ARCritique Design and Implementation ..................... 140
    6.2.1 Design and Rationale ................................ 140
    6.2.2 Implementation ....................................... 146
  6.3 Evaluation .................................................... 149
    6.3.1 Tasks .................................................. 150
    6.3.2 Measures ............................................... 152
    6.3.3 Participants .......................................... 152
    6.3.4 Procedure ............................................ 153
    6.3.5 Analyses .............................................. 153
  6.4 Findings ..................................................... 154
<table>
<thead>
<tr>
<th>Section</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>A.6.1</td>
<td>User Consent Form</td>
<td>291</td>
</tr>
<tr>
<td>A.6.2</td>
<td>Questionnaires</td>
<td>296</td>
</tr>
<tr>
<td>A.6.3</td>
<td>IRB Approval Letter</td>
<td>318</td>
</tr>
</tbody>
</table>
List of Figures

1.1 Ford designers inspect and make adjustments on car model using AR headsets. Image source: link .................................................. 2

1.2 A hybrid example of collaborative AR with collocated and remote users (connected via AR and conventional video conferencing), interacting with virtual content. Image source: link .................................................. 3

1.3 Real-time groupware Time-Space Matrix adapted from Johansen, 1988 [70] .................................................. 6

1.4 Airbus partners with Microsoft on AR developer program in factories. Image source: link .................................................. 7

1.5 AR used to help first responders. Notably, the data is using computer vision on a drone, instead of accurate 3D capture of the scene. As a result, the firefighters are less informed of the situation. Image from: link .................................................. 10

2.1 Milgram and Kishino’s Reality-Virtuality Continuum from Milgram, 1994 [101] .................................................. 18

2.2 FOV of common AR eyewear in relation to human FOV, from Trepkowski, 2019 [143] .................................................. 20

2.3 Leonardo da Vinci recorded one of the earliest observations on the relation between object distance and brightness, from “The notebooks of Leonardo da Vinci” [32] .................................................. 22
2.4 ARKit world tracking combines information from the iOS device’s motion sensing hardware with computer vision analysis of the scene visible to the device’s camera. Image from: https://developer.apple.com/documentation/arkit/arworldtrackingconfiguration.

2.5 Example of KinectFusion with the censor orbiting the table, from Newcombe et al., 2011 [110] .................................................. 25

2.6 Shared gaze is represented as a yellow cursor moving over the scene. Target (only visible to one of the participants) is indicated by a red dot. The collaborators are working under time pressure to convey the location of the target. From Neider et al., 2010 [108] .................................................. 28

2.7 The 3D referencing techniques presented by Oda and Feiner [113]. (Left) The user’s view (for the indicator) through an AR HWD while initially specifying a spherical volume with a tracked Wii remote. The volume is placed at the intersection of a ray with the live depth map of the physical environment. Objects are overlaid with numerical annotations used only in the user study. (Middle) The texture-mapped depth mesh contained within the selection sphere is brought closer to the indicator. (Right) The indicator points directly with her finger at the representation within the close-up mesh of the physical object to which she intends to refer (in this case, the purple cylinder). 33

2.8 Effectiveness of various depth perception cues over distance (Adapted from Piekarski and Thomas [120]) ................................. 35

3.1 Incorrect occlusion in face-to-face communication. ................... 38
3.2 Experiment setup: The chairs were aligned with tape on the ground to form 180° and 90° configurations for the physical, face-to-face condition (up) and the physical, side-by-side condition (below), respectively. The target cubes were placed at fixed locations on the table. The laptop and monitor were adjusted by the participants for ease of use. The image was taken with a GoPro fisheye camera.

3.3 Illustration of the experimental setup. We measure the observer’s understanding (1 or 2) of the spatial references produced by the pointer. ◊ denotes the pointer’s egocentric perspective. Left: A face-to-face layout. Right: A side-by-side layout. This setup simulates a group of people surrounding shared small-scale AR content (e.g., an AR-based board game, medical training).

3.4 Flowchart of experiment UI for one experimental trial, starting from the upper-left pair of screens.

3.5 We displayed purple stripes to illustrate the field of view to participants. The virtual content inside the rectangle formed by the four stripes is what the viewer can see. Outside the rectangle, participants could see the real world, but not the virtual cubes. We designed the layout of the virtual cubes such that the FOV from the seating positions was sufficient to see the whole set of cubes, minimizing the influence of the limited FOV.

3.6 Overall results for accuracy rate (left), completion time (middle) and confidence values (right). Red lines with data labels indicate mean values, while black lines indicate median values.
3.7 Heatmaps of accuracy rate for the AR condition in two spatial configurations. (Left) Heatmap of accuracy rate in the AR/face-to-face condition and (Right) heatmap of accuracy rate in the AR/side-by-side condition.  

3.8 The thickness of each arrow represents the number of errors. The head and tail of the arrow indicate the observer’s selection and what the correct answer was. For example, a thick left arrow pointing from 2 to 1 visualizes the number of observers who answered Cube 1 when pointers were pointing at Cube 2. All errors are horizontal in the AR/face-to-face condition (Left), whereas there are some diagonal and vertical errors in the AR/side-by-side condition (Right).  

3.9 (Strategy I) pointing straight down from the top (Left), (Strategy II) penetrating into the virtual cube (Middle), (Other) grabbing a virtual cube (Right).  

3.10 Average accuracy rate, average completion time, and average confidence across trials in the AR condition (face-to-face, side-by-side). There was no apparent trend of improvement over repeated trials.  

4.1 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.  

4.2 First person view of different pointing rays (a) with or (b) without occlusion.
4.3 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.

4.4 The pointer’s view using the pointing interface to specify: A single ray (a& b), and two rays (c). The user presses and holds the “B” button to sample 100 rays. The crosshairs in (c) are controlled using the two joysticks to align with the top and the bottom of a lamppost simultaneously.

4.5 Example of an experiment trial of the participant’s first person view.

4.6 Levels of visual and spatial ambiguity used in the experiment.

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

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

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

4.10 Average error rate for Single and Double Ray under different VA combinations. Error bars represent standard error. Brackets indicate statistical significance with \( p < 0.0001 \).

4.11 Circular menu for the Observer. Left joystick is used to navigate through the options and the ‘A’ Button is used to confirm the selection.

4.12 The time that observers used to identify targets for different ray techniques.

\* means \( p \leq 0.05 \), \** means \( p \leq 0.01 \), \*** means \( p \leq 0.001 \).
4.13 Average accuracy the observers achieved using different pointing ray techniques. Accuracy was calculated across targets and participants. . . . . . . . 106

4.14 Pointing time spent by the pointers using Double Ray and Single Ray techniques. 109

4.15 Pointing error by the pointers using Double Ray and Single Ray techniques. Errors are measured in degrees. . . . . . . . . . . . . . . . . . . 110

4.16 The relation between pointer’s pointing error and the observer’s accuracy in the collaborative target identification task. . . . . . . . . . 111

5.1 Incorrect occlusion in face-to-face communication. . . . . . . . . . . . . 118

5.2 User’s view of naïve geometric AR marking implementation. . . . . . . 119

5.3 The VectorCloud technique. Users aim at the target from different locations while the rays are recorded. All pairs from two sets of rays are then intersected to obtain several sample positions. . . . . . . . . . . . . . . . . . . . 120

5.4 User’s view of ImageRefinement in the HoloLens. The user adjusts the ray by moving the crosshair in the 2D image. . . . . . . . . . . . . . . . . . 122

5.5 User’s view of ImageRefinement in the HoloLens. The user adjusts the ray by moving the crosshair in the 2D image. . . . . . . . . . . . . . . . . . 125

5.6 Marking error in the AR simulation experiment. . . . . . . . . . . . . . . 128

5.7 Hit rate in the AR simulation experiment. . . . . . . . . . . . . . . . . . . 129

5.8 Marking error in the outdoor AR experiment. . . . . . . . . . . . . . . . 134

5.9 Hit rate in the outdoor AR experiment. . . . . . . . . . . . . . . . . . . . 135

xxi
6.1 Two remote users holding smartphones view and interact with the same virtual object through ARCritique in a collaborative virtual environment. AR-Critique allows them to point to the virtual model while maintaining awareness of collaborators. 140

6.2 The workflow of ARCritique in an iterative design cycle. 1. The student scans the physical mockup using their phone; 2. The student views the scanned model on their phone to check model quality; 3. The application uploads a qualified scanned model onto a cloud server and shares it with remote users; 4. Multiple users can view the model in a synchronized shared AR space where they can independently point to and draw on the virtual model. 141

6.3 Overall procedure of mobile KinectFusion implementation. [*] is inspired by the work of Shen [129]. 147

6.4 An example of a critique session when the student is using the ray to guide the instructor’s attention. On the model, there are drawings from prior discussion. (a). Third-person perspective captured by the experimenter; (b). First-person perspective shared by the instructor; the instructor drew with small green markers and the student pointed with a red ray attaching to their phone avatar; (c & d). Spatially separated student and instructor synchronously in shared AR space. 151

6.5 Visualization for spatial referencing conversation time and word count. Jittering points are added to provide detailed distribution information. Significance is calculated using Kruskal-Wallis test with power = 0.99. ** means $p \leq 0.01$, *** means $p \leq 0.001$, **** means $p \leq 0.0001$. 164
List of Tables

5.1 The mean time for marking a target in VR for each target & technique, as well as overall means across targets or techniques. .......................... 129

5.2 The mean time for marking a target in AR for each target & technique, as well as overall means across targets or techniques. .......................... 136

6.1 Examples of spatial referencing conversations. ........................................ 163

7.1 Research dimensions covered in this dissertation. ................................. 173
List of Abbreviations

AR Augmented Reality: a technology that register virtual content into the real world to enhance or augment reality

CAD Computer-Aided Design: a technology for automated design process, usually involving creation and modification of 3D models

CSCW Computer Supported Cooperative Work: software tools and technology that supports a group of individuals working on shared projects

DOF Degrees of Freedom: number of dimensions that can change independently

FOV Field of View: the extent of things that can be seen through the eye or a camera, usually measured in degrees

HWD Head Worn Display: a device that is used to display virtual information and is worn on the user’s head

ID Industrial Design: a major studying the design process of physical objects that are meant for mass production

LiDAR Light Detection and Ranging: a technology determining ranges by targeting an object with a laser and measuring the time for the reflected light to return to the receiver

OST Optical See-Through: combine computer-generated imagery with semi-transparent mirrors through which the user can perceive the real world

RQ Research Question: the key question that a given research aims to seek answer to
SLAM Simultaneous Localization And Mapping: an approach of constructing a map of an unknown environment while simultaneously keeping tracking within it

TOF Time-of-Flight: the time of a light wave travel from source to an object and back

UI User Interface: a medium through which human user and the computer exchange and communicate information

VR Virtual Reality: a technology that surrounds the user with computer-generated virtual graphics, sounds, and other virtual information to immerse the user into a different world
Chapter 1

Introduction

1.1 Background & Motivation

1.1.1 The Potential of AR Collaboration

Although Augmented Reality (AR) research has been an active academic domain for over three decades, it is currently on the edge of its debut to the general public. With recent advancements in commodity technologies, it is now possible to apply AR to real-world problems. Tech giants like Apple and Microsoft are investing a significant amount of resources in the novel and reliable solutions (hardware and software) to win an upper hand in this promising market.

“Among the possible applications in AR, it is widely viewed that collaborative systems are to be among its killer applications.” [38] With AR aiming to enhance beyond the physical world’s limits, the technology opens the door to many potentials extending from our conventional collaborative activities. For instance, Ford, the automaker, announced its adoption on Microsoft’s HoloLens AR headsets in the design process. By utilizing AR, the auto designers can collaboratively stand in front of primitive car models and see 3D vehicle elements digitally overlaid onto the models, leading to fast evaluation and alters of new car designs, as shown in Figure 1.1. Similarly, scientists can also use AR to visualize and interact with datasets in 3D. The 3D visualization allows users to uncover trends and patterns that may
Figure 1.1: Ford designers inspect and make adjustments on car model using AR headsets. Image source: [link](#).

not be immediately visible with 2D visualizations while collaboratively interacting in 3D with AR brings a natural and intuitive data understanding. Furthermore, scientific visualization frequently requires experts with different backgrounds to cooperate closely because many valuable insights only occur in face-to-face discussions [44].

Previous research has demonstrated that collaborative AR can improve effectiveness for remote collaboration and creates a high degree of social presence compared to conventional video conferencing [13]. Similarly, prior user evaluations also conclude that AR could significantly enhance face-to-face collaboration by merging the task space and communication
space that are initially separated in desktop systems [14]. Through combined shared space in AR, participants can use the same non-verbal cues used in face-to-face conversations while also interacting with extensive AR content. As shown in Figure 1.2, AR could bridge co-located and remote users through a hybrid workspace. Better than conventional remote collaboration tools such as video conference, AR could employ highly detailed 3D representations for the remote users to convey accurate non-verbal cues like gaze direction, hand gestures to facilitate the communication in a shared virtual space.

In parallel, research that focuses on technologies supporting general forms of collaborations also has a long history. When first given the term groupware, early collaborative software denoted the initial endeavor in studying digital shared workspace. Over 30 years later, groupware has culminated into a rich domain consisting of theories about the nature of col-
laborations, the roles of the collaborators, and how collaboration can be beyond human’s past experiences. This domain also gains a different widely accepted name as Computer Supported Cooperative Work (CSCW). Over the past three decades, the priority for the AR researchers was not about enriching shared experience with nearby or remote collaborators but addressing pressing issues with implementation. The technological and hardware challenges have hindered AR systems from practical applications. The AR community only recently started to catch up with the CSCW providing new theories and design possibilities for collaboration [38].

The advancements in AR, in turn, affect our understanding of collaboration as well. For example, the early CSCW community developed collaborator embodiment theories based on the technology of the time. Nevertheless, modern AR can provide designers the ability to present collaborators with a much richer sense of presence. As noted by Ens et al. [38], there is a clear trend that AR-related collaboration research from the past twenty years has progressed from a focus on solving initial technical challenges in AR toward more meaningful investigations of collaboration. For instance, the improved network connectivity helps to the birth of many remote collaborative systems, whereas advanced sensing technologies allow for an easily shared local user environment in remote expert systems. With new capabilities, it is expected to observe a shift from an initial focus on improving the AR collaborative system from a technical perspective to the more in-depth explorations of collaboration in the collaborative AR community.

### 1.1.2 Spatial Referencing

Spatial referencing is one of the critical characteristics in CSCW that AR could potentially facilitate. Spatial referencing is used to describe the communication and confirmation of a
1.1. Background & Motivation

location in space. Referencing a target’s location to achieve joint attention on an object of mutual interest is a common requirement in many collaborative tasks ([7, 8, 27]). For example, when meeting together to inspect machines, factory workers may use their hands to point to different components. Similarly, at a larger distance, a firefighter on the ground may need to pinpoint where a firefighting aircraft pilot should drop chemicals on a wildfire.

Olson and Olson [117] introduced the concept of the spatiality of reference. In their study of collocated synchronous collaboration, they observed particular importance in the spatiality of human interactions. People and objects involved in the collaboration are located in space, and their roles in an ongoing discussion can be indexed by their locations. During the collaborative session, such locations are frequently referenced in the communication exchanging key ideas. Olson and Olson emphasized the importance of spatiality in groupware interactivity. However, limited by the technology at the time, they predicted poor support for this crucial element in the future.

Even though the concept of spatiality was brought from collocated face-to-face interactions, it can also be applied to a broader range of collaborations. Johansen [70] proposed the time-space matrix to categorize collaborative interactions in two orthogonal dimensions (Figure 1.3). The matrix delineates collaborations into four quadrants depending on when users work together (at the same or different times – synchronous vs. asynchronous), and the physical arrangement of where users work (in the same place or different places – co-located vs. remote). It should be expected that the spatiality of reference will play an essential role across all four quadrants. For example, in a video conferencing between a factory worker and a remote expert (the Remote-Synchronous quadrant), the expert needs to frequently refer to various objects on the workspace in front of the worker to give detailed instructions. Similarly, in an asynchronous art design lecture recording (the Asynchronous quadrants), the teacher might also refer to different parts of the art model to give detailed instructions.
1.1.3 AR for Spatial Referencing

As highlighted in the above section, spatial referencing is an essential factor in various collaborations. AR can enhance spatial referencing through interaction and visualization in 3D. Through 3D interaction, AR could provide the collaborator with a natural and intuitive way to initiate the reference of spatiality, such as highlighted hand pointing. As for the viewer, AR could also help reduce the efforts needed for understanding the reference significantly by converting it into apparent visual cues.

Assisting spatial referencing can be useful to support synchronous, collocated AR collaboration. This is mostly applicable to scenarios where the collaborators use AR head-worn displays (HWDs) to view, create and interact with virtual content to support tasks ranging from brainstorming and medical training to CAD design and factory inspection. In these cases, AR displays virtual content on top of the physical context. This shared spatiality is linked to the co-located space in which the objects and participants’ roles can be indexed by their locations, providing efficiency in communication. For instance, in a joint factory
1.1. BACKGROUND & MOTIVATION

Figure 1.4: Airbus partners with Microsoft on AR developer program in factories. Image source: link

Inspection, instead of excessive use of descriptive words like “part number 15”, the worker can simply reach out their hand to point to the referenced virtual content and say, “this one.” By highlighting such pointing through 3D visualization, the partner can quickly and accurately understand the reference, as shown in Figure 1.4.

The potential effectiveness of assisting spatial referencing in AR becomes greater as the distance between the collaborators increases or as the collaboration happens asynchronously. In synchronous collaborations, 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 with their hand 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 challenging to use at considerable distances.
Meanwhile, the collaborators need an excessive amount of descriptive vocal communication about the space information to reach consensus. In this case, AR displays can be used to communicate pointing direction and target information from one collaborator to another through virtual pointing rays. If the two collaborators’ relative position and orientation are known and the AR system has access to geometric information from the physical world, a ray indicating pointing direction can be drawn on the collaborator’s display. It is then up to the AR system to correctly visualize the virtual pointing ray to successfully establishing mutual understanding between users. Furthermore, applying AR for spatial referencing in asynchronous collaboration needs to ensure consistency over the time dimension. However, the core idea remains the same.

1.2 Challenges

Section 1.1 illustrates the potential of using AR to facilitate collaboration through spatial referencing. However, there are significant challenges before this future can come true.

First and foremost, 3D information about the environment containing the target is not always available. Obtaining 3D information of the physical world is a primary requirement for many AR systems because most, if not all, AR systems need it to register digital content in the space, provide tracking functionality for the user, render realistic 3D graphics, ensure realistic interactivity, and so on. With current technology, capturing the real world around the user can be achieved through a combination of pre-built models, on-board sensors, user interactions, and 3D model reconstruction algorithms such as Simultaneous Localization And Mapping (SLAM). However, even at near distances, reconstruction could fail.

Let us start with the most common collaboration scenario: synchronous, co-located, face-to-face communication. A unique advantage of face-to-face communication is visibility, defined
1.2. Challenges

as “being able to see each other” [26]. The advantages of co-located, in-person communication include nonverbal communication such as gestures, gaze awareness, and eye contact [117]. Among these non-verbal cues, deictic gestures are powerful for the spatial referencing tasks for two reasons: 1) it is almost an instinct for people to reach out to the target; 2) viewers of the deictic gestures are used to interpret the hand’s direction and look for the first physical object along the direction. However, barehanded referencing in collaborative AR may be problematic, since most AR HWD systems do not provide correct occlusion cues. The mis-occlusion is because capturing an accurate real-time model of the user’s hands can be difficult. Extreme lighting conditions, like in bright sunlight, can interfere with the capturing sensors. The use of AR headset indicates that the wearing user has to keep their hands in front of the headset, limiting their poses. Finally, the fingers are one of the most agile joints, and the dynamic nature brings severe challenges to capturing frequency. According to Cutting and Vishton [31], occlusion is the most dominant depth cue. Even though there have been multiple works investigating the incorrect occlusion cue in the AR community, the assessments and solutions focus on the egocentric view and pay little attention to the observers with a different perspective. However, this incorrect occlusion cue will not only confuse the user and jeopardize depth perception in AR, but also impede other users’ spatial understanding in a collaborative AR system. Figure 3.1 demonstrates the visual effect of incorrect occlusion from the collaborator’s view.

Then let us extend the interaction space and focus on co-located but distant synchronous collaboration. AR system working without any environment model is known as model-free AR [28]. While model-free AR may not be a common use case, it is likely to occur in high-stakes scenarios under extreme uncertainty conditions. 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 survivors’ coordinates.
after a natural disaster, as shown in Figure 1.5. 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. It is safe to assume that, in the foreseeable future, completely getting rid of model-free AR through the advancement in hardware is highly unlikely. In a model-free setup, a crude implementation of the natural spatial referencing approaches, such as laser pointing ray, will likely fail because it will extend to infinity and will not have correct occlusion with the physical environment, as shown in Figure 4.2

Thus, given the unreliable nature of the geometric information, another major challenge is how we can mitigate the negative influence through 3D User Interface (UI) design. With the system knowing little about the referenced target’s 3D information, replicating the natural-
istic occlusion effect people are used to in the real world becomes problematic. The virtual content that the HWDs renders will inevitably appear covering everything else in the physical space, leading to an incorrect occlusion cue to the end-users. As a result, users may need to resolve the ambiguity through conversation, and misunderstood spatial referencing can have significant negative effects in some applications (e.g., medical training). One possible solution to missing 3D information is to design 3D UIs to help the user directly specify the target’s location without knowledge from the environment.

Furthermore, even though the usefulness of spatial referencing in the conventional collaborative setup has been well studied, there is a lack of evidence that good spatial referencing has a positive impact on collaborative AR systems. There should be a systematic assessment of an AR collaborative tool that supports efficient spatial referencing and a formal evaluation of such a tool compared with traditional collaboration approaches such as Zoom meetings in a Remote-Synchronous setting.

## 1.3 Research Overview

Because spatial referencing is a fundamental activity among various collaborative activities, supporting efficient spatial referencing can be beneficial or even critical in different collaborative AR systems. This dissertation aims to address the above challenges related to spatial referencing in AR through multiple research projects. We first ask how the lack of occlusion cues in model-free AR affects understanding in collaboration (RQ1). Having established the understanding of significant negative effects, we then study how these effects can be mitigated through the design of AR pointing rays (RQ2) and interaction techniques for authoring spatial referencing (RQ3). Finally, we extend the dissertation into a more realistic setting by allowing verbal communication and incorporating imperfect models from the physical object
for spatial referencing (RQ4).

The four research questions described in this dissertation cover real-time groupware of various communication spaces: co-located, remote, and wide-area. From the simplest collaboration scenario (RQ1), we illustrate how poor spatial referencing caused by the missing model can harm collaboration. Next, instead of pursuing solutions directly in the co-located collaboration, we take one step forward into a more challenging wide-area space where both the need for an environment model and the difficulty of such model acquisition increase drastically. Under the model-free constraint, we pursue two solutions to address distant object referencing problems through RQ2&3. Lastly, we withdraw the model-free constraint and focus on supporting spatial referencing with current technology in a remote collaborative task requiring frequent spatial consensus (RQ4). By utilizing imperfect models and various spatial communication channels, we aim to better understand the influence of spatial referencing in a realistic collaborative design critique.

1.3.1 Research Questions

RQ1. How do occlusion cues and spatial relationships among users affect the accuracy of understanding near-field barehanded spatial reference?

In collocated synchronous collaboration, spatial referencing of target objects with bare-hand pointing is a natural way of communication in the real-world. The same approach of spatially referencing targets of interest could also happen in AR, where the referenced objects are virtual. In the physical world, users can perceive the depth of hands and objects correctly through various depth cues, among which occlusion is the most effective. In AR, however, without precise tracking of users’ hands, occlusion cues between the hands and virtual targets will not be present, compromising spatial communication. Little quantitative evaluation of
this problem has been conducted. This research primarily aims to study the impact of missing occlusion cues on the understanding of near-field bare-hand spatial referencing in AR. We are also interested in how users adapt referencing methods, such as changing pointing gestures, to mitigate the influence of missing occlusion cues.

**Contributions**

- Evaluation of the extent to which incorrect occlusion impacts the performance of bare-handed referencing in a shared, model-free AR scene, in comparison with the correct occlusion seen during referencing of physical objects

- An empirical understanding of how users perceive and address the challenge of incorrect occlusion for bare-handed referencing

**RQ2. How do pointing rays in model-free AR facilitate an accurate understanding of a spatial reference at a distance?**

It is often essential for users to understand their collaborators’ gaze direction or gazed target. With an AR display, a virtual 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 great distances, due to the lack of occlusion cues when a model of the environment is unavailable.

Thus, we propose to design novel pointing techniques to improve gaze ray effectiveness. These approaches should be model-independent and only rely on AR devices’ local tracking capability. Through a controlled experiment in simulated AR, we plan to illustrate the effectiveness of communicating locations of objects of interest at a distance through these visualization without any environment model. Based on the findings, we can further extend the study to a more ecologically valid outdoor AR setup to evaluate the pointing techniques’
usability in the real world for both the observer and pointer.

**Contributions**

- 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

- With perfect tracking and accurate collaborator, the Double Ray visualization has a significant positive effect on collaboration effectiveness.

- A method for synchronizing the localization of two AR users in a large-scale outdoor environment

- New insight into the trade-off between the pointing technique usability and observer performance

- An understanding of the implications of our findings for future model-free AR spatial referencing interface design

**RQ3. How can we create accurate 3D spatial references of position when the environment model does not exist?**

Accurate 3D spatial referencing requires knowledge of the target position in space. If the AR system has a reliable geometric model of the physical environment, the system can extract position information easily. Unfortunately, a reliable environment model is not always available. Assuming that the AR device is capable of self-tracking, geometric triangulation could determine an accurate 3D location of the target point by intersecting two or more 3D rays that pass through the target without the need for any environment model (even in model-free AR). However, a naïve approach that directly generates 3D rays based on the
1.3. Research Overview

user’s head position and orientation could easily suffer from human motor control issues and therefore leaves room for improvement.

This research aims to explore ways to enhance geometric triangulation techniques to specify the location of an arbitrary point in space and evaluate their effectiveness and usability in ecologically valid settings. The findings could guide us to prove that this approach could be used to create accurate spatial reference points even without any environment model.

Contributions

- Evidence of reasonable accuracy of the geometric marking techniques at distances up to 85 meters
- The design of multiple enhanced geometric marking techniques
- The evaluation of enhanced techniques in both simulated AR and outdoor AR
- The benefit to AR applications where the environment model information is unavailable or unreliable.

RQ4. How does the ability to do collaborative spatial referencing in shared AR space impact the effectiveness of communication about 3D objects?

Due to Covid-19, people are advised to maintain social distancing, and students and faculty are suggested to take limited contact during lectures and research activities. This constraint has a greater impact on some disciplines, such as Industrial Design (ID), where a common form of teaching and research takes place as face-to-face critique sessions. During the stay-home order period, students and faculty could only use Zoom video conferences to carry out critique sessions. After interviewing with ID faculty, we observed complaints regarding the use of excessive descriptive verbal communication for spatial referencing during the Zoom calls, leading to decreased efficiency.
This research aims to design a collaborative AR application to address the spatial referencing issue during Zoom meetings, a conventional approach for critique based collaboration. The AR application, namely ARCritique, utilizes an imperfect model of the physical object from scanning, and it aims to support discussions around the model for remote collaborators. We conduct an ecologically valid user experiment where we invite novice ID students and faculty to compare our AR application with Zoom on the communication of 3D objects to help us evaluate our application. By allowing verbal communication and other AR referencing techniques, RQ4 helps us understand the impact of spatial referencing in a more realistic collaborative task where reaching spatial consensus is critical.

**Contributions**

- Concept, design, and implementation of ARCritique, a novel mobile collaborative AR application to support remote critique sessions around physical objects
- Findings from a user study ($n = 10$), evaluating the benefit of ARCritique in supporting spatial communication
- Discussion of the implications for future AR collaborative interface design
- Illustrating that collaborative tasks requiring frequent spatial consensus can benefit from different spatial referencing channels that current technology allows

### 1.3.2 Organization

This dissertation is structured to address the research questions directly. After discussing related work, the third chapter addresses RQ1. A completed study investigates how incorrect occlusion impacts referencing performance and seeks to understand how users adapt to the challenge. Chapter 4 focuses on RQ2 and the design of different pointing ray techniques.
The chapter then discusses two completed studies in both simulated AR and outdoor AR. Chapter 5 deals with RQ3 and describes how different geometric triangulation techniques were designed. It then discusses the implementations as well as two completed evaluation studies in simulated AR and outdoor AR. Chapter 6 discusses RQ4, where we design and implement a novel collaborative AR application for remote design critiques centering around physical objects. The application is evaluated through an ecologically valid user experiment. Chapter 7 provides a summary of this dissertation and future directions for supporting spatial referencing in collaborative AR.
Chapter 2

Related Work

2.1 Augmented Reality

Since this dissertation is rooted in collaborative AR research, it is necessary to briefly review AR technology development as it serves as the technological foundation. Because the dissertation employs Virtual Reality (VR) as a proxy to AR and they are linked in the Reality-Virtuality spectrum [101] (Figure 2.1), it is worth noting some of the VR related technologies as well.

![Figure 2.1: Milgram and Kishino’s Reality-Virtuality Continuum from Milgram, 1994 [101]](image-url)
2.1.1 Displays

From the human-computer-interaction point of view, the term “display” has a broader range than what people usually understand [87]. As a necessary component of any computer system that communicates to the user, the display devices are designed to present information to one or more of the user’s senses through the human perceptual system. These senses typically include but are not limited to: visual, auditory, haptic, olfactory, and so on. Among these senses, visual takes roughly 70% of the brain’s overall sensory information and serves as the dominant sense that both AR and VR have mostly concentrated on [59]. The dissertation is no exception and uses visual displays as the primary output system.

Ivan Sutherland delineated AR and VR as “the ultimate displays” in 1965 [135], where he described the future as: “The ultimate display would, of course, be a room within which the computer can control the existence of matter. A chair displayed in such a room would be good enough to sit in. Handcuffs displayed in such a room would be confining, and a bullet displayed in such a room would be fatal. With appropriate programming such a display could literally be the Wonderland into which Alice walked.” Sutherland proceeded in the direction towards “the ultimate display” and presented one of the world’s earliest AR Head-Mounted Displays (HMDs) in 1968 [136]. The so-called “Sword of Damocles” system had one key primitive functionality: the display was tracked so that the viewer could see virtual objects as if they were real in the physical space. After years of endeavor, modern AR/VR displays even allow the users to interact with virtual content while navigating around in space. The major distinction between AR and VR displays is that while VR enables the user to be in a completely virtual world and experience beyond physical limits, AR allows the user to enjoy the real world with virtual content created to enhance ordinary life [127]. Milgram and Kishino presented the widely accepted Reality-Virtuality Continuum [101] explaining the design space that encapsulates AR and VR (Figure 2.1). The continuum has only one
Figure 2.2: FOV of common AR eyewear in relation to human FOV, from Trepkowski, 2019 [143]

dimension with two extremes: reality (the purely physical world) and virtuality (the purely virtual world). While it is clear that VR resides at the virtuality end, AR appears in-between the spectrum, closer to the reality end, introducing virtual objects into real-world scenes [5]. Although there have been efforts to use VR headsets to detect and display real-world objects, AR still allows users to see and understand the physical environment in a much easier manner. However, AR displays, especially optical see-through (OST) displays, face a more significant implementation challenge and are much limited by the hardware, leading to worse characteristics.

Several essential characteristics must be considered when discussing visual displays in the dissertation.

**Field of View:** The field of view (FOV) refers to the maximum number of degrees of visual angle that can be seen instantaneously on a display. OST AR displays commonly suffer from
2.1. **Augmented Reality**

a narrow FOV. For example, Microsoft HoloLens 1 has a diagonal FOV of 35 degrees, the HoloLens 2 has a diagonal FOV of 52 degrees, and the Magic Leap One has a diagonal FOV of approximately 50 degrees. Comparatively, the human visual system has a binocular FOV exceeding 210 degrees horizontally and 150 degrees vertically [72]. Consequently, the augmented virtual contents only appear in a small portion of human vision, as illustrated in Figure 2.2. Comparatively, VR devices generally have considerably larger FOVs. Studies have shown that limited FOV causes visual perceptual performance decrements in both real and virtual environments [9]. The majority of studies on FOV have been performed with VR and have shown that the restrictions from FOV could degrade the ability to develop spatial knowledge and navigate ( [2, 68, 151]), while also result in decreased task performance in searching and locating a target [143]. While there are relatively fewer studies on the effect of FOV using AR devices, several researchers have identified the potential negative impact of narrow FOV on AR-based visual search tasks. Kishishita et al. [78] used a wide FOV AR display and analyzed secondary task performance in a search task between in-view and in-situ annotations with different FOVs. They found that the difference in these conditions decreased as the FOV exceeded 100 degrees. Ren et al. [122] performed a wide FOV AR study where users had to find information through annotations connecting far apart objects with different FOVs. The results indicated a smaller FOV resulted in slower task completion than full FOV. Recently, Trepkowski et al. [143] focused on the available technology and analyzed the effect of narrow FOVs in visual search performance. Their findings suggested that visual search performance could significantly benefit from increased FOV, even in a limited, narrow FOV range. In summary, it should be noted that the change in FOV in simulated AR and outdoor AR with OST HWDs will likely cause a difference in performance on visual search and spatial reasoning tasks. Hence in our studies, we often conduct studies in simulated AR to gain initial insights and proceed to replicate the study in outdoor AR to validate our results in more ecologically valid settings.
CHAPTER 2. RELATED WORK

Figure 2.3: Leonardo da Vinci recorded one of the earliest observations on the relation between object distance and brightness, from “The notebooks of Leonardo da Vinci” [32]

Brightness and Contrast: Maintaining sufficient contrast in OST displays is generally challenging. In particular, most displays cannot produce enough brightness to fight against natural light. This problem is more pronounced in the outdoor environment where the display’s maximum brightness must compete with the real world’s brightness level, making it very difficult to obtain an acceptable contrast level. Apart from the obvious impact on virtual content visibility, the brightness level will also affect depth perception. Given the same object size and distance, the brighter objects will appear closer than the dimmer. This effect is well known in art, and even Leonardo Da Vinci discussed it in his notebook [32] (Figure 2.3). The phenomenon has been studied thoroughly in psychology and found consistent results at both near ([4, 41]) and medium [29] fields. Singh et al. [131] replicated the phenomenon using an AR haploscope. They used a between-subjects design and invited different groups of ten observers to see and match the real targets with the bright AR targets and the dim AR targets. The study found the most accurate matches for dim AR targets, which were
more closely matched the brightness of real targets. One of the theories that explain this effect is that brighter objects stimulate a larger area on the retina and that brighter objects affect pupil size, which biases other near triad reflexes. In general, it is notable that the brightness of AR displays should also impact spatial referencing interfaces through visual and depth perception. In the outdoor studies described in this dissertation, we considered the limitations in brightness and try to mitigate the influence from real-world lighting conditions to ensure visibility.

### 2.1.2 Tracking

Both AR and VR systems need to obtain pose information in a coordinate system to display virtual contents correctly at their registered locations in that coordinate system. The operation of sensing and measuring such pose information is known as tracking. Tracking usually runs in real-time to ensure low latency response in the AR and VR systems. Degrees of freedom (DOF) is a concept used to describe the number of independent dimensions needed to specify a measuring system. Registering a virtual object in 3D space usually requires 6 DOF: three DOF for the position and three DOF for the orientation [127]. There are different sensing technologies developed throughout the years to support tracking, including mechanical sensing, acoustic sensing, magnetic sensing, etc. Since each tracking technology has its distinct advantages and disadvantages, state-of-the-art mobile AR/VR devices normally obtain the best results by simultaneously using all available sensors, as shown in Figure 2.4. Tracking technologies are essential throughout this dissertation, as it enables fundamental functionalities including synchronization of collaborating users, registering content in space, and 3D geometric calculations.
Environment reconstruction is often used in AR systems for different purposes. Environment reconstruction can be used for tracking. These methods assume there is no reference model (hence model-free) and does not rely on markers (hence marker-free). SLAM, which is sometimes called visual odometry [112], can achieve continuous 6DOF tracking of a camera pose in an unknown environment. The approach originates from robotics and is designed to support incremental tracking. However, along its pipeline, SLAM constructs a model of the environment that the camera captures. Apart from robotics, SLAM found its application in AR due to the globally consistent 3D reconstruction it produces [21]. Combining with other sensors, SLAM is widely used on mobile AR HWDs like Microsoft HoloLens and Magic Leap One to track the device in 6DOF in the physical environment.

Besides tracking, AR could also directly benefit from environment reconstruction to obtain...
3D information from the physical world. KinectFusion [110] is a good example. Named after the Microsoft Kinect device, the system uses light-emitting depth-sensing time-of-flight (TOF) cameras. The system can generate a real-time mapping of complex and arbitrary indoor scenes in variable lighting conditions. Figure 2.5 shows an example of simultaneous reconstruction of a physical scene, around which the Kinect orbits. However, the reconstruction range is restricted by the light source and constrained to a conservative range of 0.4 to 8 meters. Since then, there has been no lack of similar work in the field. Erat et al. [39] combined the use of drones and Microsoft HoloLens and let the user control the drone via gaze direction from an exocentric perspective on the drone. The drone is equipped with on-board sensors to continuously map the surrounding environment and streams it to the pilot standing behind a wall. Based on the spatial understanding, the pilot looks in the direction they want the drone to go. Given that the drone has great mobility, this approach is capable of reconstructing a distant environment. However, the lighting conditions will still interfere with the reconstruction. Model reconstruction is particularly important to Chapter 6 where we implemented KinectFusion to generate a virtual model based on a physical artifact.
2.2 Collaboration

2.2.1 Awareness

Research on facilitating multi-user collaboration dates back to early synchronous shared window systems. Lauwers et al. [86] highlighted the concept of “collaboration awareness” and emphasized its importance in meeting usability requirements. Gutwin et al. [54] adopted the idea and developed techniques to highlight collaborator actions.

In VR and AR, collaborative awareness is needed to understand other users’ perspectives sharing the same virtual or real space. Researchers have reported that better understanding of other users’ perspectives can lead to increased performance or usability in collaborative object-focused interaction tasks. For example, Hindmarsh et al. ([60, 61]) investigated object-focused interaction in collaborative virtual environments through a task where participants were asked to arrange the layout of furniture in a virtual room collaboratively and agree upon a single design. The participants were given conflicting priorities to encourage debate and discussion. The studies reveal that the organization and coordination of much co-present work are facilitated by the ability to ‘monitor others’ activities. This awareness of others’ actions and perspectives was not well supported by technology at the time. Yang and Olson [156] designed a collaborative navigation task in a shared virtual environment and found evidence to support the benefit for the users to understand others’ perspectives. In a more recent endeavor, Luff et al. [95] reported a series of collaborative experiments with t-Room, an immersive system that presents full-scale, real-time images of remote co-participants. Their studies highlight the need to understand how in everyday work settings, participants can produce recognizable actions and make sense of those of others, even when the resources they draw upon and their location do not remain stable. Tang et al. [139] concluded that knowing where one’s collaborators are looking helps support effortless joint
references, leading to easier collaboration.

2.2.2 Referencing

Mutual awareness depends on communication. When communicating about work objects in a visual environment, “what is shown and how it is shown is crucial.” \[153\] Referencing is often the essential component that bridges understanding between participants in groupware design \[35\]. Deictic gestures, gaze visualizations, and ray pointing are commonly used in human collaboration.

Deictic gesture: According to Dix \[35\], deictic gestures are explicit hand gestures used in cooperation to indicate different objects or artifacts relevant to the shared work. Typical examples of deictic gestures include pointing (indicate objects, areas, locations, and directions \[141\]), drawing (showing paths and shapes \[10\]), and describing (indicating orientations, distances, or sizes \[10\]). Deictic gestures can help people convey things that are difficult to put into words, hence simplifying verbal communication and improving collaboration efficiency. These gestures are pervasive in face-to-face collaborations where people have an excellent visual of each other’s hands. The deictic gestures contribute to establishing a frame of reference, a coordinate system where certain objects, actions, and phenomena are gauged \[124\]. For instance, by seeing other’s pointing gestures, the viewer maintains an understanding of a referent via either an absolute frame of reference (i.e., the pointer points northward) or a relative frame of reference (i.e., the pointer points to their left) \[92\]. This ability to interpret deictic gestures is also essential to people’s daily life and is known to be more accurate than spatial language in encoding spatial information \[132\]. The requirement of a close-range collaboration to ensure deictic gestures’ visibility can be waived through visual representations of the users’ arms and bodies \[11, 77\], or abstract pointers and im-
Figure 2.6: Shared gaze is represented as a yellow cursor moving over the scene. Target (only visible to one of the participants) is indicated by a red dot. The collaborators are working under time pressure to convey the location of the target. From Neider et al., 2010 [108]

ages [51, 155]. In summary, there is abundant evidence that supports using hand gestures to facilitate human collaboration where spatial information is exchanged.

**Gaze Visualization:** Similar to gestures, the gaze information can also help convey spatial information. Sharing of the user’s gaze information (shared gaze) in collaborative systems as a means of referencing has been well supported. Brennan et al. [18] demonstrated that using shared gaze is more efficient than verbal communication as a rapid approach to exchange spatial information. Similarly, Neider et al. [108] set up a collaborative task where one user
needs to convey spatial information to the other user under time pressure. They found that the shared gaze was both more effective and efficient than speech. Cherubini et al. [24] proposed an algorithm that combines the movements of the collaborators’ eyes in the shared workspace, the users’ utterances, and the explicit referencing gestures to detect misunderstanding for collaboration at a distance. The work revealed the possible connection between a user’s gaze information and their linguistic expression and showed the potential of using simple interaction mechanisms like explicit referencing gestures to sustain collaborative work at a distance.

**Pointing Rays:** Pointing rays are widely used in VR for different selection tasks at a distance [71]. 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 [15, 30]. The technique is based on pointing gestures that are pervasive in the user’s daily life [146]. Compared with other conventional 2D interfaces such as mouse-based pointing, ray pointing does not rely on a physical surface [142]. 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 [118].

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 [115, 116]. In 3D object referencing, pointing rays generally suffer from low accuracy because a small angular error can be significantly magnified as distance increases and often require enhancements [33, 82, 87]. The pointing ray generally appeals to the observer as well, as researchers reported it to be simple, naturalistic, and easy to understand [71, 123].
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 ensure correct intersection with real world objects and accurate depth cues such as occlusion, especially at large distances.

### 2.2.3 Communication Space

Johansen [70] proposed the time-space matrix to categorize collaborative interactions in two orthogonal dimensions - space and time. The space dimension divides groupware into two categories: collocated (participants are in the same place) and remote (participants are in different places). Schmalstieg and Höllerer [127] discussed the connection between communication space and task space (where actual work happens) and proposed that these two factors combined could have a significant impact on communicating and performing a collaborative task. In this dissertation, we focus on three different communication spaces.

**Co-located Collaboration:** This scenario usually denotes near-field collaborative tasks. The object of mutual interest can be virtual [12, 52], where the AR system displays the content at the same location for the collaborators to see. The target object can also be physical. In these cases, the AR system projects annotations and augmentations onto the physical object to enable collaboration opportunities beyond physical constraints [102, 144]. A great advantage of near-field co-located collaboration is that the users share the same spatial frame, and therefore, it is usually easier to achieve spatial consensus with well-supported interfaces. Other advantages of near-field co-located collaborative AR systems include an easier understanding of 3D structures than screen-based presentation [44, 75] and no requirement for networking connection [74].
Remote Collaboration: This scenario is quite the opposite of co-located collaboration in the sense that the collaborators reside in different spaces and are synchronized through shared virtual information. One often studied use case is the “remote expert”, which typically involves a remote knowledgeable person guiding a local person in a physical task [38]. Commonly used guiding annotations include drawing and pointing. Gauglitz et al. [49] introduced a system in which a remote expert can see a local user’s spatial context and annotate on top of it. The system then displayed the annotations anchored to real-world objects in the local worker’s space. The authors demonstrated the effectiveness and efficiency of conveying reference information using free-hand drawing. Pointing, as widely used for selection tasks [87] in virtual environments, can also be used to pinpoint the point of interest. Gurevich et al. [53] implemented TeleAdvisor to enhance the remote assistance experience by combining a pointer and drawing with a teleoperated robotic arm. Oda et al. [114] explored the use of virtual replicas for remote assembly assistance in both VR and AR. They presented two approaches using simple 3D markers as annotation and virtual lines to indicate the alignment and correspondence between the physical object and its virtual replica.

Wide-area Collaboration: This scenario is not covered in the original time-space matrix by Johansen [70]. It usually happens in an outdoor environment with sufficiently large space and is gaining more and more attention with increasing effort to bring AR into outdoor spaces [119]. A unique feature that distinguishes wide-area collaboration from the co-located collaboration is that, while the collaborators reside in the same space but with a distance between them, the visibility of the collaborator and the target object reduces drastically, causing severe problems in communication and understanding of spatial frames. Even though researchers have explored various methods to enable AR in the outdoor environment for several decades [43, 64, 65], building a wide-area collaborative AR system remains difficult. A major challenge comes from establishing a reliable connection and synchronization
mechanism between distant collaborators [99]. Without a perfect model from the physical environment, such a challenge will remain critical and inspire researchers to investigate different solutions. However, the implementation of wide-area collaboration will be essential to use AR in the wild, meet people’s social needs, and enable new collaboration opportunities.

2.3 Spatial Referencing in Collaborative AR

2.3.1 Existing Approaches

AR researchers have explored various systems to enhance collaboration. Chastine et al. [22] studied the virtual pointer as a visualization for reference cues in collaborative AR. They used hand tracking to allow the pointer to specify a referencing arrow to guide a remote collaborator. To the authors’ surprise, the virtual interactive arrow pointer alone was perceived as an invaluable reference tool and less favorable than a static 2D virtual grid, resulting in the extra need for verbal clarification. Moreover, combining the arrow pointer and the 2D grid yielded the best collaborative experience. They concluded that poor reference cues would generate referential ambiguity, incurring such additional communication costs as time and computational resources, while carefully designed interaction techniques would significantly help collaborate. Stafford et al. [134] presented the Hand of God metaphor on a world-in-miniature to support the first collaborative world-in-miniature system. It supported a table-top display user to collaborate with a VR user through natural gestures and speech. The system was capable of providing guidance on object manipulation but was also limited by issues such as scale and reconstruction drawbacks. Oda and Feiner [113] explored 3D referencing techniques in shared AR. Using a depth camera to capture the user’s pointing direction, they compared their hand-gesture referencing technique with other controller-
2.3. Spatial Referencing in Collaborative AR

Figure 2.7: The 3D referencing techniques presented by Oda and Feiner [113]. (Left) The user’s view (for the indicator) through an AR HWD while initially specifying a spherical volume with a tracked Wii remote. The volume is placed at the intersection of a ray with the live depth map of the physical environment. Objects are overlaid with numerical annotations used only in the user study. (Middle) The texture-mapped depth mesh contained within the selection sphere is brought closer to the indicator. (Right) The indicator points directly with her finger at the representation within the close-up mesh of the physical object to which she intends to refer (in this case, the purple cylinder).

based techniques and found that their gestural referencing technique was significantly more accurate when the participants had sufficiently different views of the shared scene. They suggested an advantage in the use of hands for referencing. However, their approach heavily relied on an external depth camera to capture the user’s hand and the physical objects. More recently, Kim et al. [77] 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 that only relied on hand tracking.

While interaction techniques exist for referencing virtual and physical targets in AR-based co-located collaboration ([22, 77, 91, 113]), these systems typically require users to use tracked controllers, gloves, or a stylus. These controllers limit the expressiveness afforded by bare-handed interaction. For example, holding controllers could prevent users from performing certain tasks that require their hands, such as making particular gestures or typing on a keyboard. Thus, it is beneficial to explore solutions that support object referencing
with bare hands in collaborative AR. Despite prior research, there is still no complete and reliable solution to the problem in the case of dynamic objects like hands, nor is there an understanding of the extent to which incorrect occlusion cues decrease the effectiveness of bare-handed referencing in AR.

There have been studies exploring various gaze visualization techniques facilitating the understanding of distributed collaborator’s perspectives. A commonly adopted method for providing gaze awareness is to determine the object one user is looking at and then place a virtual element such as a dot or a cursor on that object in the other user’s view ([1, 145]). 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 does not apply to model-free AR due to the lack of information about the real-world environment. Another approach is to present an image of one user’s view to another user ([45, 73]). Ideally, after the viewer inspects the image, they 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.

Some systems use a radar view to facilitate collaboration awareness ([37, 126]). 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.

To summarize, conventional solutions to assist spatial referencing often require extra input device or a good understanding of the task environment [105]. Little research has been
conducted to investigate the issues in spatial referencing in AR collaboration, especially under the model-free constraint.

2.3.2 Influence of Spatial Cues

Spatial referencing’s effectiveness depends on the user’s spatial perception, or depth perception, to be more precise. According to the taxonomy presented by Cutting and Vishton [31] (Figure 2.8), 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). Occlusion
is the strongest depth cue across all distance ranges. But only three other depth cues are reasonably effective in vista space. The 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. [84] 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 [42], 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 [80]. Researchers have explored other depth cues to improve perception accuracy. Diaz et al. [34] 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 [137]. Our dissertation can add to the literature on spatial cues in the vista space.
Chapter 3

RQ1: Effect of Occlusion on Barehanded Referencing

3.1 Chapter Introduction

In this chapter, we investigate to what extent user performance is negatively influenced without proper support for spatial referencing in a simplistic co-located collaborative task. Face-to-face communication plays an important role in our daily lives and workspaces [157]. During such communication, joint attention is often achieved via dietic gestures. When bare-handed referencing is used to denote nearby physical objects, the intent of the pointing gesture is usually very clear in the real-world; however, this may not be true when collaborative AR systems are used. Specifically, when using model-free AR, the system cannot visualize virtual content being occluded by real-world objects. This incorrect occlusion cue will not only confuse the user and jeopardize depth perception in AR, but also impede the spatial understanding of other users in a collaborative AR system. Thus, to understand how to design effective collaborative AR systems, we first need to develop a solid understanding of the challenge posed by incorrect visual occlusion of the user’s hand gestures, which is the heart of RQ1.
3.2 Experiment Design

In order to evaluate the effect of incorrect occlusion cues on bare-handed referencing during collaboration in AR, we conducted a controlled experiment wherein one participant played the role of a pointer who needed to refer to a given target with their hands, and the other played the role of an observer who needed to identify the correct target. The target could either be a physical object or a virtual object displayed in the AR environment. Figure 3.1 shows the observer’s view in the physical (left) and AR (right) conditions. Additionally, we varied the spatial relationship between the participants and the targets, because we believed that differences between the pointer’s and observer’s perspectives might lead to different visual perceptions for the observer.

We designed the experiment to investigate two RQs and associated sub-questions:

- **RQ1-A**: How do the incorrect occlusion cues in model-free AR affect performance in understanding another user’s barehanded spatial references?
  
  - **RQ1-A-i**: To what extent is the understanding of spatial references negatively influenced by model-free AR?
3.2. **Experiment Design**

- **RQ1-A-ii:** How do spatial configurations of collaborators and target locations affect understanding in model-free AR?

- **RQ1-B:** What strategies do collaborators develop to overcome incorrect occlusion cues?
  - **RQ1-B-i:** What strategies do pointers adopt to better communicate spatial references in model-free AR?
  - **RQ1-B-ii:** What strategies do observers adopt to better understand spatial references in model-free AR?

### 3.2.1 Experiment Task

There were two roles involved in the study: the pointer and the observer. The referencing targets were sixteen cubes, arranged in a four-by-four grid shape. The target cubes were either solid wood cubes or virtual cubes displayed in AR HWDs (as shown in Figure 3.1). Regardless of the cube type, the cubes’ upper surfaces had red labels from one to sixteen as identifiers. We also made sure that the virtual cubes were correctly aligned for both the pointer and the observer by using Vuforia’s image recognition algorithm\(^1\) along with a printed picture to place the virtual cubes in constant positions across sessions. Apart from changing the cube type, we also varied the collaborators’ spatial relationships throughout the experiment. Participants sat side-by-side (i.e., at 90° angles to each other with respect to the cubes) or face-to-face (i.e., at 180° angles to each other). The primary task in the experiment was for the pointer, who was given a number from one to sixteen, to point to the target cube with their dominant hand; the observer was asked to identify the target cube’s label by interpreting the referential gesture. The participants were asked to perform this task as accurately as possible, with speed of task performance being a secondary goal. The

\(^1\)https://library.vuforia.com/features/images/image-targets.html
Figure 3.2: Experiment setup: The chairs were aligned with tape on the ground to form 180° and 90° configurations for the physical, face-to-face condition (up) and the physical, side-by-side condition (below), respectively. The target cubes were placed at fixed locations on the table. The laptop and monitor were adjusted by the participants for ease of use. The image was taken with a GoPro fisheye camera.

In summary, we investigated the following factors’ impacts on referencing objects with the hands.

- Cube type: physical vs. AR
- Seating position: face-to-face vs. side-by-side
- Target position: the position of the target cube in the 4×4 grid layout

In each study session, one participant was assigned to the role of the pointer, while the other participant was the observer. In order to minimize variance, we ensured that all pointers
were right-handed. We did not switch participants’ roles during the experiment to avoid bias introduced by learning.

The goal of the study was to understand to what extent and how spatial referencing in AR is influenced by model-free AR (which has incorrect occlusion) by comparing performance and user behavior in an AR environment to the same aspects of spatial referencing in a physical environment. We decided to use pointing at physical cubes as the base condition because this represents the gold standard for spatial referencing in real-life situations with correct occlusion effect, and the scenario that AR systems attempt to simulate. While changing from the physical condition to the model-free AR condition introduces differences other than incorrect occlusion from the observer’s perspective (e.g., texture and shadows), based on what we observed in the preliminary study and what has been reported in the literature [31, 79, 90, 93], incorrect occlusion remains the primary difference between the two conditions. Therefore, comparing model-free AR with a physical condition can help us gain insight into the effect of incorrect occlusion on collaborative spatial referencing.

### 3.2.2 Experiment Design

The experiment followed a 2 (cube type: physical or AR) × 2 (seating position: face-to-face or side-by-side) within-subjects design for a total of four conditions, as depicted in Figure 3.3. The two seating positions tested were selected to model various settings where people sit around a table or share the same perspective with respect to physical artifacts. Similar settings have been studied in prior CSCW works concerning the spatial arrangements of remote communication systems or tabletop interfaces [57, 139, 154]. In each condition, the participant pair needed to complete a set of 20 trials (one trial for each of the 16 targets, plus four additional trials as decoys to prevent participants from predicting
CHAPTER 3. RQ1: EFFECT OF OCCLUSION ON BAREHANDED REFERENCING

Figure 3.3: Illustration of the experimental setup. We measure the observer’s understanding (1 or 2) of the spatial references produced by the pointer. @ denotes the pointer’s egocentric perspective. Left: A face-to-face layout. Right: A side-by-side layout. This setup simulates a group of people surrounding shared small-scale AR content (e.g., an AR-based board game, medical training).

remaining targets). In deciding condition orders, we grouped virtual and physical conditions to minimize configuration changes made between trials, resulting in a total of eight condition orders. We also prepared a total of eight distinct pseudorandom target sequences of cube numbers. We counterbalanced conditions and target sequences across all participants to avoid bias.

We measured task success (correct vs. incorrect responses), task completion time, and the observer’s self-reported confidence level. We also gathered users’ subjective feedback through an exit interview and filmed the experiment sessions for post-study observation (as shown in Figure 3.2). We computed the accuracy rate as the percentage of correct trials based on task success. To compute task completion time, we measured the elapsed time from when the pointer pressed ENTER key to display the target cube’s number on their screen (Timestamp 1 in Figure 3.4) until the observer clicked a button to submit their final selection (Timestamp 2 in Figure 3.4). We did not separate the pointing and observing time because there was no reliable way to determine the completion of a pointing gesture for different users without complicating the experiment procedure (e.g., extra button click). Meanwhile, we observed from our pilot study that the pointing time did not contribute to variance in the overall
3.2. Experiment Design

The pointer presses ENTER key.
(Timestamp 1)
The observer presses the touchpad.
The observer confirms the number.
(Timestamp 2)
The observer confirms the confidence level.

Figure 3.4: Flowchart of experiment UI for one experimental trial, starting from the upper-left pair of screens.

performance. Therefore, we decided to compute the total task time, starting from the time when the pointer saw the number and ending when the observer made their final decision. We also recorded the observer’s confidence level on a scale from one to seven, where one meant “not at all confident”, and seven indicated “very confident”.

In the exit interview, we focused on gathering subjective feedback from the participants to understand how cube type (physical vs. AR) influenced participants to consciously change their referencing gestures, and to see if participants had any specific strategies to cope with the visual occlusion problem. The questions were as follows:

- Pointer:
  1. “Did you point to a virtual object in the same way as you did to a physical object? If so, why?”
  2. “Did you adapt your pointing gesture across trials? If so, why?”

- Observer:
1. “How easy was it to recognize the target? How was recognizing AR targets different from physical objects?”

2. “What strategy did you use?”

3.2.3 Participants

Thirty-two participants (14 females, 18 males, mean age: 22.53 years) were recruited from the authors’ university. The participants assigned to the pointer role were all right-handed. Seventeen of the participants had experienced virtual or augmented reality before. All of them had normal or corrected-to-normal vision. This experiment was approved by the university’s Institutional Review Board. If participants agreed to participate, they were offered $12 in appreciation for their efforts.

3.2.4 Procedure

Participants were welcomed upon arrival and asked to read and sign an informed consent form. Then, they were invited to take a demographic survey. After the survey, the moderator introduced the experiment and tasks to the participants. The participants were reminded that in this experiment, only deictic gestures were allowed to reference the target object. Having confirmed that the participants understood the experiment, the moderator then explained the experiment procedures in detail and asked participants to practice the tasks in both experimental conditions until they felt confident.

We developed a program which gives instructions to the collaborators and measures the performance of the observer. The task procedure is illustrated in Figure 3.4, where the screenshots of the program are connected through user actions. To record the starting time, we put a keyboard in front of the pointer. When the pointer pressed the Enter key, a new trial
3.2. Experiment Design

started and a target number was shown on the screen (the lower-left screen in Figure 3.4). The pointer was asked to point at the target cube using any gesture they preferred. The observer was instructed to view the pointer’s gesture to determine which cube was being referenced. Then, the observer selected the target number from a radial menu designed to minimize selection time (the lower-right screen in Figure 3.4). We recorded the end time when the observer clicked on the “Confirm” button in the center. After confirming their selection, they were asked to rate their confidence level on a scale of 1 to 7, as described earlier in this section (and as shown in the upper-right screen in Figure 3.4). Each trial ended when the observer entered their confidence level. The control would then switch back to the standby UI for the pointer to start another trial (the upper-left screen in Figure 3.4). The participants completed 20 trials for each experiment condition. After the participants finished all four conditions, we conducted an exit interview.

3.2.5 Apparatus

In the physical conditions, the targets were sixteen 3.6 cm wooden cubes in a $4 \times 4$ layout glued on a wooden base (the image on the left in Figure 3.1). The layout of multiple cubes is modeled after applications that have small-scale visual context shared among users (e.g., board game, playing cards, sticky notes, urban planning, etc.) In the AR conditions, the virtual cubes were rendered at the same size and arranged in the exact same way as the wooden cubes. We used two Microsoft HoloLens 1 devices as the AR displays. To ensure that the virtual cubes were correctly aligned in both participants’ views, we used the HWD’s front camera and Vuforia to recognize a computer-generated, image-recognition-friendly picture we placed on the table in front of the participants. Once the HWDs detected the picture on the table, the AR system could then render the virtual cubes at the position of the recognized picture. Since both HWDs looked at the same image, it was guaranteed that the virtual cubes
were seen in the same location by both users. Once the virtual cubes were stable relative to the image target, the moderator used a controller to turn off the image recognition function and freeze the cubes in space. After the image target was no longer needed, it was removed from the table to avoid the participants seeing it during the AR conditions. To change the participants’ spatial relationships, we invited them to change seating positions from face-to-face seating to side-by-side seating. We added four purple stripes at the borders of the AR headset’s screen to show the wearer the small effective field of view, as shown in Figure 3.5.

The seating positions and the image target position were carefully chosen so that when the participants sat in their assigned seats, they were able to see all sixteen virtual cubes inside the display’s field of view. To ensure consistency, we marked the positions of the chairs on the ground and image target on the table with tape. Between the AR conditions, after the participants changed their seating positions, we performed a recalibration process using the image target to again ensure that each participant saw the virtual cubes in the same physical locations.

The study program (Figure 3.4) was coded as a web-based application consisting of three UIs: the pointer UI, the observer UI, and the moderator UI, where moderators recorded the study information and controlled trials. During the study, the web program ran locally with XAMPP 2 on a 13-inch Apple MacBook Pro. It was accessed via two Chrome windows: a window on the main monitor displayed the observer UI and the moderator UI; a window on an external monitor displayed the pointer UI. The moderator UI was only opened in between conditions for study setup; during trials, the window on the main monitor displayed the observer UI instead. We used a 12-inch Apple iPad Pro as the external monitor, connected with a cable to ensure both screens had similar size and visual quality. We chose to run both the observer and pointer UI on one computer with two different displays to minimize

2https://www.apachefriends.org/index.html
3.2. Experiment Design

Figure 3.5: We displayed purple stripes to illustrate the field of view to participants. The virtual content inside the rectangle formed by the four stripes is what the viewer can see. Outside the rectangle, participants could see the real world, but not the virtual cubes. We designed the layout of the virtual cubes such that the FOV from the seating positions was sufficient to see the whole set of cubes, minimizing the influence of the limited FOV.

The effect of network delays on measured completion time.

The external monitor and an external keyboard connected by a cable were used for the pointer side, while the MacBook Pro and its built-in trackpad were used for the observer side. A two-second tone was played when pointers pressed the Enter key to start pointing and when observers clicked the trackpad to finish observing. These tones provided audio feedback to the participants and experimenters. To analyze the participants’ behavioral patterns and interview responses, we used a GoPro 7 video camera to record the entire session (See Figure 3.3.)
Figure 3.6: Overall results for accuracy rate (left), completion time (middle) and confidence values (right). Red lines with data labels indicate mean values, while black lines indicate median values.

3.3 Results

3.3.1 RQ1-A: Performance Decreases with Incorrect Occlusion

Our results indicate that the model-free AR condition negatively affects performance in referencing objects with bare hands compared to real-world referencing. The participants in the AR condition had lower accuracy rates, took more time to recognize targets, and were less confident about their choices. The spatial configuration (seating and target positions) affected participants’ performance in the model-free AR condition.

Overall results are shown in Figure 3.6. A power anova test in the R package *pwr* and effect size analysis (cliff.delta and cohen.d) in the R package *effsize* indicated sufficient power of our results. We used power calculations for balanced one-way analysis of variance tests (ANOVA power analysis reported 99.3%).
3.3. Results

Accuracy Rate

As expected, in the AR condition, the average accuracy rates across 16 pairs for face-to-face and side-by-side seating positions (62.5% with $\sigma = 0.49$ and 66.0% with $\sigma = 0.47$, respectively) were lower than those in the physical condition (99.6% with $\sigma = 0.06$ and 100.0% with $\sigma = 0.00$, respectively), as shown in Fig 3.6 (middle). In the case of the physical condition, there was only one inaccurate answer out of 512 trials, in which the observer selected cube 14 when the correct target was cube 4. We ran a bias reduction generalized linear model (brglm) in the R package brglm\textsuperscript{3}, which is a flexible generalization of ordinary linear regression, to test the effects of cube type and seating position on accuracy rate \[109\]. We applied brglm for accuracy rate analysis due to the following reasons: 1) the response, accuracy data, is binomial distribution, 2) generalized linear model provides benefits in small data sets, and 3) brglm has improvement over traditional maximum likelihood. The test result indicated that cube type had a statistically significant effect on accuracy ($p < 0.001$, $d = 0.36$), while seating position did not.

The accuracy rate varied depending on the target position. A heatmap of accuracy is shown in Figure 3.7. As the performance per target position (16 positions) highly depended on the cube types (2 types), the generalized linear model performed poorly due to multicollinearity. Instead, we used a bias reduction generalized linear model on target position in each AR condition. We ran the effect of target position in three different ways: column (4 levels), row (4 levels), and Manhattan distance from the center (3 levels: four central cubes ($D = 1$ for 6, 7, 10, and 11), four corner cubes ($D = 3$ for 1, 4, 13, 16), and eight lateral cubes ($D = 2$)) for both conditions. In the AR/face-to-face condition, the effects of cube column ($p < 0.001$, $d = 0.25$) and Manhattan distance from the center ($p < 0.05$, $d = 0.11$) on accuracy rate were significant, with accuracy being best for the first column and worst for

\textsuperscript{3}https://www.rdocumentation.org/packages/brglm/versions/0.7.2/topics/brglm
CHAPTER 3. RQ1: EFFECT OF OCCLUSION ON BAREHANDED REFERENCING

Figure 3.7: Heatmaps of accuracy rate for the AR condition in two spatial configurations. (Left) Heatmap of accuracy rate in the AR/face-to-face condition and (Right) heatmap of accuracy rate in the AR/side-by-side condition.

In the AR/side-by-side condition, the effect of cube column \( (p < 0.01, d = 0.21) \) and Manhattan distance from the center \( (p < 0.01, d = 0.16) \) on accuracy rate were significant, with accuracy being best for the corner cubes and worst for the third column.

Figure 3.8 visualizes all the errors (incorrect answers) that the participants made in the grid of cubes. Each arrow points from the correct answer to an observer’s answer, and the thickness of the arrow indicates the number of observers who made the same mistake; for example, the thin arrow at the bottom of Figure 3.8-Left indicates that one observer chose 3 when the pointer pointed at 1 in the AR/face-to-face condition. The biggest arrow in Figure 3.8-Left represents six errors, and the biggest arrow in Figure 3.8-Right represents five errors. In the case of the AR/face-to-face condition, all errors are horizontal, and all of them except two cases are off-by-one errors.

In general, the left-facing arrows are broader than the right-facing arrows. This indicates that more observers made errors in which they chose a cube further away from them, picking cubes closer to the pointers than the ones that the pointers were actually pointing at. One
Figure 3.8: The thickness of each arrow represents the number of errors. The head and tail of the arrow indicate the observer’s selection and what the correct answer was. For example, a thick left arrow pointing from 2 to 1 visualizes the number of observers who answered Cube 1 when pointers were pointing at Cube 2. All errors are horizontal in the AR/face-to-face condition (Left), whereas there are some diagonal and vertical errors in the AR/side-by-side condition (Right).

Possible explanation for this trend is that pointers may have referenced cubes by the direction of their finger (i.e., using an imaginary ray extending from a fingertip to point at a cube). In the meantime, from the other side, observers may not have been able to judge the position and direction of the observer’s fingertip clearly due to the incorrect occlusion, so they made decisions based on which target the fingertip appeared to be in. This effect can be seen in Figure 3.1, where the fingertip in the AR condition appears to be in cube 2 or 3, while it is actually indicating a ray pointing at cube 4.

In the case of the AR/side-by-side condition, much more complex patterns appeared, with more diagonal and vertical (upward) errors. In general, we anticipated that in the side-by-side condition, observers would have more cues that they could use from the lateral view: for example, arm direction or the length of the arm. However, the result did not show any systematic patterns of errors that are apparent.
CHAPTER 3. RQ1: Effect of Occlusion on Barehanded Referencing

Completion Time

Overall, the average completion time for all AR conditions was 6.08 seconds with $\sigma = 1.77$, and the average time for all physical conditions was 4.64 seconds with $\sigma = 0.93$. In the AR condition, the average completion times across 16 pairs (512 trials) for face-to-face and side-by-side seating (5.99 seconds with $\sigma = 1.25$ and 6.19 seconds with $\sigma = 2.18$, respectively) were greater than those of the physical condition (4.79 seconds with $\sigma = 1.00$ and 4.49 seconds with $\sigma = 0.83$), as shown in Fig 3.6 (middle). Because the time distribution does not satisfy the assumption of normality, we performed a log-transformation of all completion time data points and confirmed the normality of the result. We used a linear mixed-effect model, LMER, with the lme4 package in R [25]. In contrast to traditional approaches, LMER allows controlling for the variance associated with random factors without data aggregation. Besides, we used the lmer for completion time analysis because completion time data is numerical and continuous. And this model could fit better with small sample size data and multiple parameters. We discovered a significant interaction between cube type and seating position ($F(1, 1008) = 9.09, p < 0.01, d = 0.17$), indicating that the differences in completion time cannot be explained by considering cube types and seating positions separately. The result of a post-hoc analysis with LMER revealed that the effect of cube type on completion time was significant regardless of seating position ($F(1, 1008) = 492.34, p < 0.001, d = 1.00$). The effect of seating position was significant only in the physical condition ($F(1, 496) = 18.33, p < 0.01, d = 0.32$), where the side-by-side position was slightly faster than the face-to-face condition. We could not find any systematic pattern for the effect of target position on completion time, except that Cube 1, which is closest to the pointer, was the fastest across all conditions.
3.3.2 RQ1-B: Collaborators Develop Compensatory Strategies in the AR Condition

We aim to understand how participants perceive and cope with the challenge posed by incorrect occlusion. We used a thematic analysis approach to analyze the interview transcripts and observed study recordings in order to identify (1) challenges which pointers and observers faced during collaboration in the AR condition, (2) pointing strategies that pointers used in the AR condition, and (3) identification strategies that observers used in the AR condition.

Pointers’ Strategies

A few pointers reported difficulty in making a pointing gesture. Three pointers reported that incorrect occlusion caused difficulty when cubes overlapped their fingers. A majority of the pointers (9 out of 16) reported no difficulty in pointing, although their observers did not perform any better than the average (average accuracy rate: 61.1%).

- Incorrect occlusion: “My finger in the screen was overlapped.” (P10)

- Incorrect occlusion: “It feels like the block is being projected over my finger.” (P13)

Five pointers reported that the lack of touch caused difficulty (average accuracy rate: 69.3%). They mentioned that the lack of physical feedback was an obstacle in knowing whether their fingers had reached the correct cube.

- Lack of touch: “It was sometimes a little more difficult because I felt like my finger was going, like, through the block layer.” (P6)
• Lack of touch: “It doesn’t feel so certain that my finger is actually hitting the block at the exact spot that I am trying to hit it on.” (P13)

While it would have been helpful if there was a haptic feedback for a pointer, we believe that this haptic feedback was the secondary cue that pointers were looking for given the incorrect occlusion. For example, in the physical cube condition, we did not see any consistent behaviors of pointers touching the cubes. In addition, observers should have not been able to see the touch anyway because of the incorrect occlusion even if there were haptic feedback.

We asked if pointers had any strategies in the AR condition, and 7 out of 16 pointers answered that they did not change their pointing gestures in the AR condition compared to the physical condition.

• “I guess similar. Solid, I was touching it, but then this one I can’t touch, it’s almost like trying to touch it. Like, best I can.” (P2)

• “Over time, I think I’d followed a pretty consistent strategy.” (P3)

Meanwhile, nine pointers answered that they used unique referencing strategies to cope with the challenges they faced in the AR condition, where they pointed differently from the physical condition. Based on what we observed in our study, we found two pointing gestures that pointers in the study used commonly. Some pointers used both strategies.

[Strategy I: Pointing from the Top (6/16)] Based on our observations from the recordings, one strategy involved hovering the hand relatively higher, keeping the palm facing down, and positioning only the primary reference finger (nearly) vertically; that is, nearly perpendicular to the top surface of a cube. We found six pointers used this gesture, based on the interview and the video recordings, uniquely in the AR condition, and their average accuracy rate was 69.2%, which is better than the overall average (64.3%). They believed
3.3. Results

Figure 3.9: (Strategy I) pointing straight down from the top (Left), (Strategy II) penetrating into the virtual cube (Middle), (Other) grabbing a virtual cube (Right).

this strategy would make their intentions clearer by avoiding confusion when cubes occluded the referencing finger, or when other cubes occluded the target cube.

[Strategy II: Penetrating into the Virtual Cube (4/16)] This strategy involved intentionally positioning the endpoint of the referencing finger inside a virtual cube. Four pointers answered they used this gesture uniquely in the AR condition. Naturally, this strategy was impossible to use with physical cubes. Identification of this strategy was strictly based on the observer’s perception—that is, it was reported in interviews. We cannot truly evaluate whether a participant’s fingertip was inside a virtual cube (as our recordings do not have virtual cubes rendered and we did not track pointers’ hands). The average accuracy rate for participant pairs that used this strategy was 74.2%, yielding a better overall accuracy ratio than the average.

- “When we were in AR, I tried to have my finger partially intersecting the block...I figured that if I was having my finger on top of the block, I’m not exactly sure based on his viewpoints how that may or may not look, since a lot of these things are kind of transparently looking.” (P12)

One interesting strategy we found in the recorded videos was a pointer’s (P8) gesture of grabbing the cube (the rightmost picture in Figure 3.9) This strategy provides two visual cues, using the thumb and the index finger to simulate grabbing a physical cube. The pointer
only used this strategy to reference cubes located around the edges of the grid, not all the time. The accuracy rate for this pair (pair 8) was 100% for all trials.

Observers’ Strategies

In summary, most of the observers expressed difficulty in recognizing what the pointers were pointing at. Only two observers reported that the observing experience was not hard. Nine observers reported that incorrect occlusion (the perceived overlap between virtual cubes and pointers’ fingers) caused difficulty.

Eight observers believed there was a discrepancy between the virtual cubes they saw and the cubes their pointing partners saw, even though this was not the case due to the calibration step conducted before the AR condition.

- Incorrect occlusion: “I couldn’t tell [whether] she was pretty pointing to this and blocking this or touching this and pointing to this.” (O10)

- Incorrect occlusion: “It was kind of hard to tell because sometimes, like, your finger would go, like, through a block. So it’s kind of hard to tell, like, if he was pointing to this or this one.” (O9)

- Perceived discrepancy: “There is some kind of misalignment between the pointer and the observer. So if he’s pointing to the exact [top] of block, I think that he’s finally at the center of four cubes.” (O1)

- Perceived discrepancy: “I feel like my view was shifted a bit, because, like, what I saw was, like, in the middle.” (O12)

Overall, the observers were less confident about their answers in AR condition. In the AR condition, the average confidence levels across 16 pairs for the face-to-face and side-by-side
seating positions (5.70 with $\sigma = 1.18$ and 5.47 with $\sigma = 1.33$, respectively) were lower than those for the physical condition (6.93 with $\sigma = 0.32$ and 6.97 with $\sigma = 0.20$), as shown in Fig. 3.6 (right). The confidence value in the physical condition was close to the maximum value (7), illustrating that essentially all observers felt that their answers were correct without a doubt. We ran an ordinal regression model in the R package MASS ⁴ for ordinal dependent variables to analyze this effect and found that cube type had a significant effect on self-reported confidence ($p < 0.001$, $d = 1.55)$[98]. The effect of seating position was not significant. We utilized the ordinal logistic regression for confidence analysis because confidence level is categorical data from 1 to 7. And, it works to predict the dependent variable with ‘ordered’ multiple categories and independent variables.

When asked if they had employed any particular strategies, 14 observers said that they did so to overcome visual occlusion problems. Here, we present two themes that emerged in the thematic analysis of the interviews. The groups represented in these two themes were mutually exclusive.

[Learning the Gesture Pattern (10/16)] This strategy involved learning the pointer’s hand gesture pattern as it emerged during the study, and using that knowledge as supplementary cues to help guess and identify the cubes. This strategy is based on an assumption that the observer’s previous identification is roughly correct. Ten observers compared their partners’ referencing gestures with gestures made in previous trials that they identified. In addition, they all reported that their identification grew faster as they learned their partners’ patterns.

- “I did start to, like, learn, you know, especially when I realized, like, how his was working if he was [pointing] between them. I was like, okay, he probably is pointing at the previous one. So I did get faster.”(O5)

⁴https://cran.r-project.org/web/packages/MASS/MASS.pdf
Figure 3.10: Average accuracy rate, average completion time, and average confidence across trials in the AR condition (face-to-face, side-by-side). There was no apparent trend of improvement over repeated trials.

- “I think as we went along, I sort of figured out a way, for example, oh, if it’s really low, he is just pointing to the block of the first row.” (O3)

To verify their belief, we reviewed these participants’ performance (accuracy rate and completion time) and their confidence over trials. However, we could not find any evidence that the participants were learning over trials. As we did not give any feedback on whether observers’ answers were correct or not, they believed that they were getting better over time, when in fact, they were not, as demonstrated in Figure 3.10.

[Envisioning the Pointer’s Perspective (4/16)] This strategy involved envisioning the partner’s perspective and choosing a cube that was not necessarily the closest to the referencing finger physically, but the cube that an observer imagined the pointer was pointing at. One observer imagined her partner’s finger ended on the cubes that were one row to the right of the cubes they saw. Another observer used certain virtual cubes as anchors and mentally aligned the pointer’s finger based on direction and how far his finger was from the anchor cubes.

- “I was starting to think about his perspective... if you, like, if you point over here to the right a little bit, then I would like, like, shift everything to the right.” (O2)
3.4. Discussion

3.4.1 Research Questions

Assuming that the incorrect occlusion would negatively influence the collaborative performance of barehanded referencing in model-free AR, we wanted to know to what extent performance would be degraded (RQ1-A-i). The experiment results demonstrated that the average accuracy rate was reduced by 35.6% on average, and the task completion time was increased by 1.44 seconds (31.0% increase). Without verbal communication, in model-free AR, the visual cues become less reliable: targets are no longer correctly occluded by the pointer’s hand, and the pointer’s hand shadow no longer interacts with the targets. We speculated that observers would need to seek other information to complete the task, such as pointers’ gesture patterns. As such information is neither as salient nor as reliable as occlusion, observers must spend time interpreting them and cognitive load dealing with the uncertainty, resulting in lower accuracy and slower decision-making. While these cues resulted in higher accuracy than random guessing (1/16), this accuracy is still far below that of the physical condition, implying that the strategies that the pointers take are insufficient to support barehanded referencing in model-free AR, at least on the scale of our tests. Our findings indicate the potential to exploit various pointing gestures to better support barehanded referencing in AR. For example, based on the fact that pair 8 achieved 100% for

• “Use the virtual cubes as kind of, like, the anchor...Yeah, the same way that you use physical cubes, like, as the anchor, like it’s this far in front of four.” (O10)

In general, the observers seemed to expend more cognitive effort than would have been necessary were the targets occluded correctly.
all trials, a follow-up study could be carried to evaluate the grabbing gesture against other pointing gestures.

Additionally, we hypothesized that the spatial configuration (viz., the positions of the collaborators and the target cubes) within model-free AR might be relevant to the accuracy loss and analyzed its impact on the collaborators’ performance (RQ1-A-ii). Our analysis revealed that users performed significantly better in some spatial configurations than in others, even with incorrect occlusion. We reasoned that under different spatial configurations, the observers would encounter different occlusion scenarios, with some of them (e.g., virtual objects closer to the pointer) potentially being easier to resolve than the others. For instance, in the face-to-face condition, referencing cubes 5 and 9 should be easier than referencing cubes 8 and 12, because in the former case, only one cube is likely to occlude the pointer’s hand. Our analysis did not find that seating positions alone had any influence, but it did support the influence of cube positions. Because the change in perspectives caused by different seating positions did not seem to have an impact on referencing performance in our study, this result suggests that viewing angle does not significantly influence the effectiveness of strategies that the pointers take. However, we expect that in more complicated model-free AR scenarios, references to virtual targets on edges and corners, or to isolated targets, will be easier for observers to understand no matter where the collaborators are located.

We expected that pointers would adapt their pointing gestures in the model-free AR condition in order to mitigate the difficulties posed by the incorrect occlusion (RQ1-B-i). The interview provided valuable insight into the strategies they developed from their perspectives. In summary, 9 out of 16 participants reported changing their pointing gestures, specifically in the model-free AR condition. Their motivation was mostly related to a lack of sensation of touch on the target cubes. Due to the fact that the pointers’ own hands were incorrectly occluded, they had problems knowing if their hands were aligned properly with the target
virtual cubes. In their comments, the pointers indicated that since they could not touch the virtual cubes, they could not perform the same pointing gestures as they did in the physical condition. However, the lack of tangible feedback should not directly impact the observer viewing the pointing gesture, since the observers were not making judgements based on tangible sensations. It is worth noting that the pointers used two diametrically opposed mitigation strategies. Some tried to minimize virtual content wrongly occluding their hands by pointing at cubes from above, or even grabbing cubes. Others decided to penetrate the virtual cubes to help their collaborators. We should also note that because participants did not switch roles, the pointers may not have been fully able to understand what the observers were seeing. Since the pointers did not see the incorrect occlusion from the observers’ perspective, it could have been difficult for them to develop a strategy to refine their pointing gestures and thereby improve collaborative performance. Based on these findings, we plan to run a follow-up study to evaluate the two diametrically opposed mitigation strategies that will allow the collaborators to switch roles to establish a better understanding of the incorrect occlusion.

Finally, we expected to see changes from the observers’ side as well (RQ1-B-ii). In sum, the majority of the observers admitted seeking other information through learning. Since we did not tell observers whether their answers were correct, we believe that the observers tried to learn about the effectiveness of pointing gestures from easier trials, which they took as sources of ground truth. However, this learning effect was not present in our performance analysis, which contradicts the findings from Chastine and Zhu [23]. The mismatch between the observers’ responses and objective results could have been caused by a lack of feedback. Without knowing if their responses were correct, observers were not equipped to learn from the trials. In practice, resolving referential ambiguity constantly through correction may actually help observers improve their accuracy rates. This can be tested by repeating the
study and provide feedback to the observers.

## 3.4.2 Design Implications

Generally speaking, the results indicate that, with incorrect occlusion of the user’s hand, deixis will be severely restricted and collaborative tasks will be jeopardized. Our work also demonstrates the role of spatial configurations in collaborative referencing tasks and sheds light on alternative pointing gestures that can provide useful information when correct occlusion is missing. Here, we discuss a few design implications that can alleviate the reference problem.

As we observed during the study, the cubes in the column nearest the pointer in the face-to-face AR condition were significantly easier for the observers to identify. Designers should therefore be able to reduce referencing ambiguity caused by incorrect occlusion by changing the spatial layout of virtual content. In particular, designers can avoid placing virtual content along the collaborators’ view directions, rather dispersing the content in front of them. Another approach is to increase the distance between adjacent objects to reduce possible overlap between users’ deictic gestures and non-target objects. A similar effect can be achieved by dynamically changing the scale of the virtual content during referencing actions. The idea can be further strengthened with an automatic layout adjustment. We plan to explore a system that supports dynamic arrangement of the referenced targets based on the collaborators’ spatial arrangement.

Furthermore, based on the results from our experiment, one potential solution to provide more information to observers is to share views among collaborators, as pointed out by other researchers [23]. Although we observed that the pointers adopted various pointing gestures in response to the incorrect occlusion, their strategies were limited by their perspectives. If
the pointers were able to see the observers’ views, it could have helped the pointers develop new ways of referencing virtual targets in model-free AR that are effective for observers. One typical approach for sharing perspectives among collaborators is to stream the other users’ views in the shared virtual space.

The approaches above do not apply when virtual objects are continuous, such as terrain or buildings. For instance, sharing collaborator’s view alone will not help an observer pinpoint a specific location on the map. In these cases, designers use additional tracking technologies to support robust referencing communication. If the user’s hand can be tracked with six degrees of freedom (both position and orientation), but it is not practical to reconstruct an accurate hand model, one method designers can use is to define a virtual pointing ray emanating from the hand in 3D. This ray can then intersect with target locations. In this way, the observer needs only to look for an intersection between the ray and the target. We observed this kind of behavior in the user study, where pointers used an index finger to indicate a pointing direction. However, this strategy was ineffective, since observers found it hard to interpret the pointing direction without a virtual ray. Kim et al. [77] adopted a similar approach and found its benefit in referencing tasks.

### 3.4.3 Limitations

There are some limitations to our study that necessitate further work. First, the correctness of occlusion cues was not the only difference between the AR and physical conditions, as mentioned in 3.2.1. Other factors, mostly stemming from the HWD, might have influenced the results. The limited FoV in the AR condition constrained both collaborators by limiting the area in which they could see the virtual cubes. Another factor was the partial transparency of the virtual cubes when seen through the HWD. Several participants complained
of blurry vision when wearing the headset, potentially leading to less clear perceptions of pointers’ hands. Some participants also reported drifting of the virtual cubes when they turned their heads. However, we did recalibrate the virtual cubes between trials using the image target to minimize possible drift. These self-reports of drift could also have been caused by perceptual errors related to the incorrect occlusion cues, the translucency of the virtual objects, or minor device movements over time, causing the cubes to not appear fixed to the table. Therefore, it is possible that the performance decrease we observed was not entirely due to incorrect occlusion in the AR condition, though we believe it to be the most significant factor for barehanded referencing.

Moreover, there could be bias in the participants’ demographic backgrounds, since we only recruited university students from our campus. To verify if the findings of this research would apply to the broader population, we need a larger-scale study with a more diverse demographic. Moreover, the target layout we explored was kept simple and discrete in order to isolate spatial factors. This layout does not fully reflect the complexity of real-world tasks, where targets might be locations on or features of an object, rather than discrete objects. In a future study, we could observe users' spatial referencing behavior in more ecologically valid settings (e.g., a brainstorming application).

Another factor worth noting is the use of Optical See-Through (OST) HWDs. Prior research has identified that using OST HWDs can lead to overestimation of virtual target distances [104, 138, 148]. According to Swan et al. [138], when the AR targets are 50 cm away, the overestimation should be about 2 cm, which is at the same magnitude as the target cube in our study. However, we argue that the impact from OST display is minimal. If there were a major influence of distance mis-estimation on both the pointer and the observer, there would be a consistent pattern in Figure 3.8 as a result of such perceptual discrepancies. Especially for the face-to-face condition, a systematic overestimation would tend to result
in observers selecting cubes that are closer to themselves. The most reasonable explanation for our results is that incorrect occlusion is the dominant factor causing errors in our study.

3.5 Summary

Spatial referencing of virtual objects is important in many close-range, co-located, collaborative AR scenarios. Among various referencing methods, referencing with the user’s bare hand is a naturalistic and effective way of interaction in the real world. However, barehanded referencing will likely be influenced in model-free AR settings when correct occlusion cues are missing. In this chapter, we studied the effects of model-free AR on barehanded referencing. We found that participants’ performance was indeed reduced in model-free AR settings, and that the participants used various mitigation strategies to accomplish the task, though these were not effective. The experiment revealed that spatial configurations of the targets relative to the collaborators significantly influenced performance. In short, the chapter revealed the danger of improper support for spatial referencing and how it might negatively impact a simplistic collaborative task at close range.

Our research’s principal contributions include the analysis of major factors affecting collaboration performance in a model-free AR condition, the results and implications of our controlled empirical study, and design implications for collaborative AR systems involving barehanded spatial referencing.
Chapter 4

RQ2: Pointing Rays for Distant ObjectReferencing

4.1 Chapter Introduction

With a better understanding of the importance of spatialreferencing from Chapter 3, we will look into a more challenging scenario and propose two approaches to address the issues. Exchanging an object’s spatial information is more problematic at greater distances where visual cues of the collaborators’ body pose and the visibility of the target become harder to see. Moreover, in an outdoor co-located but distant scenario, acquiring an environment model becomes unreliable, leading to the use of model-free AR. This is at the heart of RQ2: How do pointing rays in model-free AR facilitate an accurate understanding of a spatial reference at a distance? To address this question, we have designed two visualization techniques and performed a controlled study in simulated AR to evaluate their effectiveness from the observer’s perspective. We found promising results in one of the designed techniques and continued to evaluate them in a more ecologically valid outdoor AR setting. This chapter will be broken into two parts to present these two studies.
4.2 Pointing Ray Techniques

In this section, we will describe the pointing ray techniques: the baseline Single Ray technique, the Double Ray technique, and the Parallel Bars technique. We also describe the pointing interface we designed to help the pointer to specify accurate ray(s) at the target of interest.

4.2.1 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 4.2 (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 4.2 (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.
4.2.2 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 the Double Ray technique 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 4.1 (b).

The obvious downside of the Double Ray technique is the added workload on the pointer. We discuss our pointer interface design in Section 4.2.4.

4.2.3 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) that are 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 4.3. In the simulated AR study, we use empirical parameters to fix the height, lengths, and intra-bar distance from the pilot experiment.

Comparatively, we add a couple of features to further enhance the technique in the outdoor AR study. First, we allow the user to configure the height and length of the Parallel Bars and their interval 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.
4.2. Pointing Ray Techniques

Figure 4.4: The pointer’s view using the pointing interface to specify: A single ray (a& b), and two rays (c). The user presses and holds the “B” button to sample 100 rays. The crosshairs in (c) are controlled using the two joysticks to align with the top and the bottom of a lamppost simultaneously.

direction, as shown in Figure 4.1 (a) and Figure 4.3. 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 4.3 (b). In this case, we extend the original ray with a different color to the closest point.

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.

4.2.4 Pointing Interface

The goal of the pointer interface design is to help the pointer to specify the target accurately and easily in the outdoor AR study. 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 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. [85], 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 same progressive refinement approach from prior work [83, 85] 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. This procedure is illustrated in Figure 4.4 (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 4.4 (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.
4.3 Visualization Techniques in Simulated AR

4.3.1 Ambiguities

In this study, we first introduce the concept of two types of ambiguities. Many collaborative VR applications use pointing rays and highlight objects intersected by these rays to facilitate visual consensus [6]. The pointing ray can be specified accurately by various input devices, such as hand trackers, head trackers, and eye trackers ([62, 85]). Head tracking is widely supported in mainstream VR and AR 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 section, 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. However, in model-free AR, the 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 4.2b.

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 [31] 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.

4.3.2 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 4.5 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 [17, 46, 88].

4.3.3 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
4.3. Visualization Techniques in Simulated AR

Figure 4.5: Example of an experiment trial of the participant’s first person view.

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 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.

### 4.3.4 Experiment Design

Our experiment followed a $2(\text{rays}) \times 2(\text{bars}) \times 3(\text{VA}) \times 2(\text{SA})$ 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 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. Visualization Techniques in Simulated AR

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: \textit{rays} (single or double) and \textit{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.

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.

\textbf{Visual Ambiguity (VA)} \quad \textit{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 4.1, top). To systematically study how VA affects the effectiveness of our gaze ray techniques, we defined three levels of VA as follows:}

- \textbf{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 4.6a. 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 4.6b. 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 4.6c.

Figure 4.6: Levels of visual and spatial ambiguity used in the experiment.

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 4.6d).

- **High SA**: One and only one object exists in the SA region (10-15 degrees away from the gaze ray; Figure 4.6e).

VA and SA were manipulated independently in the experiment, leading to a total of six 

\((2 \times 3)\) 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 \((4 \times 6)\) 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.

**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 thoughts about the four gaze ray techniques and their strategies for completing the task.

Hypotheses

We tested five hypotheses in the experiment:

*RQ2-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.

*RQ2-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.

*RQ2-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.

**RQ2-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.

**RQ2-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 [58]. 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 [58]. 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 under-
standing of the gaze ray’s direction through the parallel bars and therefore have a better
chance to correctly identify the target.

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.

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 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) mea-
surement 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.
Figure 4.7: Average error rate for different ray techniques under different VA and SA combinations. Error bars represent standard error.

4.3.5 Results

Figure 4.7 plots the average error rate by task condition for the four ray techniques. There are 576 data points \(24(p\text{articipants}) \times 2(SA) \times 3(VA) \times 2(r\text{ays}) \times 2(b\text{ars})\). 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 4.7.

Figure 4.8 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 RQ2-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 [111] to analyze the fixed effects of VA, rays, and the interaction between them. The model (Nagelkerke $R^2 = 0.503$) [107] 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 RQ2-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 4.9.

The analysis for RQ2-H2 was similar to the analysis for RQ2-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 ($\chi^2(2) = 42.855, p < 0.0001$). 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 4.10 shows the average error rates for the Single Ray and Double Ray techniques at the three different levels of VA conditions.

When testing RQ2-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

![Figure 4.10: 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.](image-url)
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 RQ2-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 RQ2-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 ($\chi^2(1) = 3.097, p = 0.078$).

### 4.3.6 Discussion

We hypothesized that the Double Ray would effectively reduce visual ambiguity, resulting in higher accuracy and greater speed than the single ray (RQ2-H1). Our results support RQ2-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 “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 (RQ2-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 4.10. 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 (RQ2-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 RQ2-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 RQ2-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 (RQ2-H5), we proposed that spatial orientation was an important individual characteristic that would affect users’ ability to make correct selections using Parallel Bars techniques. RQ2-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.

4.3.7 Summary of Simulated AR Pointing Ray Study

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.

4.4 Pointing Ray Techniques in Model-Free Outdoor AR

4.4.1 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. [67], 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. [99] 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.

\(^{1}\)https://www.ptc.com/en/products/vuforia
\(^{2}\)https://www.photonengine.com/
Moreover, instead of using a laser to align image targets, McGill et al. [99] 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 participants. The result is illustrated in Figure 4.1.

4.4.2 Experiment Design

We conducted a user study with 32 participants from our university to replicate and improve upon our prior study. Unlike the prior study that was conducted in simulated AR, our experiment used outdoor AR to evaluate the ray techniques. Lee et al. [89] 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 perfect pointing accuracy 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):

- **RQ2-A.** 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?

- **RQ2-B.** 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?

- **RQ2-C.** 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?

- **RQ2-D.** 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:

- **RQ2-H6.** 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.

- **RQ2-H7.** 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 participants interpret the orientation and lead to better accuracy.

- **RQ2-H8.** 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 Double Ray and the Parallel Bars will provide observers with the most information, increasing accuracy and confidence, at the cost of increased observing time.
• **RQ2-H9.** 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.

**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.

**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 [40]. 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. We started from a template created for the HL1, and modified the shape to fit the HL2 visor. We chose to use a tinted window film because of its no-adhesive, static-cling installation method. We attached the custom-shaped tinted film to the front of the HL2 visor and made two loops with Velcro on either side of the sunscreen that were slid through the gap on either side of the HL2 visor to ensure further secure attachment.

---

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

4.4. Pointing Ray Techniques in Model-Free Outdoor AR

Figure 4.11: Circular menu for the Observer. Left joystick is used to navigate through the options and the ‘A’ Button is used to confirm the selection.

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 [97], 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 4.4). Observers used the left joystick to navigate through the circular menu and the ‘A’ button to confirm selections (Figure 4.11).
Task

The study took place in a large outdoor environment with two participants positioned at a distance of 70 meters. Seven targets, the closest one being 29.7 meters away and the furthest one being 89.7 meters away, resulted in an average of 85° range of gazing direction for each participant. The distance between the closest and the furthest target was 84.64 meters. To enhance visibility in AR at such a distance, each participant’s head was represented by a semi-transparent blue sphere and a yellow cone in front of the sphere was synchronized with user’s facing direction (Figure 4.1 b).

We chose seven lampposts as the targets for the study. 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. [85] to mark these positions in a fixed coordinate system. We used the measured positions not only to measure pointing accuracy, 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. The markers also ensured that the pointers and observers perceived the same point as the target location. Otherwise, the pointer and observer might have a different understanding of where the top and bottom of the lampposts were, causing confusion in the study. 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
4.4. Pointing Ray Techniques in Model-Free Outdoor AR

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 developed eight (2 roles using 4 different techniques) randomized orderings of the 14 trials. 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)). Since we took a within-subject design, each participant needed to complete $2(\text{Number of Rays}) \times 2(\text{Orientation Cues}) \times 2(\text{roles}) \times 7(\text{targets}) \times 2(\text{repetitions}) = 112$ trials.

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. Upon their arrival, we welcomed both participants and asked them to read and sign an informed consent form (approved by the Institutional Review Board of the university).

Introduction We gave the participants an introductory presentation on our experiment background and the experiment task. Following a short Q&A session, we helped the participants put on the HL2 headset and ran the calibration program to calibrate the headset for
optimal viewing experience.

**Formal Study**  After the participants reported they were ready, the experimenters accompanied the participants and escorted them to different designated locations. We asked the participants to stand on a pre-defined marker so that they were 70 meters away from each other. We provided a chair for participants and encouraged them to take rest whenever needed. From these positions, the participants learned, practiced, and completed trials with each of the four techniques.

For each technique, there were five phases: overview, training with simulated collaborator (second and third), and experimental trials with real collaborator (fourth and fifth).

In the **first** phase, we presented an overview of the technique by showing a picture with a top-down view of the current scenario and explained the technique. The **second** and **third** phases had training programs where participants were trained as both the pointer and the observer independently with a simulated collaborator. The simulated collaborator had the same appearance as an actual collaborator and its movement mirrored the participant’s head movement at the location of their actual collaborator. In each of these phases, the participants had to successfully complete at least 14 trials to familiarize themselves with the technique. Success or failure for each trial was conveyed to the participant through auditory feedback.

For pointer training phase, the participant 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. In this way, the participant could see the difference between these rays 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 for the training sessions and not available
4.4. Pointing Ray Techniques in Model-Free Outdoor AR
during the actual study sessions.

For the observer training phase, the simulated collaborator would play the role of the pointer. The virtual pointer’s head was fixed once the pointing ray(s) came out of the yellow cone. In addition to identifying the target pointed at by the virtual pointer, the participant also learned how to use the circular menu to confirm their selection. The circular menu had the numbers one through seven in clockwise sequence representing the seven targets. The circular design was chosen as all the numbers were equally accessible from the center and reduced bias to any particular target.

During the observer training phase involving the Parallel Bars techniques, participants were able to manipulate the length, intra-bar distance, and height of the parallel bars using the controller. All the parameters were measured in meters, and displayed to the observer on their screen. The participants 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 (fourth phase), where one participant was assigned the role of pointer while the other acted as the observer. The participants collaborated with each other to complete a set of 14 trials using the technique they familiarized in the prior phase. 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 this phase, 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 (details in the next section).

In the fifth phase, the participants switched their roles and completed another set of 14
CHAPTER 4. RQ2: POINTING RAYS FOR DISTANT OBJECT REFERENCING

trials using the same technique. After completing these trials, we again collected the TLX and SUS questionnaires.

We repeated these five phases for each of the four techniques. 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.

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.

We also recorded the accuracy for both pointer and observer. The pointing error was defined as the angular distance of the ray from 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 participants successfully identified the target. We did not give the participants any feedback on accuracy, to avoid learning effects.

Finally, during the experiment, we gathered subjective feedback through Raw NASA TLX [56] and a SUS questionnaire modified from the original questionnaire [19]. The participants completed the NASA TLX questionnaire immediately after completing a set of trials as the pointer for a pointing technique, as described in previous section. We collected only two sets of TLX data from each participant (since the pointing task does not change when Parallel Bars are added). We gave participants the SUS questionnaire immediately after completing a session as an observer. 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 participants 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.

4.4.3 Study Results

The study ended with 1792 data points \((32(\text{participants}) \times 2(\text{repetitions}) \times 2(\text{Number of Rays}) \times 2(\text{Orientation Cue}) \times 7(\text{targets}))\). 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.

RQ2-A. Effects of Double Ray Technique on Observer Performance and Subjective Experience

Figure 4.12 presents the observing time by pointing techniques. We started with a linear mixed model (LMM) using Welch-Satterthwaite t-Test [150] with participant pairs as random effects to test Number of Rays, Orientation Cue, and their interaction. We found the interaction significant \((t(1632) = -2.01, p = 0.044)\). 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, the model revealed a significant effect of the Double Ray technique on the observing time \((t(824) = -3.1, p = 0.002)\). When applying the same analysis when the Orientation Cue was present, another LMM found no significant difference between the two techniques \((t(794) = -0.06, p = 0.95)\). 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 4.13 plots the average accuracy that the observers achieved using different
pointing techniques. While the average observer accuracy was calculated across different targets and participants, 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 [69] to test Number of Rays, Orientation Cue, and their interaction. Since we found no significant effect from their interaction ($z = -1.31$, $p = 0.19$), we could use separate GLMMs to analyze the effects of Number of Rays and Orientation Cue independently. We found a main effect of Number of Rays on observer accuracy ($z = -5.857$, $p < 0.0001$). 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
Figure 4.13: Average accuracy the observers achieved using different pointing ray techniques. Accuracy was calculated across targets and participants.

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 ($t(94.0) = -2.19, p = 0.031$), higher confidence ($t(31.0) = -4.05, p = 0.0003$ with Orientation Cue absent, $t(31.0) = -2.26, p = 0.031$ with Orientation Cue present), and less complexity when the Orientation Cue was present ($t(31.0) = -2.27, p = 0.03$) than the Single Ray techniques. We found no significant effect of the Number of Rays on the overall modified SUS score ($t(94.0) = -1.45, p = 0.149$). 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.

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 participants preferred the Double Ray over the Single Ray technique. For the techniques with Orientation Cue present, 91% of the participants preferred Double Ray over Single Ray. The participants 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 with Parallel Bars, 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 participants’ prioritization of accuracy over speed, especially when they consider working with a collaborator.

RQ2-B. 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 performed a series of similar analyses as the last section. 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 ($t(792.4) = -6.059, p < 0.0001$). With the Double Ray, there was also a significant effect of Orientation Cue ($t(825.6) = -3.286, p = 0.001$). 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 ($z = -1.156, p = 0.248$), 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 \( t(94.0) = 3.456, p = 0.0008 \). Moreover, a closer look into the sub-scores revealed that the Parallel Bars techniques were harder to learn \( t(94.0) = 3.586, p = 0.0005 \), harder to understand \( t(94.0) = 3.124, p = 0.002 \), and required more learning effort \( t(94.0) = 3.606, p = 0.0005 \) and more mental effort \( t(94.0) = 2.222, p = 0.0287 \). 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 participants 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 participants preferred to have Single Ray with Parallel Bars over just Single Ray. For Double Ray, however, 53% of the participants found the Parallel Bars redundant and time consuming. As a result, only 59% of the participants preferred to have Double Ray with Parallel Bars over Double Ray.

### 4.4.4 RQ2-C. Pointer Performance and Subjective Experience with Double Ray Technique

Figure 4.14 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 pointing interaction occurred). An LMM indicated that Number of Rays had statistically significant effect on pointing time \( t(1682.1) = -33.83, p < 0.0001 \). The increased pointing time using the Double Ray technique is expected, as the pointer must align two crosshairs with the controller.

Figure 4.15 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 ($t(1718.3) = -7.963, p < 0.0001$). However, the relative difference in error was not large between the two techniques, with an estimate of $0.043^\circ$.

Regarding the NASA TLX score, the Double Ray technique was found to have higher workload ($t(31.0) = -2.066, p = 0.0473$). A closer look into the sub-scores only revealed trending significance of the effects from Double Ray technique on mental demand ($t(31.0) = -1.986, p = 0.056$) and physical demand ($t(31.0) = -1.962, p = 0.059$).

When asked to rank the two techniques from a Pointer perspective, there was a mixed response. 53% of the participants chose the Double Ray as their preferred technique. Overall, then, while pointing with the Double Ray is demonstrably slower, more error-prone, and more
CHAPTER 4. RQ2: POINTING RAYS FOR DISTANT OBJECT REFERENCING

Figure 4.15: Pointing error by the pointers using Double Ray and Single Ray techniques. Errors are measured in degrees.

demanding, these drawbacks were not seen as fatal flaws by the participants. We discuss this further in the next section.

4.4.5 RQ2-D. Trade-off between Pointer and Observer

The Double Ray technique benefits observer accuracy 4.4.3 but reduces some aspects of pointer experience 4.4.4. 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 ($z = 1.55$, $p = 0.121$). However, pointing error was found to be a statistically significant covariate to the observer accuracy ($z = -3.983$, $p < 0.0001$), as shown in Figure 4.16. These results suggest that the more erroneous the pointer is, the less likely the observer can make a correct target identification, whereas spending more pointing time does not help the observer become more accurate.

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 4.4.4, 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 less-accurate pointing ray. 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 the previous section, 53% of the participants 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.

### 4.4.6 Discussion

We hypothesized that the Double Ray technique would significantly improve some aspects of the observer experience compared with the Single Ray technique (RQ2-H6). We found good evidence to support this hypothesis since the Double Ray technique was easier to learn, led to better observer accuracy, user confidence, and was favored by most participants 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 RQ2-H6 based on the findings of our previous experiment 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. We speculate that this was because of the human pointer. Unlike the prior work, which used a simulated pointer that was capable of casting perfectly accurate rays towards the target object, the pointers in our 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 study, the Parallel Bars enhancement did not contribute to a higher observer accuracy, even though it increased observing time. They 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 participants to adjust the Parallel Bars parameters to avoid visual clutter. We hypothesized RQ2-H7 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 participants 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 RQ2-H8 as a natural extension to RQ2-H6 and RQ2-H7, 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 participants. The participants’ 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 (RQ2-H9). RQ2-H9 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 participants chose the Double Ray technique, despite the decreased observer performance. We are inclined to believe that when asked about the experience, the participants also considered the observer benefit.

We wanted to know understand the trade-off that the pointers have to make to help observers (RQ2-D), 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 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. On the flip side of the discussion, 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.

4.4.7 Summary of Outdoor AR Pointing Ray Study

Working under model-free AR, the naïve usage of virtual pointing rays can lead to ambiguity and confusion. In this work, 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 from our prior simulated AR study, but also added new understanding because of our use of true outdoor AR and a two-user collaborative task. 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.
4.5 Summary

This chapter proposes two pointing ray techniques for distant object referencing that do not rely on the environment model. Through one simulated AR study and an outdoor study built upon the results from the previous one, we have demonstrated the potential of enhancing visual match conditions (the Double Ray technique) that can significantly benefit the observer’s performance. Our results indicated that even though the Double Ray technique causes performance loss on the pointer’s side, the users are willing to sacrifice in the hope of the greater good. Comparatively, the Parallel Bars technique is proven unsuccessful in both studies. We have to explore other 3D UI designs to efficiently provide meaningful spatial information to the user. In short, this chapter indicates that through 3D UI design, we can overcome the problem introduced by a lack of environment information in collaborative spatial referencing.
Chapter 5

RQ3: Point Marking Techniques

5.1 Chapter Introduction

From the first two RQs, we have seen evidence of negative impact on spatial referencing in model-free AR and presented visualization techniques that could attenuate the negative impact. RQ3 aims to address the problem directly from its origin: since the system doesn’t have information about the referencing point, can we develop 3D UIs that can help the user to specify (or mark) the 3D point location directly? From prior work [20], we observed the potential in geometric triangulation marking techniques as they do not rely on human’s depth perception, as illustrated in Figure 5.1. Unlike alternative approaches that rely on one ray and let the user specify a distance along that direction using a “fishing reel” technique with a mouse wheel or similar instrument [20, 130], this geometric approach does not rely on human depth perception. Yet, the naïve implementation may be severely limited by human motor control issues (such as head or hand tremor) and/or technical issues (such as tracking jitter or drift). We designed multiple enhanced techniques and conducted controlled user studies both in simulated AR and outdoor AR to evaluate the techniques. We found that it is possible to reliably specify an arbitrary point’s location up to 85 meters. The work is a continuation of Wallace S. Lages’s doctoral dissertation: “Walk-Centric User Interfaces for Mixed Reality” [125]. Notably, the marking techniques are methods for the pointer to create annotations at arbitrary 3D positions in wide-area collaboration, unlike those in Chapter 4,
which focused on both the pointer and observer. We briefly discuss the use of the point markings by an observer in the conclusion of this chapter.

5.2 Interaction Methods

5.2.1 VectorCloud

We implemented a naïve version of the geometric triangulation approach (Figure 5.1) using the Microsoft HoloLens. To specify a 3D ray, the user rotates her head to align the center of a virtual crosshair with the point of interest in the real world. The crosshair is centered in the user view, as shown in Figure 5.2. Thus, 3D rays are defined by the user’s head position and orientation, as tracked by the HoloLens SLAM system. Users press a button on a handheld controller to fix the first ray, then walk to a new location and repeat the process. The location of the target 3D point is calculated as the midpoint between the closest points on each ray, since two 3D rays may not intersect.
We soon observed that marking was fairly accurate (the average of many attempts to mark the same point was near the target), but imprecise (there was high variability in the position of marked points over multiple attempts)\(^1\). This issue became worse at larger distances. We suspect that as distances increase, small angular errors in head orientation due to head tremor and noise in the tracking data can introduce substantial errors in ray direction. Although it is easy to fixate the eyes on a distant target due to the vestibulo-ocular reflex \([94]\), the same is not true of the head. Without the option to use eye tracking, our crosshair was attached to the head, and jitter of the crosshair relative to distant targets was obvious.

The accuracy of casting an individual ray depends on the stability of the user’s head and their ability to press the button at a time when the crosshair is very close to the target. We thus set out to reduce the influence of human and system jitter by casting many rays in a

\(^1\)We use the term *accuracy* to refer to the error, relative to ground truth, of an individual marked point or the average error over many attempts, and the term *precision* to refer to the variability in point locations over many attempts.
multi-sampling approach. Assuming that the pointing error is random, consecutive attempts to mark the same target will lead to a cloud of points surrounding the desired target position. Therefore, computing a location estimate of the entire cloud should lead to a better estimate of the target point. We call this technique VectorCloud. The insight behind the VectorCloud technique is to assume that rays gathered from two different positions are independent, and so can be intersected in any order. Instead of computing the intersection of every pair of rays in sequence, we first store several samples of the target direction from each position. Then, we compute all possible intersections between the two sets of stored rays (Figure 5.3). We use the midpoint between the closest points on the two rays as the intersection.

By computing all intersections, the technique allows the user to mark from any number of arbitrary positions in space or even continuously sample as they walk. However, special care is needed in the case of continuous sampling and discrete sampling from more than two locations. When sampling from two discrete locations, ray intersections between pairs of rays from each location will always generate useful information, since every intersection will be valid. However, when using more than two discrete locations or continuous sampling it is necessary to choose which subsets to intersect. We implemented a simple solution that works...
in both cases: we subtract all pairs of consecutive vectors and split the samples into two sets at the point of maximal absolute difference. This solution is simple and works even when the user completely circles a target while sampling. However, it is not optimal, as it may intersect groups of rays with a short baseline. To prevent this, more sophisticated solutions could be used, like clustering the rays by angular difference. We leave this for future work.

Finally, once all intersections are computed, we can use any location measure to estimate the target position. The exact 3D shape of the distribution depends on the sampling error, the sampling positions, and the distance between them. We have explored different methods to estimate the location of the target, including using the mean, median, and mode; and performing asymmetric trimming before computing the mean. We settled with the mean in our implementation, since it is well defined in every case.

From the user’s perspective, VectorCloud follows the same steps as the naïve geometric technique, except that the user needs to hold down the controller button for a few seconds at each location to gather multiple ray samples. Since we did not impose a limit on how many samples the user can gather, we assigned a second button to finalize the overall marking process.

### 5.2.2 ImageRefinement

From prior work comparing VectorCloud against naïve geometric, we found that the key determinant of accuracy in geometric marking is the precision of specifying the ray directions for triangulation. The precision of ray specification in the naïve geometric technique suffers because of head and tracker jitter. While VectorCloud reduces the impact of head tremor, it does not eliminate it completely. Another way to improve geometric point marking is to decouple the specification of the rays’ directions from the position and orientation of the
user’s head, so that each ray is more accurate. This led us to the creation of a technique based on progressive refinement [83].

In the ImageRefinement technique, the user first aims roughly at the target using the head and presses the controller button. Instead of generating a ray, the technique captures an image using the AR headset’s onboard camera. We crop and enlarge the image so that only the central region around the targeting reticle is shown, resulting in an image with the appearance of a 2.5x zoom. We display the image in the headset at a comfortable distance and display a crosshair at the center of the image (Figure 5.4). Next, in a refinement stage, the user adjusts the direction of the ray by manipulating the crosshair in 2D image coordinates to specify the location of the target in the image. Crosshair manipulation uses an analog joystick on a handheld input device.
5.3 Simulated AR Study

The origin of the ray is the position of the user’s head at the time the image was taken. The direction of the ray is computed by considering the forward vector of the camera at the time the image was captured, the horizontal and vertical offset of the cursor in pixels from the center of the image and the camera’s horizontal and vertical fields of view. In our implementation, we define a vector in camera coordinates from the camera to the offset cursor. Transforming this vector into world coordinates results in the refined ray. The user then repeats the process from a second location, and the intersection between the two rays is calculated as in the naïve geometric technique. We call this technique ImageRefinement (IR).

IR results in two high-quality rays, and gives users control over the final direction of each ray. Because of the refinement step, it does not require precise aiming using the head. It also allows users to rest their head and neck muscles while refining the ray, since the system recalls the user’s head position when they took the snapshot. Unlike VectorCloud, users of IR do not have to keep their heads in the same position for an extended period of time.

5.3 Simulated AR Study

Having established that it is possible to increase the precision of geometric marking (Section 4), we also wanted to compare our two enhanced techniques to explore the trade-offs between them. We hypothesized that IR would result in higher precision than VectorCloud due to its ability to make fine-grained adjustments to ray direction and its lack of reliance on head stability. However, we questioned how large this effect would be, and whether IR would result in a higher “hit rate” when we defined an accuracy threshold around targets.

The use of the VectorCloud technique is simple. Similar to the naïve geometric technique, it involves only looking at the target and pressing buttons to define the ray. IR, on the other
hand, is more complex, as it requires an additional refinement step for each ray. Based on these observations, we hypothesized that VectorCloud would be faster than IR, and that some users might prefer its simplicity. However, the refinement step also gives IR users more direct control over the result of marking (potentially leading to higher confidence), and it does not require users to keep their heads still during use (potentially leading to greater comfort). Thus, we hypothesized that most users would prefer IR overall.

Additionally, we wanted to understand how the two main features of IR—decoupling head position from ray direction and providing a zoomed image—affected performance individually. We therefore created an ImageRefinement “no zoom” variant (IRNZ) in which the image shown to the user was not digitally zoomed. The three techniques (VectorCloud, IR, IRNZ) were then evaluated in two controlled experiments.

5.3.1 Experiment Setup

In VR, we rendered the virtual crosshair and other simulated AR graphics with a customized shader to prevent occlusion by other geometry and with a transparency value close to their actual appearance in AR. We built a testing virtual environment containing six target lampposts at distances of 26.7m, 55.2m, and 85.5m (targets 2, 4, and 6 from previous study) since we had already learned how distance affects performance on the marking task. These distances were chosen to replicate a real-world scene we used for the AR evaluation. Since all techniques require the user to sample one or more rays from two separate locations, we marked two spots on the ground that were 1.92m from each other as the marking locations. The length of the baseline was not varied during the experiment because its effect on the techniques’ performance is equivalent to that of target distance: using a longer baseline is equivalent to marking a closer target and vice versa. The VR scene is shown in Figure 5.5.
Due to limited resolution and dynamic range of the VR display, we placed a small red sphere at the tip of each lamppost to indicate the exact point that should be marked.

Our experiment was a $3 \times 3$ within-subjects repeated-measures experiment. The dependent variables were marking error (absolute distance from marked point to target), hit rate (percentage of marking attempts with an error less than 4.25 meters, which is 5% of the farthest target’s distance), time to mark the target, and usability as measured by a modified SUS questionnaire.
5.3.2 Apparatus

We used a consumer version HTC Vive HMD with Lighthouse tracking and a wireless Xbox controller for input. IR and IRNZ were both controlled by pressing the ‘RB’ button to take a picture of the scene. The image was captured by a virtual camera positioned halfway between the left- and right-eye cameras used to render the scene. We rendered the image so that when it was displayed in the HMD, detail in the image was maintained; it was always a $640 \times 360$ image regardless of zoom level. Users could then use the left joystick to control the cursor. VectorCloud was controlled by holding down the RB button until the system indicated that the user had collected 300 samples (by playing a sound) and stopped collecting more samples. The number of samples was chosen to match the earlier VectorCloud experiment. Taking 300 samples took 2.5 seconds. After the second set of samples was gathered, the user pressed the ‘A’ button to confirm the input and complete the marking. The software was developed in Unity3D.

5.3.3 Participants and Procedure

This experiment gathered data from 24 participants (five female) with a mean age of 24.5 (standard deviation of 4.8). All of them had prior experience with AR or VR, with eight of the participants utilizing these technologies regularly. Six participants used contact lenses, nine used glasses, and the remaining nine had good uncorrected vision. The experiment was approved by the university’s Institutional Review Board.

Participants first read and signed an informed consent describing the experiment and possible side-effects. After that, the participants filled out a background questionnaire on their previous experiences with AR/VR as well as demographic data. Next, we gathered participants’ physiological data by measuring their interpupillary distance (IPD) and determining
their dominant eye, using the same tools as in section 4.2.

We instructed participants that they should emphasize accuracy over speed. We then presented the first technique (order of technique presentation was counter-balanced using a Latin square). A training procedure guided the participants through marking three targets. Once they confirmed that they were comfortable with the technique’s use, we started the formal trials. Participants marked the virtual lampposts in a pseudorandom order that included marking targets 2, 4, and 6 six times each for a total of eighteen trials. If a marking was more than 4.25 meters away from the actual position of the target, the system indicated a “miss” by playing a sound. Otherwise, a positive feedback tone was played.

After each technique, participants completed a modified system usability scale (SUS) questionnaire [19] about the technique they just finished. We replaced question 5 on the original SUS (“I found the various functions in this system were well integrated”) with “I was physically comfortable using this technique” since the original question did not apply to an individual technique. We also asked participants what strategies they used and why, and what comments they had about the technique, if any.

After completing the same steps for the second technique, participants filled out a final questionnaire asking them to specify which technique they preferred, and to list any issues they had while using the techniques.

**5.3.4 Results**

Figure 5.6 shows the marking error for each target and technique combination. We used a linear mixed model to analyze the effects of the independent variables, including fixed effects for technique and target and random effects for subject. The subject variance in the fitted model was small (0.2285) but indicated some benefit of including subject as a random effect.
We first analyzed the data using a model including an interaction term and found significant effects of technique and target, but no interaction effect. When we removed the interaction term from the model, we again found significant main effects of technique ($F(2, 34.5) = 35.85, p = 3.8e^{-09}$) and target ($F(2, 48.3) = 43.37, p = 1.7e^{-11}$). Furthermore, with post-hoc pairwise analysis we found significance between all pairs of techniques and all pairs of targets ($p < .001$). All t-test statistics were adjusted using Satterthwaite’s method to estimate the effective degrees of freedom. The differences between the techniques are visible in Figure 5.6, with IR achieving the highest accuracy, followed by IRNZ, and then VectorCloud. The hit rate data (Figure 5.7) also reflects this, with IR achieving a 98% hit rate at 85.5 m.

Timing measurements were only considered for attempts that resulted in a ‘hit’. This ensured we were only measuring attempts where the participants were working to be as accurate as possible. The mean times for hits with each technique and target combination can be seen in...
5.3. Simulated AR Study

Figure 5.7: Hit rate in the AR simulation experiment.

<table>
<thead>
<tr>
<th>Technique</th>
<th>Target 2</th>
<th>Target 4</th>
<th>Target 6</th>
<th>Overall</th>
</tr>
</thead>
<tbody>
<tr>
<td>IR</td>
<td>13.30</td>
<td>13.46</td>
<td>14.24</td>
<td>13.66</td>
</tr>
<tr>
<td>IRNZ</td>
<td>13.82</td>
<td>14.5</td>
<td>16.72</td>
<td>15.02</td>
</tr>
<tr>
<td>VC</td>
<td>16.31</td>
<td>17.13</td>
<td>17.80</td>
<td>17.08</td>
</tr>
<tr>
<td>Overall</td>
<td>14.48</td>
<td>15.03</td>
<td>16.26</td>
<td></td>
</tr>
</tbody>
</table>

Table 5.1: The mean time for marking a target in VR for each target & technique, as well as overall means across targets or techniques.

Table 5.1, along with overall means for each technique. After performing a similar analysis as for marking error, we found a significant main effect of both technique \((F(2, 23.2) = 27.9, p = 6.7e^{-07})\) and target \((F(2, 23.2) = 11.5, p = 3.5e^{-04})\) on the time to mark a target. Post-hoc analysis with Satterthwaite’s method revealed significant differences between all pairs of techniques \((p < .05)\).

We also performed an ANOVA on the effects of technique on total SUS score. We found that the techniques were significantly different from each other in terms of SUS score \((F(2, 69) = 12.37, p = 2.56e^{-05})\). Furthermore, post-hoc comparisons with Tukey’s adjustments revealed that all pairs of techniques were significantly different. The mean score for IR was 89.7,
while the means for IRNZ and VC were 74.8 and 63.2 respectively.

5.3.5 Discussion

Our findings show that ImageRefinement results in significantly higher accuracy than VectorCloud. This can be seen even when the benefit of zooming is removed from the technique. However, including zoom does have an effect on the technique, as IR is significantly more accurate than IRNZ, and the difference between the two is larger than the difference between IRNZ and VectorCloud.

The timing data show that despite its two-step process, ImageRefinement can still be faster than VectorCloud. However, it should be noted that we tuned these techniques for marking accuracy; VectorCloud could potentially be faster if we used a lower number of samples.

IR also appealed more to users. Users gave it a higher score than VectorCloud by an average of 26 points on the 100-point SUS scale, with 23 of the 24 participants indicating that IR was their preferred technique. The last participant indicated they preferred VectorCloud, but their SUS data gave IR a score of 97.5 and VectorCloud a score of 90. Users commented that they liked how IR allowed “the freedom to fail,” since they could refine their rays using the crosshair on the image. The qualitative feedback for VectorCloud indicated that nine users tried to “hold their breath” while using the technique. Furthermore, two expressed frustration with the technique since it did not give them feedback while using it, and two stated that the technique caused neck pain.

From this experiment we conclude that, in an idealized setting, the ImageRefinement approach was better than VectorCloud for marking tasks requiring high levels of accuracy, and that IR also improved the overall user experience, producing a feeling of greater control. The zoom feature in this study was also beneficial, allowing rays to be even more accurate.
5.4 Outdoor AR Study

Since the AR simulation in the study described in Section 5.3 was less than realistic in some ways (e.g., high-resolution camera images, opaque display, highly accurate tracking), we also wanted to explore how our enhanced marking techniques work in a real-world AR setting. We therefore ported the VR application code to work on the HoloLens. While VectorCloud ported over easily, we had to adjust the method of calculating the rays for IR and IRNZ, since we now were working with a real camera. We used the HoloLens API to transform pixel coordinates from the camera image into a ray based on the factory-measured camera intrinsic parameters and view transformation. In all other respects, we designed this experiment to be as similar as possible to the AR simulation experiment with a 3 (technique) × 3 (distance) within-subjects repeated-measures experiment design so that we could compare the differences between the two.

5.4.1 Apparatus

We used the Microsoft HoloLens along with a bluetooth Microsoft Xbox One controller. The HoloLens has a 2.4-megapixel camera, meaning that zoomed images are quite blurry. To maintain consistency with the previous VR evaluation, we cropped the images from the HoloLens’ camera to simulate zooming. However, since the camera’s resolution is only 2.4 megapixels, this results in a fairly blurry “zoomed” image. In the VR implementation, we generated a 640 × 480 pixel image for the user to refine their target. In AR, however, we had to crop the camera’s output to a 400 × 225 pixel image for IRNZ, while IR cropped the output to a 208 × 117 pixel image.

Due to issues with the version of Unity we were using at the time, in combination with the HoloLens and the controller, we had to change the control of the cursor in the ImageRefine-
ment techniques to use the ‘A’, ‘B’, ‘X’, and ‘Y’ buttons as directional inputs (for ‘down’, ‘right’, ‘left’, and ‘up’ respectively). Capturing an image with ImageRefinement techniques took about 0.5 seconds, as opposed to the instantaneous image capture in the AR simulation. Otherwise, the control scheme was the same as the VR experiment. VectorCloud was implemented in the same way as in section 4, except that we required 200 samples instead of 300, since it took the same amount of time to collect 200 samples on the HoloLens as it took to collect 300 samples in the VR experiment.

To improve the contrast of the HoloLens display outdoors, we built a “sunglasses” adapter which allowed the crosshair and the experiment instructions to be visible in daylight. The adapter was attached to the exterior of the device and was built using duct tape, Lego pieces, and shaded film.

In addition, we used an Apple iPad to communicate with the HoloLens during the experiment. This gave the researcher feedback from the system on how the participant was progressing in the experiment through a text console. The iPad also gave the experimenter a way to reset the HoloLens spatial anchor (a known position that is used by the HoloLens to help maintain a fixed coordinate system) and correct any large drift caused by using the HoloLens outdoors.

5.4.2 Participants and Procedure

We gathered data from eighteen participants (four female) with a mean age of 21.5 (standard deviation of 2.18). All of them had a background in engineering and only one of the eighteen had no experience with AR/VR displays. Five participants used contact lenses, seven used glasses, and six had perfect vision. The experiment was approved by the university’s Institutional Review Board.
Participants first read and signed an informed consent describing the experiment and potential side-effects. After that, the participants filled out a background questionnaire on their previous experiences with AR/VR as well as demographic data. Once these were completed, the participants completed the Microsoft setup program for the HoloLens that calibrates the system for each individual user and gets them accustomed to the display.

Participants were taken to the testing area, which was in a nearby courtyard (Figure 5.4). We defined two positions on the ground with a baseline of 1.92m for the user to stand on while marking the target. The lamp posts found in this area were used as targets, and were at the same distances as in the AR simulation experiment (26.7, 55.2, and 85.5 meters). The participants were instructed to emphasize accuracy over speed for all trials. We then introduced the first technique (order of techniques was counterbalanced using a Latin square) and guided them through a training period where they marked three targets. Once the participants felt comfortable with the technique, we began the formal trials. As in the AR simulation experiment, participants had to mark three targets six times each, and they were given auditory feedback by the system when they hit or missed the target. For this experiment the range considered to be a ‘hit’ was extended to be 10% of the maximum target distance (8.5 meters), because pilot testing revealed that there would be excessive misses with the 5% threshold used in the prior experiment. Between every other trial, participants were asked to gaze at a specific target while standing at a specific position, so that the experimenter could reset the virtual anchor. This was done to actively correct any HoloLens tracking drift and reset the coordinate system so that we could minimize errors due to tracking.

After each technique, participants filled out the same questionnaire as in the previous experiment. After all three techniques had been completed, they completed the same post-experiment questionnaire as in the prior study.
5.4.3 Results

Figure 5.8 shows the marking error for each technique target pair. Again, we used a linear mixed-model to analyze the effects of the independent variables. With the interaction term included in the model, we found only a significant main effect of target ($F(2, 137.5) = 349.4, p = 2.2e^{-16}$). As before, the interaction term between technique and target was not significant, so we removed it from the model for further analysis. Without the interaction term, we found significant main effects of both technique ($F(2, 28.4) = 6.35, p = .0052$) and target ($F(2, 932.3) = 353.8, p = 2.2e^{-16}$). A post-hoc pairwise analysis found that IR was significantly more accurate than VectorCloud ($T(24.78) = 4.199, p = 2.93e^{-05}$), and that IRNZ was significantly more accurate than VectorCloud ($T(24.78) = 2.5, p = .0126$). However, IR was not significantly different from IRNZ ($p = .0893$). Hit rates for each
target-technique combination can be seen in Figure 5.9.

As in the prior experiment, timing measurements were only considered for attempts that resulted in a ‘hit.’ Mean times for a hit for each technique and target combination can be seen in Table 5.2, along with overall means for each technique and target separately. A mixed model using random slopes for target and technique indicated there was a significant main effect of both technique \( (F(2, 17.5) = 14.2, p = .00022) \) and target \( (F(2, 33.07) = 80.94, p = 11.8e^{-13}) \) on marking time. Post-hoc analysis with Satterthwaite’s method revealed that IR was significantly slower than VectorCloud \( (T(17) = 3.57, p = .00235) \) and that IRNZ was also significantly slower than VectorCloud \( (T(16.9) = 4.6, p = .000248) \). As in the AR simulation experiment, it should be noted that marking time for VectorCloud is linked to the number of samples it needs from each position.

On the SUS questionnaire, IR scored a mean of 84.7 with a standard deviation of 14.4, IRNZ scored a mean of 73.5 with a standard deviation of 17.3, and VectorCloud scored a mean of
### Table 5.2: The mean time for marking a target in AR for each target & technique, as well as overall means across targets or techniques.

<table>
<thead>
<tr>
<th></th>
<th>Target 2</th>
<th>Target 4</th>
<th>Target 6</th>
<th>Overall</th>
</tr>
</thead>
<tbody>
<tr>
<td>IR</td>
<td>19.77</td>
<td>23.65</td>
<td>25.46</td>
<td>22.96</td>
</tr>
<tr>
<td>IRNZ</td>
<td>20.53</td>
<td>24.47</td>
<td>29.40</td>
<td>24.80</td>
</tr>
<tr>
<td>VC</td>
<td>16.36</td>
<td>18.59</td>
<td>20.05</td>
<td>18.33</td>
</tr>
<tr>
<td>Overall</td>
<td>18.89</td>
<td>22.24</td>
<td>24.97</td>
<td></td>
</tr>
</tbody>
</table>

72.4 with a standard deviation of 14.5. An ANOVA found a significant effect of technique on score \(F(2, 51) = 3.53, p = .0367\). However, post-hoc comparisons with Tukey’s adjustments found no significant difference between any pair of the three techniques.

### 5.4.4 Discussion

As in the AR simulation experiment, the ImageRefinement approach resulted in significantly better accuracy than VectorCloud. However, the differences in this study were much smaller, likely due to the limitations of current real-world AR systems. Four participants complained about the image resolution by saying the image was fuzzy or that IRNZ had a sharper image than IR. In addition, the general performance decrease compared to the VR experiment seems to indicate that the HoloLens still had issues tracking the scene correctly. We tried to compensate for this by regularly resetting the anchor, but some small amount of error likely occurred regardless. The effects of tracking issues could potentially be isolated from the effects of poor camera resolution by looking into how tracking affected VectorCloud, since it does not use the camera at all. However, there might be additional factors also affecting both techniques. We leave this for future work.

The marking time for VectorCloud was similar to the prior experiment, as expected. However, both ImageRefinement techniques took much longer, which resulted in VectorCloud being significantly faster in this study. We attribute the increased time for the ImageRefine-
ment techniques to the slow speed of the HoloLens image capture function, increased difficulty in seeing whether the cursor was over the target with the fuzzy and semi-transparent image, and the button-based control scheme. Again, we note that VectorCloud could be tuned for quicker marking (but lower precision) by using fewer samples.

We did not find significant differences among the techniques based on the SUS data. However, participants’ qualitative feedback was similar to the AR simulation study. Three participants indicated that they had to strain to keep VectorCloud on target and five participants indicated that they held their breath while using VectorCloud. Furthermore, most participants (13) still preferred IR overall, while three participants indicated a preference for VectorCloud, and two preferred IRNZ. Preference was not as overwhelmingly in favor of IR as in the prior experiment, which might be explained by the change in camera resolution and control scheme.

5.5 Summary

The specification of 3D points is a fundamental task in AR systems, and it is not always possible to rely on accurate models of the real-world environment to aid 3D marking. Current AR devices can build an environment model using depth sensing, but only within a limited range. In this work, we have explored three variants of the geometric triangulation approach to 3D marking that rely on neither an environment model nor human depth perception.

We know from prior work that the naïve geometric point marking technique suffers from a lack of precision, especially at larger distances [125]. Our findings indicate that geometric marking can be reasonably accurate at distances up to 85 meters. VectorCloud is more precise than the basic geometric technique, but the lack of user control over the final ray direction leaves room for further improvement. Users perform better with the ImageRefine-
ment approach and also prefer it to VectorCloud.

Our results can be applied to indoor or outdoor AR applications where environment model information is unavailable or unreliable. Specifically, pointing marking can help with collaborative spatial referencing without knowledge of the physical space. With the ability to specify the position of an arbitrary point in space, the pointer can place a virtual marker at that point to create a reference for the observer. Without the model of the environment, however, interpreting the location of the marker in space could still be tricky for the observer. Hence, the success of spatial referencing relies on the observer’s depth perception. We could explore other depth cues that remain effective in the vista space, such as relative size or motion parallax, to help the observer make correct depth judgment.
Chapter 6

RQ4. Supporting Spatial Referencing in Remote Critique

6.1 Chapter Introduction

So far, we have explored the impact of spatial referencing in collaborative AR and means to support spatial referencing. In the next study, we focus on RQ4: how does the ability to do collaborative spatial referencing in shared AR space impact the effectiveness of communication? We seek to strengthen the influence of spatial referencing with evidence of practical enhancement through collaborative presentation and critique around physical artifacts for remote users.

In this study, we designed and implemented ARCritique, a mobile AR application that allows its users to scan physical mockups to obtain and share the generated three-dimensional (3D) models, view the model simultaneously in a shared virtual environment regardless of the users’ locations, and use simple spatial referencing tools to point to and draw on the model in order to aid discussion during design critiques, as shown in Figure 6.1. We evaluated ARCritique through qualitative methods, including interviews and observations, and confirmed that the added support for easy spatial communication contributes to a better collaborative experience.
6.2 ARCritique Design and Implementation

This section will start with a workflow of an iterative collaborative design cycle to provide an overview of ARCritique. We will describe the concept and design decisions of ARCritique to address some of the issues in supporting collaborative critique sessions around physical objects. We also describe the technical details of our implementation.

6.2.1 Design and Rationale

Workflow

ARCritique achieves a workflow shown in Figure 6.2 in an iterative design cycle:

1. User scans the physical artifact with their smartphone, Figure 6.2 1.

2. Once the program generates a virtual model, the user inspects the model’s quality to
6.2. ARCritique Design and Implementation

Figure 6.2: The workflow of ARCritique in an iterative design cycle. 1. The student scans the physical mockup using their phone; 2. The student views the scanned model on their phone to check model quality; 3. The application uploads a qualified scanned model onto a cloud server and shares it with remote users; 4. Multiple users can view the model in a synchronized shared AR space where they can independently point to and draw on the virtual model.

1. The student scans the physical mockup using their phone.
2. The student views the scanned model on their phone to check model quality.
3. The application uploads a qualified scanned model onto a cloud server and shares it with remote users.
4. Multiple users can view the model in a synchronized shared AR space where they can independently point to and draw on the virtual model.

... decide whether they want to upload and share it or rescan and obtain a new model, Figure 6.2 2.

3. If accepted, the model is uploaded on a cloud server and is available for others to download, Figure 6.2 3.

4. Multiple users can jointly inspect, discuss, and annotate the same model in AR even when they are remote and separated, Figure 6.2 4.
3D Model for Mockups

To support natural viewing and collaboration around a physical object, ARCritique needs to be able to generate a 3D model from a physical artifact. One approach is to ask the students to create 3D models using Computer-Aided Design (CAD) software, like Maya or 3DS Max and share them with the instructors directly. While this strategy ensures high-quality 3D models, it requires the students to obtain a certain level of CAD skills, moves away from the benefits of physical mockups, and slows down the iterative design cycle with the extra efforts needed in digitizing the mockup. Moreover, making a purely digital model during the design phase may require simplification from its original form and cannot be placed in physical context to understand the scale compared to other objects (e.g., hands). An optimal solution would rely on the physical mockups created by the students to generate the 3D models and share that model with remote instructors. Thus, ARCritique introduces a scanning process that allows the student to obtain a 3D representation of their physical object.

Scanning

There are many scan algorithms that have been studied and implemented in various fields, such as KinectFusion [110], ElasticFusion [152], instant Scene Modeler [128], 3DCapture [106], and 3DLite [66]. The scanning algorithms can generally be divided into two categories: online and offline scanning. The main difference between these two categories is whether the algorithm continually refines the model as new data becomes available (online) or not (offline). An offline scanning algorithm takes all input at the start and generates the final reconstructed model after some time, making it hard to provide rapid feedback to the user during the scan process. ARCritique takes an online approach as this allows the user to
examine the model during the scan process, as shown in Figure 6.3. Such live feedback allows the users to refine a particular region if they notice flaws in the scanned model. Specifically, ARCritique implements KinectFusion [110] to create the virtual model of the students’ mockups.

The implemented reconstruction algorithm takes depth information from the camera and generates surface mesh inside a pre-defined cubic space inside the camera’s FOV. We define three cubic container sizes (100mm, 200mm, and 300mm) that contain common mockup dimensions and still fit in the depth sensor’s working range. The app can capture as much detail as possible by allowing the user to choose among the three sizes depending on the physical object. Because the position of this cubic volume relative to the real world is determined on the first depth frame, the scanning process requires the user to put their phone vertically at the beginning to ensure the scanned model is not tilted.

Visualization Technology and Hardware

The next decision is how to support viewing of and interaction with the 3D model. Both AR and VR technologies have good support for immersive, realistic, and accurate model inspection [147, 149]. While both AR and VR can be used to support remote critique after obtaining the virtual model of the mockup, VR often requires a head-worn display and blocks users’ view of the real world. In contrast, AR integrates virtual objects with the real world, which can help ARCritique users understand the scale of the 3D models. In addition, AR can be implemented on common handheld devices, such as smartphones. Mobile AR has advantages in availability, affordability, and familiarity since smartphones have become a vital part of daily life. Thus, we selected smartphone AR as the core technology to visualize and interact with the scanned 3D model in ARCritique.
CHAPTER 6. RQ4. SUPPORTING SPATIAL REFERENCING IN REMOTE CRITIQUE

Visual Representations of Collaborator

ARCritique aims to support remote critique through a shared virtual environment. A vital factor to consider in the context of assisted collaborative learning is collaborator awareness. Here, collaborator awareness means the user being aware of the collaborator’s location, looking direction, or potential actions. A user may need to maintain a certain level of collaborator awareness to facilitate communication and comprehension of collaborators’ feedback. The use of avatars can heavily influence collaborator awareness. An avatar is ‘the representation of a user’s identity within a multi-user computer environment; a proxy for the purposes of simplifying and facilitating the process of inter-human communication in a virtual world’ [50]. While studies have suggested that a high-fidelity avatar—such as a full-body avatar with facial features—yields a higher level of awareness, co-presence, and communication [50, 121], the lack of information of the users’ head or hands poses challenges to creating a human-size high-fidelity avatar. There are limited tracking capabilities from the device to capture the user’s postures. As a consequence, the avatars in ARCritique are in the form of a virtual phone (as shown in Figure 6.1). With 6 DOF tracking, the collaborator’s phone’s position and orientation are synchronized and displayed as a virtual phone model. We also consider that the collaborator’s phone can potentially block the local user’s view of the model. Thus, we make the phone model translucent when the collaborator moves between the 3D model and the other user in the shared virtual environment.

Interactions

To support spatial referencing around the 3D model, ARCritique needs to provide interaction techniques to convey communication about a location. We design spatial referencing techniques based on pointing and drawing.
Pointing to the target with virtual rays for selection [16] and communication [3] is widely used in VR and AR. The technique has the advantage of simplicity, naturalness, and effectiveness. The pointing ray technique relies on the device to provide 6DOF tracking on the mobile AR platform. The user can tap a particular point of interest on the model on the screen to indicate the point of reference. Based on the location and orientation of the phone, the system can render a ray connecting that point to the device, as shown in Figure 6.1. When the pointer’s phone moves, ARCritique dynamically updates the ray so that one end of the ray remains attached to the center of the pointer’s phone avatar while the other end is attached to the point of reference.

The second interaction technique is drawing. After discussing with ID instructors, a drawing function was requested to facilitate discussion about design ideas and suggestions. Apart from drawing’s effectiveness in conveying abstract design ideas, it is also beneficial to provide referencing annotation in collaborative systems [140]. Similar to pointing rays, the user can drag their finger over locations on the model to draw annotations. The app performs a raycast from the current finger location towards the phone’s facing direction. If the raycast hits the 3D model, a small cube marker will appear on the model’s surface, and as the user drags their finger, many small markers form what appears to be a continuous curve, as shown in Figure 6.4 b. However, since we rely on the 3D model to find the 3D location for each marker, this implementation does not provide the ability to do mid-air annotations.

It would be difficult to draw in a legible way while the phone’s position is updating the virtual camera at the same time, so when the user selects drawing mode, we freeze the view so that they can concentrate on drawing without having to make an effort to keep their phone still. The frozen view only applies to the user who is drawing and does not affect the view of the collaborator. We also offer an erase mode, where the finger acts as an eraser for previously drawn annotations. Once the user is done drawing, they can press a button to exit drawing
mode and resume 6DOF tracking to view their drawings from different perspectives.

For both interaction techniques, to distinguish the participating users in the shared environment, each user is assigned a distinctive color for their rays and drawings.

### 6.2.2 Implementation

To implement the design, ARCritique consists of four major components:

- A KinectFusion[110]-based processing algorithm to generate point cloud based on the depth data captured by the phone
- Amazon Web Services (AWS)\(^1\) servers to generate the model based on the point cloud and to store and distribute the model among users
- ARKit \(^2\) to support 6DOF tracking and plane detection
- Photon Engine\(^3\) to provide shared viewing and interaction

We developed the ARCritique application on two iPhone 11 Pro Max smartphones with FaceID function. The scanning and model sharing feature was implemented in a native iOS application for low-level development needed for parallel computing and network communication. The collaborative AR experience was embedded on the native iOS application using the Unity game engine with the Photon Network framework.

\(^{1}\)https://aws.amazon.com/
\(^{3}\)https://www.photonengine.com/
Figure 6.3: Overall procedure of mobile KinectFusion implementation. [*] is inspired by the work of Shen [129].

**KinectFusion**

Since iPhone X, any iPhone with FaceID technology is equipped with a TrueDepth camera⁴ (a camera and infrared projector pair that can be used to create a depth map of a scene). ARCritique utilizes the TrueDepth camera to implement a simplified version of the KinectFusion algorithm.

The procedure of the algorithm is illustrated in Figure 6.3. The algorithm uses depth images from the TrueDepth camera and merges consecutive frames to maintain the surface information of the scanned mockup. Due to limited computational power on the mobile device, the algorithm runs with reduced depth resolution and voxel resolution and fewer iterations than the original KinectFusion algorithm. Although ARCritique renders a 3D model during

⁴https://support.apple.com/en-us/HT208108
the scan process to provide live feedback, the indexing of the vertices simply considers three consecutive points to be a triangle in the mesh. The precise indexing of the resulting 3D point cloud is also challenging on the mobile platform. Thus, the algorithm outputs a point cloud with normal data and uploads it to an AWS server for post-processing.

One notable drawback of our implementation is that the TrueDepth camera is located on the front side of the phone, meaning that the user has to hold the phone facing the physical object and loses access to the screen. Our solution was for the user to mirror or project the phone’s screen on a computer during the scan process. This is reasonable, because the remote critique session typically uses a videoconferencing app (e.g., Zoom\(^5\)), and the same app on the same computer can be used to mirror the phone’s display. The implementation also requires the camera to scan the physical mockup from all sides. A user can orbit around the object while keeping a constant distance and a more reliable approach is to place the object on a turntable, hold the phone still and face the object, and rotate the turntable to get a complete scan.

**Cloud Servers for Constructing and Sharing 3D Models**

ARCritique functions with two AWS servers. An Elastic Computing Cloud (EC2) server is in charge of final model generation using Poisson surface reconstruction \[^76\]. Then the generated model file is transferred to a Simple Storage Service (S3) server for storing and distribution. Meanwhile, upon the file’s generation, it is immediately downloaded to the phone (of the user who performed the scan) for preview. The user can inspect the entire model on their phone by rotating it around and zooming in with their fingers (as shown in the right of Figure 6.3). The total time for model generation, including data transfer, is approximately 30 seconds.

\[^5\]https://zoom.us/ Zoom has a share screen function for iPhone/iPad by default.
6.3 Evaluation

To share the model with remote collaborators, ARCritique can generate an email with a specific URL link. When a recipient clicks the link on a device on which the app is installed, the model will be automatically added to the app’s local library. Alternatively, the student can also share the model ID directly with their instructor, and the instructor can manually add the model.

**ARKit and Photon Engine**

The AR-based real-time, shared critique functionality is implemented through Unity\(^6\), ARKit, and Photon Engine. Once users select a virtual model, they join an online session with all other users viewing the same model. The application uses ARKit and asks the user to move their phone around to capture the user’s physical environment and detect a horizontal plane to place the virtual model. Apart from the ability to navigate around the space via 6DOF tracking, the user can also press a button to detach the object from the plane and use their finger to swipe left or right to rotate the model around its vertical axis. ARKit also handles 6DOF tracking of the device, providing the freedom of seeing the model from any perspective. ARCritique synchronizes remote users’ phone positions and orientations in the same session by communicating their poses relative to the model through Photon. The poses are used to update the avatars of each user (see section 6.2.1).

### 6.3 Evaluation

We conducted a user study with seven students and three instructors from the ID program at our university. We invited the participants to try out ARCritique in ecologically valid design critique settings by inviting them to engage in the same type of design discussions in

\(^6\)https://unity.com/unity/features/arfoundation
their educational activities with mockups from their classes. We designed the experiment to investigate the following RQs:

- **RQ 4.A.** What are the issues with conventional remote videoconferencing tools for design critique sessions?

- **RQ 4.B.** How well does ARCritique address the problems associated with conventional remote videoconferencing tools?

- **RQ 4.C.** What is the perceived usability of the ARCritique app?

### 6.3.1 Tasks

The primary task of the study was the remote collaborative design critique between the student and the instructor using ARCritique. The free-form critique session centered around the student presenting their mockups to the instructor remotely through ARCritique for viewing and interacting with the model. Simultaneously, they used Zoom for voice communication. Each student had three artifacts from their prior classes to discuss with the instructor, resulting in three critique sessions using ARCritique. Each critique session started with the student introducing the artifact to the instructor via verbal description and/or referencing tools like pointing or drawing. The instructor then asked questions for clarification, suggested improvement ideas, or introduced materials relevant to the student’s design. The task was designed to reflect an authentic and synchronous critique session that generally took place in class as an educational activity. The Zoom usage in the study was strictly limited to voice conversation with no other visual communication or interaction tools. We did not constrain the conversation, because in real life, such creative design critiques are dynamic and vary by participant. A free-form discussion could help preserve ecological validity.
A student had an additional task prior to the remote critique session: 3D model scanning. Guided by the experimenter, the student needed to use the phone to complete scans to create virtual replicas of their three physical mockups. Placing the mockup on the turntable, the student slowly and steadily rotated the model while holding the phone still. After the mockup was turned 360° horizontally, in order to capture the mockup from above, the student might need to maintain roughly the same distance between the phone and the mockup, tilt the phone, and move it up. After the model was generated, the student then inspected the quality of the scan to determine whether it was sufficient for a critique session. This continued until the student ended up with three good-quality models for the critique session.
6.3.2 Measures

We were primarily interested in qualitative insights that can help us understand the difficulties the users may face in current remote communication tools and their perceived benefits of using ARCritique. We analyzed their communication, observed spatial referencing occurrences, and inquired their perceived usability. Each critique session was video and audio recorded by the experimenter using Zoom. During the critique session, the instructor was asked to share their phone’s screen in Zoom for the purpose of recording (Figure 6.4 b). The experimenter joined the student and the instructor in the same shared AR space and observed their interactions from a third perspective (Figure 6.4 a). Finally, an audio-recorded exit interview and a post-study questionnaire helped gather subjective feedback. The questions asked during the interview and the questionnaire are listed in the Appendix. Both the participants’ conversations and the interview were transcribed into text using Microsoft Word. The first author manually went through the videos to correct any errors in the initial transcriptions.

6.3.3 Participants

Seven students (three female, mean age 20) and three instructors (one female, mean age 50.67) from the ID program were recruited. Instructor 1 (I1) participated with two students (S1 and S2), I2 participated with four students (S3 to S6), and I3 participated with one student (S7), forming seven pairs. All participants were reasonably familiar with Zoom-based online classes (had more than ten prior experiences). Three students and the instructors had used Zoom for remote critiques around physical mockups more than ten times, while the other students had experienced them more than three times. Five students and one instructor had

\footnote{https://support.microsoft.com/en-us/office/transcribe-your-recordings-7fc2efec-245e-45f0-b053-2a97531ecf57}
tried AR no more than three times, and the rest of the participants had used AR fewer than ten times.

6.3.4 Procedure

The experiment was broken down into three sections for each student-instructor pair. First, the students were given a brief introduction to the study and the phone being used in the scanning session. Then the experimenter would walk the student through a guided scan process using the phone and the turntable. The students were given 30 minutes to scan and generate three virtual models based on their mockups. Second, prior to the critique session, the experimenter would guide the student and the instructor through the process and familiarize them with the interaction techniques. Then the student and the instructor would conduct the free-form critique about the three scanned models in 30 minutes. Finally, each participant (student or instructor) ended the study with a post-study questionnaire and an exit interview. The questionnaire for the student added one section inquiring about their scanning experience while the rest focused on their critique session.

6.3.5 Analyses

We mainly employed thematic analysis [100] on the transcribed text of the interview and questionnaire. Critique session recordings and transcriptions served as the materials for observational summarization.
6.4 Findings

6.4.1 RQ 4.A. Issues with Zoom-based Remote Critique

In order to understand whether ARCrtique’s support for spatial referencing would actually be relevant to the challenges encountered by the ID students and instructors, we asked the following questions during the interview:

- Could you tell us about your experience using Zoom for discussing physical mockups?
- What major challenges do you experience when using Zoom during remote critique?

Through thematic analysis of the interview responses, we identified three major issues with Zoom-based remote critique that are relevant to spatial referencing.

Communication

There are several issues that the participants reported when communicating with each other during remote design critique via videoconferencing. The verbal communication itself was perceived as less fluid than that in an in-person critique session. For example, when asked to compare the Zoom-based critique with in-person experience, I2 commented that: “It’s a little bit confining because conversations don’t happen in the same fluid manner that they do when you’re in person.”

The communication might appear to be disconnected. I2 considered the communication around physical mockups during remote sessions as limited or disconnected from the actual artifact. They commented: “You never...sit with the work in the room. It’s never sitting next to you while you’re working on something else and sparking your memory, and so it’s really disconnected in that way.”
Meanwhile, the verbal-only communication over Zoom turned out to be insufficient for proper discussion in a critique session:

S2: “When it comes to more complex physical models, it can be difficult to fully get the idea across.”

S4: “But definitely when you’re in person, you can exactly see how the model looks and like how big it is and, nothing is lost in the translation when you’re trying to describe it.”

On the other hand, certain features from Zoom helped with the communication in remote critique. For example, I1 spoke highly of the whiteboard function during the interview: “In a physical critique, you can point, manipulate and pick up, but if you have an idea, you either have to go to a whiteboard, which hopefully there’s one in the room, or if not, you have to say, do you have a scrap piece of paper and then you can sketch, whereas with the Zoom you do have an option of a whiteboard.”

Perspective

The second major complaint that the students and instructors had about Zoom-based critiques was the limited perspective. Students said that during remote Zoom critique sessions, they mostly sent photos of their mockups from selected perspectives to their instructors. The task of taking multiple pictures was not a trivial process. The student might take a lot of pictures and choose the ones that best conveyed their mockups. Taking photos also played a vital part during this process:

S6: “The the biggest thing was you had to take a lot of pictures to convey what you want.”
S7: “Yeah, one thing that I attempted to do was take multiple pictures of the model at different angles. And so the photography I heavily depended on, the quality of the photos and the angle at which I took the photos.”

The similar extra effort of using additional camera also applied to the instructors in their classes.

Another issue about still images was that both the student and the instructor found it challenging to fully understand the artifact as the perspectives appeared disconnected, unlike the in-person session, where the viewers could inspect the physical craft from various angles to inspect its scale, details, etc.

S1: “At least in first year [when classes were in-person], where we would have like presentations and stuff, everybody could stand off towards the thing and then we could look at it from various angles.”

S2: “You can still do a decent amount over Zoom by just holding up the model to the camera, rotating it around, but there is something to having somebody be able to pick up your model and look at it closely at their own will and see the details that they want to look at in particular.”

**Human Scale**

The lack of understanding of scale and form factor was perceived as a key issue with Zoom-based remote critique. During the interview, participants mentioned that the objects they discussed in critiques were mainly those held in the hand or operated by hand, such as brushes, toasters, or toys. Thus, it was essential for them to understand the object’s scale relative to the hand or human body. However, when the understanding of scale was hard to
achieve from the still pictures, the student had to use excessive verbal description to convey that to the instructor.

S6: “And it’s like always a professor ... how a human person holds it, just so he could get a human scale of it.”

S7: “I think there were a few times where they were curious about the scale in which I would verbally explain.”

### 6.4.2 RQ 4.B. Benefits of ARCritique

To investigate the second research question about how well ARCritique mitigates the issues of videoconferencing-based remote critique sessions, we analyzed participants’ interview responses and our direct observations of their use of ARCritique.

### Interview Feedback

Based on prior literature, we first identified four features of ARCritique that might help with critique around 3D objects: unconstrained perspective, collaborator awareness, pointing, and drawing. We asked the participants about their opinions about these features, and whether they thought each feature helped or hindered the critique session.

**Unconstrained Perspective** When asked about the ability to view the model from any perspective in AR, all 10 participants agreed it was important for the remote critique experience. They thought this feature supported their discussion for different reasons.

The primary benefit was the ability to perform inspections freely. Six students expressed appreciation for being able to see the model from different angles:
S4: “yeah, I think it is [important]. Because you get to see the product in every view but still have it like in the scene and see how it will look.”

S5: “I mean there can be a lot of different things going on in the model and being able to see what’s going down the sides obviously can be important.”

Another major benefit was the added understandability. Four students and two instructors pointed out that since the mockups were made in 3D, the corresponding models should also be perceived in 3D. The ability to navigate freely in space or rotate the model with their finger helped reveal subtle information:

I3: “but these (different sides of the model) are all very subtle and it’s very three dimensional and you know, I mean, it’s not like straight on like you never see that you know, you never see that at all (from still images).”

S6: “There is an empty space on the back side, but on the front side you couldn’t really tell and so being able to turn that around, you can see that empty space and reiterate that more. But I feel like in terms of turning it around, I feel like there were really no downsides to it.”

S5 commented that while this feature was important for inspecting, the ability to unlock the object and rotate the model through their finger added ease and did not affect other person’s view.

S3 added that being able to see the model from different angles, either through moving around or rotating the model on the phone screen, helped them understand the scale of the model in relation to other objects.
Collaborator Awareness  Seeing the collaborator’s phone model in the shared virtual space received mixed responses. Seven participants agreed it was important and helpful to see their collaborator’s phone model, but for different reasons. A primary reason was that it helped understand where their collaborator was looking by seeing where their phone was and which direction the phone pointed. This was more prominent when the collaborator was looking at a specific region of the model that was not immediately visible to the user. By knowing where the phone was, the user “could move to where the phone was and be looking from the same perspective.” (S7) Such awareness could also be utilized in communication, supported by S4’s comment: “So if I could see [her] phone, I could help guide her to the area that I wanted to discuss.”

S2 and S6 had overall negative thoughts about the feature, mostly caused by the collaborator’s phone model occluding their views. S6 mentioned the phone model interfered with the drawing function during the critique session: “when I was looking at my point of view, the phone would be right in front of my view when I try to draw notes, the phone would be in the way and I couldn’t move it out of the way, so that definitely hindered a lot in terms of like me being able to draw on the physical model.” Even though we made the collaborator’s phone translucent when the collaborator moved between the critiqued object and the other user, S2 still thought the phone model was interfering with their view of the critiqued object and hindered their experience: “Even though when you pass over the other person’s phone, it becomes like translucent. It still obscures the view a little bit, which was a little bit, but overall that wasn’t too bad.”

S5 held a neutral opinion towards the collaborator’s phone as they found it to be less noticeable: “I found the other tools to be more useful. Partially it’s kind of maybe being concerned that maybe there might be a bit of lag like where exactly they are.”
Pointing  The feedback we received about the pointing ray function was generally positive, with only S4 holding a negative opinion. The major benefit of using the ray was assisting description, since pointing with the ray allows a type of deictic gesture [35] that can accompany speech when referring to a feature or location on the artifact. Five students mentioned the rays made their communication with the instructor easier when they needed to describe specific details on the model:

S3: “it allows for specific critiques like you know, 'this area is like a little bit large', you know 'you need to like slim that down' or 'this curve is wrong right here'. The instructor can point that out and show you instead of just trying to describe it, and failing to do so.”

S7: “Especially with a lot of detail in the model where maybe there were indents or there were specific parts that were darker than others. The pointing tool definitely helped in pointing out a specific area. And again, it was a better method than trying to verbally describe where that area was, so I thought it was a very helpful tool.”

S4 had a negative opinion the tool because of missed touches. They complained that: “when I was trying to click the ray it would miss my product and hit somewhere in the background.” The missed pointing caused frustration during the critique session when the student found that the ray confused the instructor and essentially slowed down the process.

Another notable point raised by S2 was that they found that “the current implementation hindered the the view of the person using the ray.” Since we made the ray dynamic and attached one end to the pointer’s phone model in the virtual environment, a ray tip appeared on the center of the pointer’s screen. However, S2 still agreed the tool was beneficial to the critique.
6.4. Findings

Drawing All ten participants agreed that drawing was an important feature that greatly supported the critique session. We proposed the drawing functionality as an annotation tool for referencing tasks. This was supported by participants’ responses stating that the drawing function simplified their description and aided their explanation and understanding as they can draw on top of under-specified regions:

S1: “I had an LED panel that was kind of stuck out with a Cardboard. So I would draw on top of that LED panel and then I’d say ‘hey this is a LED panel and these are buttons that would connect to the panel’ so I could draw arrows and stuff like that to show the significance of each feature of my toaster in that case.”

S3: “And I think that the drawing provides that ability, whereas if you didn’t have a drawing, you know, they would just have to describe it.”

Similarly, the drawing feature was also helpful in conveying new or alternative design ideas. One instructor and five students liked to use the tool to convey design suggestions or changes by drawing them out on the model. This was particularly useful as it could add information beyond the model. Since ARCritique only scans the geometric shape of the mockup, the actual texture was missing in the resulting model. Moreover, depending on the scan, the resulting model could even capture the turntable and create confusion. In these cases, the students used the drawing function to “be able to draw out some details” (S5) or “draw on features I wasn’t able to explore in my physical prototypes” (S6).

One complaint we received from the students and the instructor was that it was not easy for them to draw precisely using their fingers.

I2: “We weren’t always drawing exactly where we thought we were, but I think it’s...absolutely necessary to have a drawing feature.”
S6: “It wasn’t as accurate as I thought it would be like if I wanted a really straight line. I couldn’t do that.”

Critique Observation

A second approach we took to investigate the benefit of ARCritique was through analysis of the critique session video recordings. Since seven student-instructor pairs participated in the study, and each pair contributed to three critique sessions with different mockups, the study ended with a total of 21 critique sessions. The first author went through the critique session recordings and kept track of conversations involving spatial referencing. Specifically, the first author looked for conversations containing referencing words, such as this, that, here, on top of, etc., or explicit attention guiding requests, such as do you see where my ray is. We recorded these conversations and noted whether or not the student or the instructor used the pointing ray or drawing when these conversations happened. We distinguished between cases where the student or instructor actively placed the ray or drew an annotation and cases where they discussed the previous ray or annotation. Finally, to understand communication efficiency, for each spatial reference we measured the conversation time (in seconds) and counted the number of words used.

Table 6.1 shows some examples of the captured conversations.

We were able to identify 193 conversations related to spatial referencing. The participants applied the ray for spatial referencing 69 times (36%), with 3.19 seconds mean conversation time and 11.65 average words. 62 (32%) spatial referencing conversations involved the user drawing annotation or discussing the previous annotations, with 4.54 seconds mean duration and 12.98 average words. There were 62 occurrences (32%) where the participants did not use our referencing tools, with an average of 6.3 seconds and 18.02 average words. The result
Table 6.1: Examples of spatial referencing conversations.

<table>
<thead>
<tr>
<th>Utterance</th>
<th>Duration</th>
<th>Word Count</th>
<th>Tool Use</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>I see the edge of that turntable and I see the two planes. The bottom one is definitely a pyramid. the top one I think is a pyramid too.</em></td>
<td>12.4</td>
<td>29</td>
<td>-</td>
</tr>
<tr>
<td><em>Could you draw a part line for me between the the trailer and the box?</em></td>
<td>4.6</td>
<td>15</td>
<td>-</td>
</tr>
<tr>
<td><em>OK, so you put toast on both sides is that right?</em></td>
<td>1.9</td>
<td>11</td>
<td>Discuss Ray</td>
</tr>
<tr>
<td><em>See where I’m pointing?</em></td>
<td>0.9</td>
<td>4</td>
<td>Point Ray</td>
</tr>
<tr>
<td><em>I mean from my position, I can see like looks almost like that.</em></td>
<td>3.4</td>
<td>13</td>
<td>Draw</td>
</tr>
<tr>
<td><em>Like this is like the lip of a pot. Does that make sense?</em></td>
<td>3.1</td>
<td>13</td>
<td>Discuss Drawings</td>
</tr>
</tbody>
</table>

is shown in Figure 6.5. In short, the analysis indicated that the participants spent less time and fewer words in spatial referencing conversations when using the referencing tools, and using the virtual rays took even less time than using the drawing tool.

We initially visualized the data to see whether the participants’ interview responses about the benefits of the spatial referencing tools were supported by our observation. Because the data were not normally distributed and we had three groups to compare, we ran the Kruskal-Wallis test [63] on both time and word count. The test found significant effects of the interaction type used (None vs. Drawing vs. Ray) on both measurements. We then performed Dunn’s test [36] for non-parametric pairwise multiple comparisons and found that for time, all pairs were significantly different; while for word count, all pairs but Drawing vs. Ray were significantly different. Given that our number of data points was not large, we performed a post-hoc chi-square power approximation for the Kruskal-Wallis test and found
6.4.3 RQ 4.C. Perceived Usability

We were interested in the participants’ perception of the ARCritique app’s usability throughout the entire process, including scanning and critiquing. The post-study questionnaire (listed in the Appendix A.6.2) helped us gather the perceived usability from two themes: Scan Process, Critique Experience.

Scan Process

Overall, the scanning process did not receive positive feedback from the students. Six students found the scanning procedure tedious or difficult, at least in their first scan. We received complaints about the requirements of holding the phone vertical at first, maintaining the right distance between the phone and the mockup, and pressing buttons with the
screen facing the mockup. However, four of these six students added that their performance improved as they scanned three mockups for the critique session.

S2: “Scanning objects was also quite tedious. Needing to spin the model on a turntable while also holding the phone straight up and at just the right distance is quite awkward.”

S3: “Scanning, at first, wasn’t difficult, but wasn’t easy—having to hold the phone vertical, making sure the object was close enough, and pressing the buttons all at once was relatively hard to coordinate. After 2 or 3 test runs, though, I was able to scan without a problem.”

The remaining two students found the scanning algorithm limiting. S4 found the scanning frustrating when they had to try multiple times to get the correct model. S7 brought a large model that exceeded the camera’s field of view and experienced difficulty in scanning.

S6 claimed the process was “pretty easy” and they had no problem with their model.

Critique Experience

When asked about their critique experience, nine participants expressed that their communication went smoothly. They also thought the ARCritique app was “intuitive and easy to use.” (S6).

However, the participants agreed that the scanned models were simplistic and might even jeopardize the critique experience. The models appeared to participants to be made of clay, and did not carry any texture. Limited by the camera, the scanning precision was also not perfect. S1 thought that while the scan only captured the basic shape, the models were “all present in the scan, so it was easy to just point them out.” Other participants criticized the scanned model for the lack of details. S4 commented that: “I was not able to communicate
my mockup that well mostly because the scan left out a lot of the details on models. I found myself trying to explain what the features looked like and how they worked and spent less time explaining why I made those choices when making the models.” S2 also thought the model’s low fidelity “made it very difficult to accurately show what they were presenting.” All three instructors also wrote suggestions about the importance of a better-scanned model for the critique session.

Half of the participants thought the ARCritique experience was closer to an in-person critique session, mainly because of the ability to navigate the model and reference specific sections freely. To them, such experience was not available in a Zoom-based remote critique and appeared as a successful imitation of an in-person critique. However, some participants thought the experience was closer to a Zoom call but for different reasons. S5 acknowledged the benefit of unconstrained 3D perspectives but thought they could have done the same with Zoom-based presentations. I3 thought that the lack of details on the scanned model made the experience closer to Zoom critique. The other participants thought the experience was somewhere in between because Zoom was still used for voice communication during the study.

It was clear that most of the participants’ complaints regarding the overall experience were caused by the scan defects. The instructors all mentioned that the scanned model’s lack of texture or color details made the app suitable for the early design phase, where the focus was on shape and outer form.

6.5 Discussion

Through the study, we were able to identify key issues with videoconferencing about spatial referencing in remote design critique centered around physical mockups. The participants
reported different issues based on their experience of previous Zoom-based remote critique sessions. Some of the issues exposed insufficient support for spatial information communication via Zoom, such as the lack of shared 3D understanding of the object, limited perspective, and the loss of the object’s scale. The participants also tried to overcome these issues by taking multiple pictures or videos to help convey their mockup. Apart from the extra efforts they had to take, the images were still insufficient to support an efficient and effective critique session as the in-person experience. These issues led to a reduced understanding of the physical object from the remote collaborator’s side, hindering communication around the mockup. A similar observation was also reported by Mok and Oehlberg [103].

To address these issues, we designed and implemented ARCritique with a specific focus on improving spatial referencing communication. Using the app, the user could scan their mockup to generate its virtual replica, share the virtual model with remote collaborators, and have a synchronous critique in a collaborative AR environment with two interactive referencing techniques: pointing and drawing. Viewing the 3D model in AR with 6DOF tracking allowed independent navigation in the virtual space, resulting in unconstrained perspectives. All participants praised this feature as it provided a view method that was close to the in-person experience. The feature provided a naturalistic way to inspect the model and improved the remote collaborator’s understanding of the model’s scale and physical form.

The avatar we added to help the users establish collaborator awareness was not as successful as unconstrained navigation. While most participants found this feature important and helpful for the critique session, there were clear critics. Participants complained about the phone avatar being obtrusive when they viewed or drew on the model. The movement of the phone avatar might also be affected by internet connection and introduce lag feeling to the participant. The mixed results suggest that the need for an avatar might change throughout
different phases of the critique session, and an always-on avatar might not be necessary. The obtrusive experience happened when the user was trying to draw on the critique object, whereas the avatar appeared less noticeable during other times of the critique session.

The pointing rays and drawings were generally perceived as beneficial to the critique session. They both supported easy and natural interactions that the students and instructors were used to in in-person critiques. The negative experience was mostly about executing these interactions on the phone’s screen. The finger might not be as agile and precise as pointing and drawing tools used in in-person critiques. A more carefully designed user interface can potentially help with this problem, such as progressive refinement [83].

The benefit of the interaction techniques was also partly supported through the critique observations. We found that the spatial referencing conversation time was the shortest when the participants used virtual pointing rays. While the participants might spend longer time than a single tap on the screen using the drawing tool, their conversation time was still less than when they only communicated verbally. Similarly, significantly fewer words were spoken when the participants used ray or drawing tools since referring to the area of interest might require excessive verbal descriptions. These results indicated improved efficiency and effectiveness from provided spatial referencing interactions. Moreover, the feedback we received from the participants about the drawing functionality was inspiring. While we initially developed the drawing technique to help reference or annotate the models, the participants also used this feature to augment their prototypes and exchange design ideas.

The overall critique experience was satisfying. ARCritique was found intuitive and easy to use during the main critique session. The primary issue with ARCritique’s current implementation was about the scanning process and the quality of the scanning result. The scan only produced a gray clay-like virtual replica with less than ideal precision. While the overall shape was well preserved, the lack of texture, color, and fine details brought a significant
hurdle to the overall experience. In its current version, the instructors concluded that AR-Critique was only suitable for the early design phase when the primary focus was for the student to decide the form of their mockups. This problem was caused by both the hardware and software. A more dedicated but higher-cost 3D camera could produce a more precise virtual model. An improved model reconstruction algorithm such as Elasticfusion [152] could add more details to the resulting model missing from ARCritique.

6.5.1 Design Suggestions

Based on our observations, the support for spatial referencing is proven to be beneficial in remote critique sessions around physical models. Thus, we suggest that designers and developers consider them when designing and implementing similar collaborative AR experiences. Apart from rays and drawings, other tools and visualizations can also be explored. For example, when eye tracking is available, visualizing the user’s focal point may be a convenient and efficient approach to share referencing information. Hand tracking may also be incorporated to express referencing through gestures such as pointing. Gesture-based referencing communication should also be seen as natural and easy to use as it happens pervasively in our daily communication.

The quality of the virtual replica of the physical mockup depends on the task. As commented by the participants, models of low fidelity may be sufficient for early-stage design critiques. In such a case, a low-fidelity model preserving the shape may be appropriate from a cost-benefit perspective. On the other hand, tasks like detailed inspection of near-final prototypes may require a reconstruction process that represents as much fine detail as possible to support accurate understanding and communication.

Collaborator awareness can be achieved by the use of an avatar, as shown in the literature [55,
However, the need for such an avatar may change throughout the critique session based on the user’s activity. In particular, when the user focuses on the model and intends to perform certain tasks on it, such as annotation and modification, their need for collaborator awareness may be low, and they may even wish not to be interrupted by their collaborator. On the other hand, the user may need to be aware of their collaborator’s location, looking direction, and point of interest during a design discussion to express and exchange ideas. Hence, it is worth considering a dynamic representation depending on the user’s local activity so that the remote collaborator will not interfere with the local user. When to present, dismiss, or even highlight the collaborator avatar is a question for the designer.

Last but not least, specifically for design critique applications, augmentation can be a powerful tool to help the users convey sophisticated design ideas and explore beyond real-world limitations. For instance, the drawing tool may need more complex modes than the current ARCritique, such as different drawing styles or drawing in the air. Other augmentation tools may also improve usability, including texture attachment or 3D widgets.

6.6 Summary

In this work, we present ARCritique, a mobile AR app for remote collaborative critique sessions around physical mockups. We conducted an observational user study to understand how support for spatial referencing in a mobile AR system can address some of the limitations of videoconferencing tools. The study demonstrated that improving the effectiveness of communication about 3D objects via well-supported spatial referencing makes communication more efficient and fluid during such critique sessions. Although some parts of the application need refinement, the core experience of collaborating around a 3D model of a physical mockup in a shared AR space proved successful in an ecologically valid setting. Our
results shed light on how to enhance remote collaboration experiences that focus on physical objects. Lastly, compared with works that sharing the local users’ egocentric view with a remote expert [47, 48, 96], remote critique requires a capability for remote experts to see the 3D model from various angles. The object-centric context sharing illustrated by ARCritique is a less studied scenario with many research opportunities. In this sense, ARCritique also contributes to the community to provide initial insight into critique sessions centered around physical models.
Chapter 7

Conclusions and Future Work

7.1 Conclusions

In this dissertation, we proposed four RQs to address known issues about spatial referencing in collaborative AR. Table 7.1 provides an overview of different research dimensions that have been covered in this dissertation. The Reference Range denotes the distances between the user and the referred target that is either near-field or far from the user (in the vista space). Communication Space is the space in which the users exchange information [127]. In the study, we explored three different spaces from a typical co-located, to more challenging wide-area, and to remote spaces with increasing interest during the pandemic. In Communication Methods, we list out the investigated spatial referencing approaches. Model Availability specified the type of model used for spatial referencing and their availability in different research projects. Referred Target distinguishes whether the referred target is real or virtual.

RQ1 focused on close-range co-located barehanded referencing in AR when precise hand tracking is unavailable. We compared participants’ performance on target identification in the physical environment vs. in AR without correct occlusion cue between the user’s hand and the virtual target. We found that participants’ performance (both in time and accuracy) was indeed reduced in the AR condition, despite the participants’ efforts in coming up with various mitigation strategies. Based on the findings, we suggested multiple design implications, such as changing the spatial layout of virtual content to reduce referencing...
7.1. Conclusions

<table>
<thead>
<tr>
<th>Reference Range</th>
<th>Communication Space</th>
<th>Communication Methods</th>
<th>Model Availability</th>
<th>Referred Target</th>
</tr>
</thead>
<tbody>
<tr>
<td>RQ1. Barehanded Referencing</td>
<td>Near</td>
<td>Co-located</td>
<td>Hand pointing</td>
<td>Missing hand model</td>
</tr>
<tr>
<td>RQ2. Pointing Rays</td>
<td>Far</td>
<td>Wide-area</td>
<td>Pointing rays</td>
<td>Missing target model</td>
</tr>
<tr>
<td>RQ3. Point Marking</td>
<td>Far</td>
<td>Wide-area</td>
<td>3D position annotation</td>
<td>Missing target model</td>
</tr>
<tr>
<td>RQ4. ARCritique</td>
<td>Near</td>
<td>Remote</td>
<td>Pointing rays, drawings, verbal communication</td>
<td>With target model</td>
</tr>
</tbody>
</table>

Table 7.1: Research dimensions covered in this dissertation.

ambiguity. RQ1 provides insights into the danger of missed occlusion on the understanding of hand referencing in near-field model-free AR setup.

RQ2 involved two user studies evaluating pointing rays as a reference tool for distant objects in model-free AR. We proposed two types of enhancements to a single pointing ray - the Double Ray technique (enhancing visual match condition) and the Parallel Bars technique (providing artificial spatial cues) and compared them in both simulated AR and outdoor AR. Both experiments found the Double Ray technique helpful, and its extra cost on the pointer’s side was well paid off as the observer could more likely correctly identify the target at about 70 meters. While the Parallel Bars technique failed to provide concise information for the task, participants’ feedback approved the idea of extra spatial cues.

We developed two types of marking techniques to mitigate the problem of model-free AR in RQ3. With the marking techniques, the user could be able to specify the location of arbitrary points in space. We evaluated our proposed techniques - the VectorCloud technique (taking multiple samples to attenuate errors from pointing) and the ImageRefinement technique (using screenshots to remove pointing errors in marking calculation). Both studies illustrated the effectiveness of both techniques. While the ImageRefinement technique had better marking accuracy, the VectorCloud technique was faster between the two. RQ2 and 3 explore solutions to address the model-free AR problem in more challenging outdoor space. The techniques designed in the two projects shed light on supporting spatial referencing even in the most difficult situations without knowledge of the physical environment.
In RQ4, we presented ARCritique, a mobile AR application for remote collaborative design critique around physical mockups. An observational user experiment demonstrated that with good support for spatial referencing, ARCritique could address some limitations of videoconferencing tools. The enhanced spatial referencing contributed to more efficient and fluid communication during the critique session. In RQ4, we extend the dissertation into a more realistic setting by allowing verbal communication and incorporating imperfect models from the physical object for spatial referencing. RQ4 helps us understand the impact of spatial referencing in a realistic collaborative task where reaching spatial consensus is critical.

In summary, the dissertation’s major contributions include:

1. Providing evidence of the danger from poorly-supported spatial referencing in simplistic collaboration setup (RQ1)

2. Designing different techniques to address spatial referencing problems in the most difficult situation without knowledge of physical space (RQ2 & 3)

3. Illustrating that collaborative tasks requiring frequent spatial consensus can benefit from different spatial referencing channels that current technology allows. (RQ4)

### 7.2 Future Work

Finally, I present several directions for future work as a conclusion to this dissertation. The dissertation proposes these future research possibilities within the broader area of 3D UI design for collaborative AR.
7.2. Future Work

7.2.1 Barehanded Referencing

In the study of RQ1, we established some basic understanding of the negative impact of model-free AR in barehanded referencing. We omitted any tracking technology in favor of focusing on the baseline model-free condition because having an unreliable model could be as bad as having no model at all. While a perfect solution for real-time reliable hand capture is still unlikely to be around any time soon, with constant advancements in AR technology, we expect to see improved hand tracking and modeling quality. Additional information from a tracking system may potentially help improve collaborative performance by providing occlusion cues with errors. Notably, it is also unknown whether these errors may introduce systemic bias. Hence, designing a barehanded user interface that incorporates state-of-the-art hand tracking and modeling technology and testing this interface in a more ecologically valid task can bring significant value to AR designers and application developers.

The current experiment is somewhat more abstract than many real-world tasks. Our study assumes the targets are discrete, which is not always the case. For example, tasks like factory machine inspections or collaborative map inspections are more likely to work with continuous virtual content.

Another promising direction is to explore a collaborative AR system that automatically changes the layout and scale of content based on the users’ positions and viewing angles based on the design implications. By incorporating machine learning to make the system aware of the spatial context, the system can adjust the positions of the virtual content to minimize overlay between physical and virtual objects. In this way, collaborators may have attenuated possible ambiguities caused by incorrect occlusion.
7.2.2 Collaborative Pointing Rays

Based on the results from the collaborative pointing rays study (Section 4.4), we can propose two possible future work directions. First, we plan to design and explore other ways of enhancing spatial perception with the 3D geometrical information of the pointing ray and the users. While we observed limited effectiveness from the Parallel Bars technique, the user preference from the study still suggests promising potential in providing artificial spatial cues to address the issues in distant object referencing in model-free AR. During the user interview, one of the reasons the participants brought up for preferring the Parallel Bars technique over the Single Ray technique was the added confidence. The users said they could consult with the extra information in an ambiguous case. Even though it took the users additional time and effort to interpret the artificial cues, they felt their performance improved and found the technique better than the Single Ray technique, where they could only rely on their best guess. A candidate approach can be sharing of the pointer’s perspective, as often seen in the literature [45, 73]. However, we must carefully explore enhancements to the naïve implementation in wide-area collaborative AR where the collaborators’ perspectives can be extremely distinct.

Another future work direction is the subjective usability measurement in an asymmetrical collaborative context. We used a modified version of the SUS and the raw NASA TLX questionnaires in the experiment. We found inconsistency between the questionnaire scores and the user ranking or preference. We suspected that these questionnaires were designed to measure single-user scenarios. In our case, when asked about their preference, the participants considered both the pointer and observer experiences. This was reflected during the interview that many participants admitted even though the Single Ray technique was easier and faster to use than the Double Ray technique, they still chose the Double Ray technique as their first choice because of its benefit to the observer. A potential approach can be to
design an equation to unify subjective scores from different roles. A critical challenge is to produce a generalizable equation instead of an ad-hoc one. However, such an asymmetrical collaborative usability score should demonstrate to be helpful to help designers and practitioners make critical decisions in collaborative system design.

### 7.2.3 Point Marking

We introduced two point marking techniques and demonstrated their strength and limitation through two user studies. As discussed in Section 5.5, a direct future research direction is to study the usage of the marking techniques in a collaborative spatial referencing context. By allowing the pointer to specify the position of a distant object and place a virtual marker at the marked location, we can investigate how the observer interprets the virtual marker to help with distant object spatial referencing.

In the future, we will also seek to build on these techniques in real-world wide-area AR applications. These point marking techniques can be particularly beneficial to AR applications requiring simple position data that is hard to obtain in critical conditions, such as AR tour planning architectural massing in the field (e.g., [133]). For example, by acquiring the position of a point of interest, an AR tour planning application can attach labels and tour guide information for the tourists to view and inspect.

Additionally, we can seek further improvements to these point marking techniques by combining computer vision algorithms. Our techniques require user input to identify the target from different locations, one by casting rays to the same target (VectorCloud), the other by picking the target from images (ImageRefinement). With the help of computer vision algorithms, we hope to design and build a marking technique that can automatically match the target and reduce user input during the process so that maybe the user only needs to
initially indicate the target or correct the algorithm when it fails. Moreover, we also plan on conducting user studies to assess the benefit of a semi-automated technique and compare it against VectorCloud and ImageRefinement. We expect to see improvements to efficiency and possible accuracy, depending on the performance of the feature matching algorithms.

7.2.4 ARCritique

One major limitation of the current ARCritique is scanning. We can use new mobile devices equipped with Light Detection and Ranging (LiDAR) sensors on their back to simplify the scanning process. Unlike using the front camera, using the LiDAR sensors allows the user to hold the phone with the screen facing them like they usually do and inspect the scanning result more easily. More importantly, we hope to test other model reconstruction methods to improve the scan quality. As participants commented, the currently scanned models were only suitable for the initial design stage where shapes and geometries of the mockups mattered. The application can benefit from more information from the mockup, including improved scanning accuracy and detailed colored texture from the original physical artifact. Thus, we need to improve our scanning algorithm to generate models with higher spatial resolution and incorporate photorealistic imagery to afford more design aspects. With improved scanning, we are also able to test ARCritique in more complex critique sessions where the discussion goes beyond the shape of the artifact. We can expect to see increased use of the referencing tools throughout the discussion as the users have more information from the scanned model.

Moreover, we can add other widely used spatial referencing interactions, such as hand gestures [77] to ARCritique. With newly added hand pose capture functionality ¹, it becomes possible to support hand tracking in ARCritique. We can design and create shared virtual

7.2. Future Work

hands of the collaborators in a critique session. In this way, the user can refer to different features of the model with various hand gestures and interact with the model to move it, rotate it, or even change its shape, as they can in the real world. We expect users to find barehanded referencing beneficial to the critique session. Hence, we plan to conduct a user experiment to evaluate the application in similar ecologically valid settings.

7.2.5 Future of Spatial Referencing in AR

Throughout the dissertation, we have explored different issues with current AR collaborative systems across the space dimension delineated by Johansen et al. [70]. As we argued in Chapter 1, obtaining an always up-to-date, accurate model from the physical environment can remain challenging in the foreseeable future, especially under difficult conditions like outdoor space or dynamically changing areas. However, obtaining a partial or low-detail environment model is possible under current technology, as demonstrated in RQ4. Hence, one potential research direction for future spatial referencing is to take advantage of such imperfect model information. Going beyond the model-free scenario, we face plenty of research opportunities to refine the techniques presented in this dissertation and explore new possibilities. For instance, an interesting question can be how hand models of different levels of detail affect co-located barehanded referencing. A basic minimal hand tracking may only generate the user’s palm position, and a complicated solution may generate a simple palm model with 6 DOF tracking. A referencing method can directly use this information to intersect virtual content to produce depth cues or other visual effects. We can conduct a controlled user study to investigate the minimal level of detail needed to support naturalistic pointing gestures with this straightforward approach.

More research opportunities can emerge from different combinations of the dimensions in
Table 7.1 that have not been explored yet. For instance, we can propose a research project exploring wide-area referencing with pointing rays, partial target model, and a combination of virtual and real world reference targets. To be more precise, with the support of an imperfect environment model, we can use a pointing ray technique to indicate the referenced object by presenting correct occlusion at the target location. Since there is no guarantee that the environment model is always reliable, we need to develop a method to automatically detect the incorrect environment model or allow the user to notify the AR system when a failure happens. Moreover, the system can utilize the marking technique to let the user mark a point of interest and incorporate the marking point into the known environment model to improve its quality.

Furthermore, we can further extend the studies by supporting more spatial referencing channels to achieve higher ecological validity. Similar to RQ4, where we allowed the participants to communicate verbally besides using the pointing ray and drawing tools, we can conduct research to investigate how the users can overcome issues with our presented techniques with added communication. For instance, we can study the benefit of adding verbal communication as a disambiguate solution to the Single Ray technique in the outdoor pointing ray experiment. As we have already seen in outdoor AR, the Single Ray technique can achieve over 80% accuracy. By letting the pointer talk to the observer, the pointer may help the observer understand the target correctly at the cost of communicating time. Then, the benefit of the Double Ray may be less favorable if the Single Ray with verbal communication can achieve comparable observer accuracy.

Last, future work needs to address the time dimension across various collaborative tasks. Our work has primarily focused on real-time groupware where collaborators participate in the task simultaneously. We did not consider scenarios where the collaborators were separated in time. For example, online courses are mainly asynchronous since the instructors pre-record lectures
to distribute to attendees. In an online course involving a 3D model, spatial referencing is still needed to guide students’ attention. Little is known about how the discrepancy in time may affect the performance, let alone a solution to address the issues. Hence, a systematic investigation of asynchronous use cases can be another future work direction. To start with, we designed the ARCritique application with a *load&save* function. The users can store their drawings on the model on their phones and load them back in a different critique session. We can further add reference information to the saved message so that the student can replay the critique session. For instance, instead of showing all previous drawings onto the model at once, the application can sequentially re-draw the augmentations, guiding the viewer’s attention to different regions of the artifact. A controlled user study can be conducted to study the impact of these two re-draw methods on the student’s experience of an asynchronous remote critique session.
Bibliography


[40] Austin Erickson, Kangsoo Kim, Gerd Bruder, and Gregory F Welch. Exploring the limitations of environment lighting on optical see-through head-mounted displays. In Symposium on Spatial User Interaction, pages 1–8, 2020.


[47] Steffen Gauglitz, Cha Lee, Matthew Turk, and Tobias Höllerer. Integrating the physical environment into mobile remote collaboration. In *Proceedings of the 14th


[57] Zhenyi He, Karl Toby Rosenberg, and Ken Perlin. Exploring configuration of mixed


[75] Hannes Kaufmann, Dieter Schmalstieg, and Michael Wagner. Construct3d: A virtual reality application for mathematics and geometry education. *Education and In-


[88] C. Lee, S. Bonebrake, D. A. Bowman, and T. Höllerer. The role of latency in the


[143] Christina Trepkowski, David Eibich, Jens Maiero, Alexander Marquardt, Ernst Kruijff, and Steven Feiner. The effect of narrow field of view and information density on visual


Appendices
Appendix A

User Study Materials

A.1 Study of Barehanded Referencing

A.1.1 User Consent Form
Title of research study: IRB 19-519 Object Reference in Collaborative Augmented Reality

Principal Investigator: Sang Won Lee sangwonlee@vt.edu (540)231-4857

Other study contact(s):
- Doug A. Bowman dbowman@vt.edu (540) 231-2058
- Yuan Li vli92@vt.edu (208) 807-6797
- Donghan Hu hudh0827@vt.edu (540) 613-9449
- Boyuan Wang boyuan@vt.edu

Key Information: The following is a short summary of this study to help you decide whether or not to be a part of this study. More detailed information is listed later on in this form.

We invite you to participate in this Augmented Reality (AR) study to help us testing different collaborative techniques. This work is hoping to make AR more usable and we deeply appreciate your participation.

Detailed Information: The following is more detailed information about this study in addition to the information listed above.

Who can I talk to?
If you have questions, concerns, or complaints, or think the research has hurt you, talk to the research team at vli92@vt.edu / (208) 807-6797.

This research has been reviewed and approved by the Virginia Tech Institutional Review Board (IRB). You may communicate with them at 540-231-3732 or irb@vt.edu if:

- You have questions about your rights as a research subject
- Your questions, concerns, or complaints are not being answered by the research team
- You cannot reach the research team
- You want to talk to someone besides the research team to provide feedback about this research

How many people will be studied?
We plan to include about 30 people in this research study.

What happens if I say yes, I want to be in this research?
The study will take place in the Usability Lab 133 (D, E, F) (102 McBryde Hall). and will take approximately 60 minutes. When you arrive, you will be greeted and asked to read and sign the informed consent form after your questions (if any) are answered. Then you will be asked to clarify you have normal vision. Next you will be provided with written or verbal instructions for the experiment, and familiarized with the lab and the equipment they will be using. You will wear an augmented reality (AR) headset such as the Microsoft Hololens. Using the devices, you will then complete a series of interaction tasks in AR and in real world, using one or more 3D interaction
Consent to Take Part in a Research Study

techniques. Tasks will involve physical movements including looking around the environment, pointing to objects, and manipulating virtual objects. We will use a camera to only capture your hand gestures. After each block of tasks, you will be interviewed by the investigator to gather qualitative feedback. Breaks will be given after each usability questionnaire. After all tasks are completed, you will be interviewed about the entire experience. The entire session will be audio recorded.

To thank you for your participation, we will give out $12 electronic gift card at the end of the study.

What happens if I say yes, but I change my mind later?
You can leave the research at any time, for any reason, and it will not be held against you.
If you decide to leave the research, contact the investigator so that the investigator can terminate current experiment session. Any data collected by far will be disregarded and will not be used for further study. There will be no follow-up procedures afterwards.

Withdrawing from the experiment does not affect receiving $12 gift card compensation.

Is there any way being in this study could be bad for me? (Detailed Risks)
Using AR technology can produce symptoms of sickness or discomfort in some users. These symptoms are usually mild, and may include dizziness, nausea, eye strain, headache, or disorientation. During tasks involving physical movement, there is also some risk that you could collide with obstacles in the physical environment.

You will be given the option to take a break or quit the experiment at any time. To mitigate the risk of sickness and discomfort, we will adjust the display properly, keep task sessions short, provide frequent breaks, and ask you after each set of tasks how you are feeling. To mitigate the risk of physical obstacles and cabling, we will clear the area of obstacles, show you where the boundaries of the space are.

What happens to the information collected for the research?
We will make every effort to limit the use and disclosure of your personal information, including research study, only to people who have a need to review this information. We cannot promise complete confidentiality. Organizations that may inspect and copy your information include the IRB, Human Research Protection Program, and other authorized representatives of Virginia Tech.

All data collected during this study will be done so anonymously and stored in a password protected computer. No names, contact information, or any other identifying information will be attached to your responses to an investigator’s questions or to your results from the provided tasks. At no time will the researchers release identifiable results of the study to anyone other than individuals working on the project without your written consent. However, you may be asked to perform tasks in a public space and bystanders could recognize you.

If identifiers are removed from your private information or samples that are collected during this research, that information or those samples could be used for future research studies or distributed to another investigator for future research studies without your additional informed consent.
Consent to Take Part in a Research Study

The results of this research study may be presented in summary form at conferences, in presentations, reports to the sponsor, academic papers, and as part of a thesis/dissertation.

Can I be removed from the research without my OK?
The person in charge of the research study or the sponsor can remove you from the research study without your approval. Possible reasons for removal include technical issues.

What else do I need to know?
We will not offer to share your individual test results with you. You may accept or decline these results.
Consent to Take Part in a Research Study

Signature Block for Capable Adult

Your signature documents your permission to take part in this research. We will provide you with a signed copy of this form for your records.

________________________________________________________________________
<table>
<thead>
<tr>
<th>Signature of subject</th>
<th>Date</th>
</tr>
</thead>
</table>

________________________________________________________________________
| Printed name of subject |

________________________________________________________________________
<table>
<thead>
<tr>
<th>Signature of person obtaining consent</th>
<th>Date</th>
</tr>
</thead>
</table>

________________________________________________________________________
| Printed name of person obtaining consent |
A.1.2 Questionnaire

Background Questionnaire
Collaborative AR Barehand Study

Start of Block: Default Question Block

Q1 Gender

- Male (1)
- Female (2)

Q2 Age

Q3 Do you wear glasses or contact lenses?

- Glasses (1)
- Contact Lenses (2)
- Neither (3)

Q4 Are you:

- Right-handed (1)
- Left-handed (2)
- Ambidextrous (3)
Q5 Occupation (if student, indicate graduate or undergraduate):

________________________________________________________________

Q6 If student, major/area of specialization

________________________________________________________________

Q7 Rate your fatigue level

<table>
<thead>
<tr>
<th>Not tired at all</th>
<th>Extremely tired</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>4</td>
</tr>
</tbody>
</table>

1 ()

Q8 Rate your experience with computers:

<table>
<thead>
<tr>
<th>Beginner</th>
<th>Advanced</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>4</td>
</tr>
</tbody>
</table>

1 ()
Q9 How often do you use computers for work:
- Not at all (1)
- Once a month (2)
- Once a week (3)
- Several times per week (4)
- Daily (5)

Q10 How often do you use computers for fun:
- Not at all (1)
- Once a month (2)
- Once a week (3)
- Several times per week (4)
- Daily (5)

Q11 How often do you play video games (consoles or computer)?

<table>
<thead>
<tr>
<th>Never</th>
<th>Everyday</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>4</td>
</tr>
</tbody>
</table>

Click to write Choice 1 ()
Q13 If you have watched any 3D videos, please rate your experience with them:
Bad (headaches, dizziness, etc)  Excellent
0  4

Q14 How many times have you tried virtual or augmented reality:
  O Never used (1)
  O Once or twice (2)
  O 3 to 10 times (3)
  O More than 10 times (4)

End of Block: Default Question Block

Start of Block: Experimenter Notes

Q15 Please Stop - Experimenter Notes

Q16 Today's Date
Q19 What is (or will be) the date?
Month (1)
Day (2)
Year (3)

▼ January (1) ... December ~ 31 ~ 2020 (28194)

Q17 Internal User ID

________________________________________________________________

Q18 Experimenter

☐ Yuan (1)

☐ Donghan (2)

☐ Other (3) ________________________________________________

Q19 Condition order

________________________________________________________________

Q26 User observation and interview notes

________________________________________________________________

End of Block: Experimenter Notes
Interview Questions
o As the pointer:
  ▪ How easy it is to point?
  ▪ Do you consider your collaborator to adjust your pointing gesture?
  ▪ What, if anything, is difficult about the pointing?
    ▪ If any, Could you propose any solution to overcome the obstacles?

o As the observer:
  ▪ How easy it is to identify the object
  ▪ Do your collaborator helps to adjust gesture?
  ▪ What, if anything, is difficult about the identifying?
    ▪ If any, Could you propose any solution to overcome the obstacles?
A.1.3 IRB Approval Letter
MEMORANDUM

DATE: September 27, 2019
TO: Sang Won Lee, Yuan Li, Douglas Andrew Bowman, Donghan Hu, Boyuan Wang
FROM: Virginia Tech Institutional Review Board (FWA00000572, expires January 29, 2021)

PROTOCOL TITLE: Object Reference in Collaborative Augmented Reality
IRB NUMBER: 19-519

Effective September 27, 2019, the Virginia Tech Institution Review Board (IRB) approved the Amendment request for the above-mentioned research protocol.

This approval provides permission to begin the human subject activities outlined in the IRB-approved protocol and supporting documents.

Plans to deviate from the approved protocol and/or supporting documents must be submitted to the IRB as an amendment request and approved by the IRB prior to the implementation of any changes, regardless of how minor, except where necessary to eliminate apparent immediate hazards to the subjects. Report within 5 business days to the IRB any injuries or other unanticipated or adverse events involving risks or harms to human research subjects or others.

All investigators (listed above) are required to comply with the researcher requirements outlined at: https://secure.research.vt.edu/external/irb/responsibilities.htm
(Please review responsibilities before beginning your research.)

PROTOCOL INFORMATION:
Approved As: Expedited, under 45 CFR 46.110 category(ies) 6,7
Protocol Approval Date: July 17, 2019
Progress Review Date: July 16, 2020

ASSOCIATED FUNDING:
The table on the following page indicates whether grant proposals are related to this protocol, and which of the listed proposals, if any, have been compared to this protocol, if required.
**SPECIAL INSTRUCTIONS:**

This amendment, submitted September 22, 2019, changes the research protocol by adding Boyuan Wang's information. Research personnel were updated. Boyuan Wang's information was also added to the recruitment materials and consent form. Data collection instruments were also changed by re-phrasing interview questions.

<table>
<thead>
<tr>
<th>Date*</th>
<th>OSP Number</th>
<th>Sponsor</th>
<th>Grant Comparison Conducted?</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*Date this proposal number was compared, assessed as not requiring comparison, or comparison information was revised.

If this protocol is to cover any other grant proposals, please contact the HRPP office (irb@vt.edu) immediately.
A.2 Study of Collaborative Gaze Ray in Simulated AR

A.2.1 User Consent Form
I. Purpose of this Research Project

This research project is intended to compare a variety of methods for interacting with virtual reality (VR) and augmented reality (AR) systems. This research will help us understand the best ways to interact with data and information in VR and AR, and produce design guidelines for VR/AR user interface design. The results of this study may appear in future publications and presentations building upon this research, though all results will be anonymous. Study participants should be over the age of 18 and have normal (corrected or uncorrected) eyesight.

II. Procedures

The study will take place in the Perform Studio (Moss Arts Center room 159), the Sandbox Studio (Moss Arts Center 160), the Cube (Moss Arts Center), or in a nearby outdoor area, and will take approximately 90 minutes for each participant.

Upon signing up, the participants will be asked if they have participated in previous related studies. When participants arrive, they will be greeted and asked to read and sign the informed consent form after their questions (if any) are answered. Then they will be asked to clarify they have normal vision and screened by a stereoscopic perception program. Next they will be provided with written or verbal instructions for the experiment, and familiarized with the lab and the equipment they will be using. Participants will wear either a virtual reality (VR) headset such as the HTC Vive or an augmented reality (AR) headset such as the Microsoft Hololens. Using these devices, participants will then complete a series of interaction tasks in VR or AR, using one or more 3D interaction techniques. Tasks will involve physical movements including looking around the environment, walking to targets, pointing to objects, and manipulating virtual objects. After each block of tasks, participants will be interviewed by the investigator to
gather qualitative feedback. Breaks will be given after each usability questionnaire. After all tasks are completed, participants will be interviewed about their experience.

III. Risks

Using VR and AR technology can produce symptoms of sickness or discomfort in some users. These symptoms are usually mild, and may include dizziness, nausea, eye strain, headache, or disorientation. During tasks involving physical movement, there is also some risk that you could collide with obstacles in the physical environment, or contact the physical cables connecting the display to the computer.

You will be given the option to take a break or quit the experiment at any time. To mitigate the risk of sickness and discomfort, we will adjust the display properly, keep task sessions short, provide frequent breaks, and ask you after each set of tasks how you are feeling. To mitigate the risk of physical obstacles and cabling, we will clear the area of obstacles, show you where the boundaries of the space are, display a virtual wall when you near a physical boundary, hold the cables to keep them out of your way, and warn you if you are nearing an obstacle.

IV. Benefits

The study will improve our understanding of 3D interaction design for VR and AR, so that effective user interfaces can be designed for real-world VR and AR applications. Study participants will benefit from exposure to state-of-the-art VR and AR technologies and techniques.

No promise or guarantee of benefits has been made to encourage you to participate.

V. Extent of Anonymity and Confidentiality

All data collected during this study will be done so anonymously. No names, contact information, or any other identifying information will be attached to your responses to an investigator’s questions or to your results from the provided tasks. At no time will the researchers release identifiable results of the study to anyone other than individuals working on the project without your written consent. However, you may be asked to perform tasks in a public space and bystanders could recognize you.

The Virginia Tech (VT) Institutional Review Board (IRB) may view the study’s data for auditing purposes. The IRB is responsible for the oversight of the protection of human subjects involved in research.

VI. Compensation

The participant will be paid a flat rate of $15 for taking part in this research. If the participant is paid, the compensation will be broken down as follows:

• It is a one-time payment
• It is paid at the end of participation
• Participant will get the same amount of payment if dropping out during the experiment
VII. Freedom to Withdraw

It is important for you to know that you are free to withdraw from this study at any time without penalty. You are free not to answer any questions that you choose or respond to what is being asked of you without penalty.

Please note that there may be circumstances under which the investigator may determine that a subject should not continue as a subject.

VIII. Questions or Concerns

Should you have any questions about this study, you may contact one of the research investigators whose contact information is included at the beginning of this document.

Should you have any questions or concerns about the study’s conduct or your rights as a research participant, or need to report a research-related injury or event, you may contact the Virginia Tech Institutional Review Board at irb@vt.edu or (540) 231-3732.

IX. Subject’s Consent

I have read the Consent Form and conditions of this project. I have had all my questions answered. I hereby acknowledge the above and give my voluntary consent:

_______________________________________________ Date__________
Subject signature

_______________________________________________
Subject printed name
A.2.2 Questionnaire

AR Ray Questionnaire
Usability Questionnaire
Participant #: 
Technique: 
Please circle the appropriate response for each question:

1. I think that I would like to use this technique frequently.
   - Strongly Disagree
   - Disagree
   - Neutral
   - Agree
   - Strongly Agree

2. I found the technique unnecessarily complex.
   - Strongly Disagree
   - Disagree
   - Neutral
   - Agree
   - Strongly Agree

3. I thought the technique was easy to use.
   - Strongly Disagree
   - Disagree
   - Neutral
   - Agree
   - Strongly Agree

4. I think that I would need the support of a technical person to be able to use this technique.
   - Strongly Disagree
   - Disagree
   - Neutral
   - Agree
   - Strongly Agree

5. I thought there was too much inconsistency in this technique.
   - Strongly Disagree
   - Disagree
   - Neutral
   - Agree
   - Strongly Agree

6. I would imagine that most people would learn to use this technique very quickly.
   - Strongly Disagree
   - Disagree
   - Neutral
   - Agree
   - Strongly Agree

7. I found the technique very cumbersome to use.
   - Strongly Disagree
   - Disagree
   - Neutral
   - Agree
   - Strongly Agree

8. I felt very confident using the technique.
   - Strongly Disagree
   - Disagree
   - Neutral
   - Agree
   - Strongly Agree
9. I needed to learn a lot of things before I could get going with this technique.

   Strongly Disagree  Disagree  Neutral  Agree  Strongly Agree

10. I was physically comfortable using this technique.

   Strongly Disagree  Disagree  Neutral  Agree  Strongly Agree
Interview Questions
View management and user interface optimization for wide-area mobile augmented reality

Interview topics/questions:
1. Which of the interaction techniques that you tried today did you prefer? Why?
2. Comment on the following aspects of the interaction techniques:
   a. Ease of learning
   b. Ease of use
   c. Naturalness
   d. Comfort
   e. Speed
   f. Accuracy
   g. Fun
3. Do you think these interfaces would be usable in environments with these characteristics? Why or why not?
   a. Dark
   b. Bright
   c. Outdoors
   d. Noisy
   e. Rainy/damp
A.2.3 IRB Approval Letter
Virginia Tech Institutional Review Board (FWA00000572, expires January 29, 2021)

PROTOCOL TITLE: View management and user interface optimization for wide-area mobile augmented reality

IRB NUMBER: 17-431

Effective February 7, 2019, the Virginia Tech Institution Review Board (IRB) approved the Amendment request for the above-mentioned research protocol.

This approval provides permission to begin the human subject activities outlined in the IRB-approved protocol and supporting documents.

Plans to deviate from the approved protocol and/or supporting documents must be submitted to the IRB as an amendment request and approved by the IRB prior to the implementation of any changes, regardless of how minor, except where necessary to eliminate apparent immediate hazards to the subjects. Report within 5 business days to the IRB any injuries or other unanticipated or adverse events involving risks or harms to human research subjects or others.

All investigators (listed above) are required to comply with the researcher requirements outlined at: https://secure.research.vt.edu/external/irb/responsibilities.htm

(Please review responsibilities before beginning your research.)

PROTOCOL INFORMATION:

Approved As: Expedited, under 45 CFR 46.110 category(ies) 7
Protocol Approval Date: May 1, 2018
Protocol Expiration Date: April 30, 2019
Continuing Review Due Date*: April 16, 2019

*Date a Continuing Review application is due to the IRB office if human subject activities covered under this protocol, including data analysis, are to continue beyond the Protocol Expiration Date.

ASSOCIATED FUNDING:

The table on the following page indicates whether grant proposals are related to this IRB protocol, and which of the listed proposals, if any, have been compared to this IRB protocol, if required.
**IRB SPECIAL INSTRUCTIONS:**

This amendment, submitted January 29, 2019, changes the research protocol by extending the study time and adding compensation. These changes were also reflected in the revised recruitment materials and consent form.

<table>
<thead>
<tr>
<th>Date*</th>
<th>OSP Number</th>
<th>Sponsor</th>
<th>Grant Comparison Conducted?</th>
</tr>
</thead>
<tbody>
<tr>
<td>04/26/2017</td>
<td>PET47EWN</td>
<td>University of California, Santa Barbara (Title: View management and user interface optimization for wide-area mobile augmented real)</td>
<td>Not required (VT is not primary inst.)</td>
</tr>
</tbody>
</table>

* Date this proposal number was compared, assessed as not requiring comparison, or comparison information was revised.

If this IRB protocol is to cover any other grant proposals, please contact the HRPP office (irb@vt.edu) immediately.
A.3 Study of Collaborative Gaze Ray in Outdoor AR

A.3.1 User Consent Form
Title of research study: *IRB 19-777 Collaborative Gaze Ray in Augmented Reality*

**Principal Investigator:** Doug Bowman  
dbowman@vt.edu  
(540)231-2058

**Other study contact(s):**

- Yuan Li  
yli92@vt.edu  
(208) 807-6797
- Feiyu Lu  
feiyulu@vt.edu  
(540) 257-4562
- Lee Lisle  
llisle@vt.edu  
(540) 840-0569
- Ibrahim Tahmid  
iatahmid@vt.edu  
(540) 998-9675

**Key Information:** The following is a short summary of this study to help you decide whether or not to be a part of this study. More detailed information is listed later on in this form.

**Why am I being invited to take part in a research study?**

We invite you to take part in a research study because you are a Virginia Tech student who is 18 years old or older and has normal or corrected (glasses or contacts) vision.

**What should I know about being in a research study?**

- Someone will explain this research study to you
- Whether or not you take part is up to you
- You can choose not to participate
- You can agree to take part and later change your mind
- Your decision will not be held against you
- You can ask all the questions you want before you decide

**What should I know about this research study?**

This research project is intended to compare a variety of methods for interacting with Augmented Reality (AR) system. This research will help us understand the best ways to interact with remote human users in collaborative AR and produce design guidelines for AR user interface design. The results of this study may appear in future publications and presentations building upon this research, through all results will be presented in summary form.

This study involves two participants in one 2-hour experiment session, which can be roughly split into the following phases:

- Pre-study phase: we will introduce you to the area, gather general demographical information, introduce the concept of this research and the tools we will be using.
- Tutorial phase: we will present the software and how to use the software during the study.
- Study phase: we will ask you to perform the same task using the software shown in Tutorial phase.
- Post-study phase: we will perform an interview on your experience in the study phase.
Consent to Take Part in a Research Study

With many AR tasks, there is a chance that you will have “simulator sickness”, which is a kind of motion sickness. Otherwise, there are minimal risks during this study. While there are no personal benefits from this study, the research we perform may improve future AR collaborative software. If at any time during the study you wish to stop your participation, please let us know and we will terminate the study immediately with no repercussions.

Detailed Information: The following is more detailed information about this study in addition to the information listed above.

Who can I talk to?

If you have questions, concerns, or complaints, or think the research has hurt you, talk to the research team at ylj92@vt.edu / (208) 807-6797.

This research has been reviewed and approved by the Virginia Tech Institutional Review Board (IRB). You may communicate with them at 540-231-3732 or irb@vt.edu if:

- You have questions about your rights as a research subject
- Your questions, concerns, or complaints are not being answered by the research team
- You cannot reach the research team
- You want to talk to someone besides the research team to provide feedback about this research

How many people will be studied?

We plan to include about 30 people in this research study.

What happens if I say yes, I want to be in this research?

The study will begin in Sandbox Studio in Moss Art Center and then proceed to the backdoor of Moss Art Center near Turner Street. It will take approximately 2 hours. When you arrive at Sandbox, you will be greeted and asked to read and sign the informed consent form after their questions (if any) are answered. We will ask you to complete a Qualtrics questionnaire to help us gather some basic demographic information. Then you will be provided with written or verbal instructions for the experiment, and familiarized with the equipment you will be using. Then you will be brought to the outdoor space to complete collaborative tasks. You will wear an AR headset (Microsoft HoloLens). Using the devices, you will then complete a series of interaction tasks. The tasks will involve physical movements including looking around the environment, interacting with UIs that define AR rays in the headset. After each block of tasks, you will be invited to complete NASA task load index questionnaire and a usability questionnaire on smartphones. Breaks will be given after each questionnaire. After all tasks are completed, you will take a user interview about your experience. At the end of your participation, you will receive a $30 Amazon gift card as a gratitude for your time.

What happens if I say yes, but I change my mind later?

You can leave the research at any time, for any reason, and it will not be held against you.

If you decide to leave the research, contact the investigator so that the investigator can terminate current experiment session. Any data collected by far will be disregarded and will not be used for further study. There will be no follow-up procedures afterwards.
Consent to Take Part in a Research Study

Is there any way being in this study could be bad for me? (Detailed Risks)
Using AR technology can produce symptoms of sickness or discomfort in some users. These symptoms are usually mild, and may include dizziness, nausea, eye strain, headache, or disorientation. During tasks involving physical movement, there is also some risk that participants will collide with obstacles in the physical environment.

You will be given the option to take a break or quit the experiment at any time. To mitigate the risk of sickness and discomfort, we will adjust the display properly, keep task sessions short, provide frequent breaks, and ask you after each set of tasks how you are feeling. To mitigate the risk of physical obstacles and cabling, we will clear the area of obstacles, show you where the boundaries of the space are.

What happens to the information collected for the research?
We will make every effort to limit the use and disclosure of your personal information, including research study, only to people who have a need to review this information. We cannot promise complete confidentiality. Organizations that may inspect and copy your information include the IRB, Human Research Protection Program, and other authorized representatives of Virginia Tech.

All data collected during this study will be done so anonymously and stored in a password protected computer. No names, contact information, or any other identifying information will be attached to your responses to an investigator’s questions or to your results from the provided tasks. At no time will the researchers release identifiable results of the study to anyone other than individuals working on the project without your written consent. However, you may be asked to perform tasks in a public space and bystanders could recognize you.

If identifiers are removed from your private information or samples that are collected during this research, that information or those samples could be used for future research studies or distributed to another investigator for future research studies without your additional informed consent.

The results of this research study may be presented in summary form at conferences, in presentations, reports to the sponsor, academic papers, and as part of a thesis/dissertation.

Can I be removed from the research without my OK?
The person in charge of the research study or the sponsor can remove you from the research study without your approval. Possible reasons for removal include technical issues.

What else do I need to know?
We will not offer to share your individual test results with you. You may accept or decline these results.
Consent to Take Part in a Research Study

Signature Block for Capable Adult

Your signature documents your permission to take part in this research. We will provide you with a signed copy of this form for your records.

________________________  ______________________
Signature of subject       Date

________________________
Printed name of subject

________________________  ______________________
Signature of person obtaining consent       Date

________________________
Printed name of person obtaining consent
A.3.2 Questionnaire

Pointer NASA TLX Questionnaire
Q1 What technique did you use?

- Single Ray  (1)
- Double Ray  (2)

Q2 Mental Demand
How much mental and perceptual activity was required (for example, thinking, deciding, calculating, remembering, looking, searching, etc)? Was the task easy or demanding, simple or complex, forgiving or exacting?

Low
High

0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20
Q3 Physical Demand
How much physical activity was required (for example, pushing, pulling, turning, controlling, activating, etc)? Was the task easy or demanding, slow or brisk, slack or strenuous, restful or laborious?

<table>
<thead>
<tr>
<th>Low</th>
<th>High</th>
</tr>
</thead>
<tbody>
<tr>
<td>0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20</td>
<td></td>
</tr>
</tbody>
</table>

Q4 Temporal Demand
How much time pressure did you feel due to the rate or pace at which the tasks of task elements occurred? Was the pace slow and leisurely or rapid and frantic?

<table>
<thead>
<tr>
<th>Low</th>
<th>High</th>
</tr>
</thead>
<tbody>
<tr>
<td>0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20</td>
<td></td>
</tr>
</tbody>
</table>

Q5 Performance
How successful do you think you were in accomplishing the goals of the task set by the experimenter (or yourself)? How satisfied were you with your performance in accomplishing these goals?

<table>
<thead>
<tr>
<th>Poor</th>
<th>Good</th>
</tr>
</thead>
<tbody>
<tr>
<td>0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20</td>
<td></td>
</tr>
</tbody>
</table>
Q6 Effort
How hard did you have to work (mentally and physically) to accomplish your level of performance?

<table>
<thead>
<tr>
<th>Low</th>
<th>High</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>10</td>
</tr>
</tbody>
</table>

1 ()

Q7 Frustration Level
How insecure, discouraged, irritated, stressed, and annoyed versus secure, gratified, content, relaxed, and complacent did you feel during the task?

<table>
<thead>
<tr>
<th>Low</th>
<th>High</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>10</td>
</tr>
</tbody>
</table>

1 ()

End of Block: Default Question Block
Start of Block: Block 2

Q54 What technique did you use?

- Single Ray (1)
- Double Ray (2)
### Q55 Mental Demand
How much mental and perceptual activity was required (for example, thinking, deciding, calculating, remembering, looking, searching, etc)? Was the task easy or demanding, simple or complex, forgiving or exacting?

<table>
<thead>
<tr>
<th>Low</th>
<th>High</th>
</tr>
</thead>
<tbody>
<tr>
<td>0 1 2 3 4 5 6 7 8 9</td>
<td>10 11 12 13 14 15 16 17 18 19 20</td>
</tr>
</tbody>
</table>

1 ()

---

### Q56 Physical Demand
How much physical activity was required (for example, pushing, pulling, turning, controlling, activating, etc)? Was the task easy or demanding, slow or brisk, slack or strenuous, restful or laborious?

<table>
<thead>
<tr>
<th>Low</th>
<th>High</th>
</tr>
</thead>
<tbody>
<tr>
<td>0 1 2 3 4 5 6 7 8 9</td>
<td>10 11 12 13 14 15 16 17 18 19 20</td>
</tr>
</tbody>
</table>

1 ()

---

### Q57 Temporal Demand
How much time pressure did you feel due to the rate or pace at which the tasks of task elements occurred? Was the pace slow and leisurely or rapid and frantic?

<table>
<thead>
<tr>
<th>Low</th>
<th>High</th>
</tr>
</thead>
<tbody>
<tr>
<td>0 1 2 3 4 5 6 7 8 9</td>
<td>10 11 12 13 14 15 16 17 18 19 20</td>
</tr>
</tbody>
</table>

1 ()
Q58 Performance
How successful do you think you were in accomplishing the goals of the task set by the experimenter (or yourself)? How satisfied were you with your performance in accomplishing these goals?

<table>
<thead>
<tr>
<th>Poor</th>
<th>Good</th>
</tr>
</thead>
<tbody>
<tr>
<td>0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20</td>
<td></td>
</tr>
</tbody>
</table>

1 ()

Q59 Effort
How hard did you have to work (mentally and physically) to accomplish your level of performance?

<table>
<thead>
<tr>
<th>Low</th>
<th>High</th>
</tr>
</thead>
<tbody>
<tr>
<td>0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20</td>
<td></td>
</tr>
</tbody>
</table>

1 ()

Q60 Frustration Level
How insecure, discouraged, irritated, stressed, and annoyed versus secure, gratified, content, relaxed, and complacent did you feel during the task?

<table>
<thead>
<tr>
<th>Low</th>
<th>High</th>
</tr>
</thead>
<tbody>
<tr>
<td>0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20</td>
<td></td>
</tr>
</tbody>
</table>

1 ()
Observer Modified SUS Questionnaire
Observer_Modified_SUS

Start of Block: Block 1

Q36 Participant ID

Q37 Experiment Date

Q38 Participant Location - Gate or Far-end

End of Block: Block 1

Start of Block: Default Question Block

Q1 What technique did you use?

- Single Ray (1)
- Single Ray with Parallel Bars (2)
- Double Ray (3)
- Double Ray with Parallel Bars (4)
Q2 I would imagine that most people would learn to use this ray visualization technique very quickly.

Strongly disagree  Strongly Agree
0 1 2 3 4 5 6 7 8 9 1011121314151617181920

\[1\]

Q39 I needed to learn a lot of things before I could get going with this ray visualization technique.

Strongly disagree  Strongly Agree
0 1 2 3 4 5 6 7 8 9 1011121314151617181920

\[1\]

Q40 I felt very confident using the ray visualization technique to identify the target.

Strongly disagree  Strongly Agree
0 1 2 3 4 5 6 7 8 9 1011121314151617181920

\[1\]

Q41 I found the ray visualization technique unnecessarily complex.

Strongly disagree  Strongly Agree
0 1 2 3 4 5 6 7 8 9 1011121314151617181920
Q42 I thought the ray visualization technique was easy to understand.

<table>
<thead>
<tr>
<th>Strongly disagree</th>
<th>Strongly Agree</th>
</tr>
</thead>
<tbody>
<tr>
<td>0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20</td>
<td></td>
</tr>
</tbody>
</table>

Q43 I felt that I spent a lot of mental effort to understand the ray visualization technique.

<table>
<thead>
<tr>
<th>Strongly disagree</th>
<th>Strongly Agree</th>
</tr>
</thead>
<tbody>
<tr>
<td>0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20</td>
<td></td>
</tr>
</tbody>
</table>

End of Block: Default Question Block

Start of Block: Block 2

Q65 What technique did you use?

- Single Ray (1)
- Single Ray with Parallel Bars (2)
- Double Ray (3)
- Double Ray with Parallel Bars (4)
Q66 I would imagine that most people would learn to use this ray visualization technique very quickly.

Strongly disagree Strongly Agree
0 1 2 3 4 5 6 7 8 9 1011121314151617181920

1 ()

Q67 I needed to learn a lot of things before I could get going with this ray visualization technique.

Strongly disagree Strongly Agree
0 1 2 3 4 5 6 7 8 9 1011121314151617181920

1 ()

Q68 I felt very confident using the ray visualization technique to identify the target.

Strongly disagree Strongly Agree
0 1 2 3 4 5 6 7 8 9 1011121314151617181920

1 ()

Q69 I found the ray visualization technique unnecessarily complex.

Strongly disagree Strongly Agree
Q70 I thought the ray visualization technique was easy to understand.

Strongly disagree  Strongly Agree

Q71 I felt that I spent a lot of mental effort to understand the ray visualization technique.

Strongly disagree  Strongly Agree
Q72 What technique did you use?

- Single Ray (1)
- Single Ray with Parallel Bars (2)
- Double Ray (3)
- Double Ray with Parallel Bars (4)

Q73 I would imagine that most people would learn to use this ray visualization technique very quickly.

<table>
<thead>
<tr>
<th>Strongly disagree</th>
<th>Strongly Agree</th>
</tr>
</thead>
<tbody>
<tr>
<td>0 1 2 3 4 5 6 7 8 9 1011121314151617181920</td>
<td></td>
</tr>
</tbody>
</table>

1 ()

Q74 I needed to learn a lot of things before I could get going with this ray visualization technique.

<table>
<thead>
<tr>
<th>Strongly disagree</th>
<th>Strongly Agree</th>
</tr>
</thead>
<tbody>
<tr>
<td>0 1 2 3 4 5 6 7 8 9 1011121314151617181920</td>
<td></td>
</tr>
</tbody>
</table>

1 ()

Q75 I felt very confident using the ray visualization technique to identify the target.

<table>
<thead>
<tr>
<th>Strongly disagree</th>
<th>Strongly Agree</th>
</tr>
</thead>
<tbody>
<tr>
<td>0 1 2 3 4 5 6 7 8 9 1011121314151617181920</td>
<td></td>
</tr>
</tbody>
</table>
Q76 I found the ray visualization technique unnecessarily complex.

<table>
<thead>
<tr>
<th>Strongly disagree</th>
<th>Strongly Agree</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>1</td>
</tr>
</tbody>
</table>

Q77 I thought the ray visualization technique was easy to understand.

<table>
<thead>
<tr>
<th>Strongly disagree</th>
<th>Strongly Agree</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>1</td>
</tr>
</tbody>
</table>

Q78 I felt that I spent a lot of mental effort to understand the ray visualization technique.

<table>
<thead>
<tr>
<th>Strongly disagree</th>
<th>Strongly Agree</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>1</td>
</tr>
</tbody>
</table>

End of Block: Block 3
Start of Block: Block 4

Q79 What technique did you use?

- Single Ray (1)
- Single Ray with Parallel Bars (2)
- Double Ray (3)
- Double Ray with Parallel Bars (4)

Q80 I would imagine that most people would learn to use this ray visualization technique very quickly.

Strongly disagree  Strongly Agree

0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20

1 ()

Q81 I needed to learn a lot of things before I could get going with this ray visualization technique.

Strongly disagree  Strongly Agree

0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20

1 ()
Q82 I felt very confident using the ray visualization technique to identify the target.

<table>
<thead>
<tr>
<th>Strongly disagree</th>
<th>Strongly Agree</th>
</tr>
</thead>
<tbody>
<tr>
<td>0 1 2 3 4 5 6 7 8 9 10</td>
<td>11 12 13 14 15 16 17 18 19 20</td>
</tr>
</tbody>
</table>

1 ( )

Q83 I found the ray visualization technique unnecessarily complex.

<table>
<thead>
<tr>
<th>Strongly disagree</th>
<th>Strongly Agree</th>
</tr>
</thead>
<tbody>
<tr>
<td>0 1 2 3 4 5 6 7 8 9 10</td>
<td>11 12 13 14 15 16 17 18 19 20</td>
</tr>
</tbody>
</table>

1 ( )

Q84 I thought the ray visualization technique was easy to understand.

<table>
<thead>
<tr>
<th>Strongly disagree</th>
<th>Strongly Agree</th>
</tr>
</thead>
<tbody>
<tr>
<td>0 1 2 3 4 5 6 7 8 9 10</td>
<td>11 12 13 14 15 16 17 18 19 20</td>
</tr>
</tbody>
</table>

1 ( )

Q85 I felt that I spent a lot of mental effort to understand the ray visualization technique.

<table>
<thead>
<tr>
<th>Strongly disagree</th>
<th>Strongly Agree</th>
</tr>
</thead>
<tbody>
<tr>
<td>0 1 2 3 4 5 6 7 8 9 10</td>
<td>11 12 13 14 15 16 17 18 19 20</td>
</tr>
</tbody>
</table>

1 ( )
Interview Questions
As a pointer, please rank the two pointing techniques based on your preference.

What makes you rank XXX as your favorite technique? What makes YYY the least favorite technique?

(While one experimenter is asking above questions, the other can take the users’ feedback from Qualtrics and calculate NASA TLX scores)

So you think technique XXX has the most workload, could you explain why you think that way? So you think technique XXX has the least workload, could you explain why you think that way?

Do you think your pointing rays were accurate for your collaborator?

As an observer, please rank the four techniques based on your preference. Why?

What makes you rank XXX as your favorite technique? What makes YYY come 2nd on your preference list? What makes ZZZ the 3rd on that list? What makes XYZ the least favorite technique?

(While one experimenter is asking above questions, the other can take the users’ feedback from Qualtrics and calculate observer subjective scores)

So you think technique XXX is the most usable, could you explain why you think that way? So you think technique XXX is the least usable, could you explain why you think that way?
Do you think your partner's rays were accurate enough for you to make a good judgement about the target?
A.3.3 IRB Approval Letter
MEMORANDUM

DATE: September 14, 2021
TO: Douglas Andrew Bowman, Yuan Li, Lee Lisle, Feiyu Lu, Ibrahim Asadullah Tahmid
FROM: Virginia Tech Institutional Review Board (FWA00000572)
PROTOCOL TITLE: Collaborative Gaze Ray in Augmented Reality
IRB NUMBER: 19-777

Effective September 14, 2021, the Virginia Tech Institution Review Board (IRB) approved the Amendment request for the above-mentioned research protocol.

This approval provides permission to begin the human subject activities outlined in the IRB-approved protocol and supporting documents.

Plans to deviate from the approved protocol and/or supporting documents must be submitted to the IRB as an amendment request and approved by the IRB prior to the implementation of any changes, regardless of how minor, except where necessary to eliminate apparent immediate hazards to the subjects. Report within 5 business days to the IRB any injuries or other unanticipated or adverse events involving risks or harms to human research subjects or others.

All investigators (listed above) are required to comply with the researcher requirements outlined at: https://secure.research.vt.edu/external/irb/responsibilities.htm

(Please review responsibilities before beginning your research.)

PROTOCOL INFORMATION:

Approved As: Expedited, under 45 CFR 46.110 category(ies) 4,6,7
Protocol Approval Date: October 9, 2019
Progress Review Date: October 8, 2022

ASSOCIATED FUNDING:

The table on the following page indicates whether grant proposals are related to this protocol, and which of the listed proposals, if any, have been compared to this protocol, if required.
SPECIAL INSTRUCTIONS:

This amendment, submitted September 6, 2021, updates research protocol, recruitment materials, and consent forms to update the experiment duration from 90 minutes to 2 hours and to increase compensation from $20 to $30 in Amazon gift card.

<table>
<thead>
<tr>
<th>Date*</th>
<th>OSP Number</th>
<th>Sponsor</th>
<th>Grant Comparison Conducted?</th>
</tr>
</thead>
<tbody>
<tr>
<td>09/11/2019</td>
<td>PET47EVN</td>
<td>University of California, Santa Barbara (Title: View management and user interface optimization for wide-area mobile augmented real)</td>
<td>Compared on 09/26/2019</td>
</tr>
</tbody>
</table>

* Date this proposal number was compared, assessed as not requiring comparison, or comparison information was revised.

If this protocol is to cover any other grant proposals, please contact the HRPP office (irb@vt.edu) immediately.
A.4 Study of Point Marking in Simulated AR

A.4.1 User Consent Form
I. Purpose of this Research Project

This research project is intended to compare a variety of methods for interacting with virtual reality (VR) and augmented reality (AR) systems. This research will help us understand the best ways to interact with data and information in VR and AR, and produce design guidelines for VR/AR user interface design. The results of this study may appear in future publications and presentations building upon this research, though all results will be anonymous. Study participants should be over the age of 18 and have normal (corrected or uncorrected) eyesight.

II. Procedures

The study will take place in the Perform Studio (Moss Arts Center room 159), the Cube (Moss Arts Center), or a nearby outdoor area, and will take approximately 60 minutes for each participant.

First, you will be provided with written or verbal instructions for the experiment, and familiarized with the lab and the equipment you will be using. You will wear either a virtual reality (VR) headset such as the HTC Vive or an augmented reality (AR) headset such as the Microsoft Hololens. Using these devices, you will then complete a series of interaction tasks in VR or AR, using one or more 3D interaction techniques. Tasks will involve physical movements including looking around the environment, walking to targets, pointing to objects, and manipulating virtual objects. Short (>5 minutes) usability questionnaires will be given after each block of tasks. Breaks will be given after each usability questionnaire. After all tasks are completed, you will be interviewed about your experience.

III. Risks

Using VR and AR technology can produce symptoms of sickness or discomfort in some users. These symptoms are usually mild, and may include dizziness, nausea, eye strain, headache,
or disorientation. During tasks involving physical movement, there is also some risk that you could collide with obstacles in the physical environment, or contact the physical cables connecting the display to the computer.

You will be given the option to take a break or quit the experiment at any time. To mitigate the risk of sickness and discomfort, we will adjust the display properly, keep task sessions short, provide frequent breaks, and ask you after each set of tasks how you are feeling. To mitigate the risk of physical obstacles and cabling, we will clear the area of obstacles, show you where the boundaries of the space are, display a virtual wall when you near a physical boundary, hold the cables to keep them out of your way, and warn you if you are nearing an obstacle.

IV. Benefits

The study will improve our understanding of 3D interaction design for VR and AR, so that effective user interfaces can be designed for real-world VR and AR applications. Study participants will benefit from exposure to state-of-the-art VR and AR technologies and techniques.

No promise or guarantee of benefits has been made to encourage you to participate.

V. Extent of Anonymity and Confidentiality

All data collected during this study will be done so anonymously. No names, contact information, or any other identifying information will be attached to your responses to an investigator’s questions or to your results from the provided tasks. At no time will the researchers release identifiable results of the study to anyone other than individuals working on the project without your written consent. However, you may be asked to perform tasks in a public space and bystanders could recognize you.

The Virginia Tech (VT) Institutional Review Board (IRB) may view the study’s data for auditing purposes. The IRB is responsible for the oversight of the protection of human subjects involved in research.

VI. Compensation

Participating in this study will provide you with no compensation, monetarily or otherwise. All subjects in the study do so of their own free will and a desire to further computing technologies.

VII. Freedom to Withdraw

It is important for you to know that you are free to withdraw from this study at any time without penalty. You are free not to answer any questions that you choose or respond to what is being asked of you without penalty.

Please note that there may be circumstances under which the investigator may determine that a subject should not continue as a subject.
VIII. Questions or Concerns

Should you have any questions about this study, you may contact one of the research investigators whose contact information is included at the beginning of this document.

Should you have any questions or concerns about the study’s conduct or your rights as a research participant, or need to report a research-related injury or event, you may contact the Virginia Tech Institutional Review Board at irb@vt.edu or (540) 231-3732.

IX. Subject’s Consent

I have read the Consent Form and conditions of this project. I have had all my questions answered. I hereby acknowledge the above and give my voluntary consent:

_______________________________________________ Date__________
Subject signature

_______________________________________________
Subject printed name
A.4.2 Questionnaire

Modified SUS Questionnaire
Usability Questionnaire
Participant #:
Technique:
Please circle the appropriate response for each question:

1. I think that I would like to use this technique frequently.
   
   Strongly Disagree    Disagree    Neutral    Agree    Strongly Agree

2. I found the technique unnecessarily complex.
   
   Strongly Disagree    Disagree    Neutral    Agree    Strongly Agree

3. I thought the technique was easy to use.
   
   Strongly Disagree    Disagree    Neutral    Agree    Strongly Agree

4. I think that I would need the support of a technical person to be able to use this technique.
   
   Strongly Disagree    Disagree    Neutral    Agree    Strongly Agree

5. I thought there was too much inconsistency in this technique.
   
   Strongly Disagree    Disagree    Neutral    Agree    Strongly Agree

6. I would imagine that most people would learn to use this technique very quickly.
   
   Strongly Disagree    Disagree    Neutral    Agree    Strongly Agree

7. I found the technique very cumbersome to use.
   
   Strongly Disagree    Disagree    Neutral    Agree    Strongly Agree

8. I felt very confident using the technique.
   
   Strongly Disagree    Disagree    Neutral    Agree    Strongly Agree
9. I needed to learn a lot of things before I could get going with this technique.

   Strongly Disagree    Disagree    Neutral    Agree    Strongly Agree

10. I was physically comfortable using this technique.

   Strongly Disagree    Disagree    Neutral    Agree    Strongly Agree
Interview Questions
View management and user interface optimization for wide-area mobile augmented reality

Interview topics/questions:
1. Which of the interaction techniques that you tried today did you prefer? Why?
2. Comment on the following aspects of the interaction techniques:
   a. Ease of learning
   b. Ease of use
   c. Naturalness
   d. Comfort
   e. Speed
   f. Accuracy
   g. Fun
3. Do you think these interfaces would be usable in environments with these characteristics? Why or why not?
   a. Dark
   b. Bright
   c. Outdoors
   d. Noisy
   e. Rainy/damp
A.4.3 IRB Approval Letter
MEMORANDUM

DATE: February 16, 2018
TO: Douglas Andrew Bowman, Wallace Santos Lages, Yuan Li, Lee Lisle
FROM: Virginia Tech Institutional Review Board (FWA00000572, expires January 29, 2021)

PROTOCOL TITLE: View management and user interface optimization for wide-area mobile augmented reality
IRB NUMBER: 17-431

Effective February 16, 2018, the Virginia Tech Institution Review Board (IRB) approved the Amendment request for the above-mentioned research protocol.

This approval provides permission to begin the human subject activities outlined in the IRB-approved protocol and supporting documents.

Plans to deviate from the approved protocol and/or supporting documents must be submitted to the IRB as an amendment request and approved by the IRB prior to the implementation of any changes, regardless of how minor, except where necessary to eliminate apparent immediate hazards to the subjects. Report within 5 business days to the IRB any injuries or other unanticipated or adverse events involving risks or harms to human research subjects or others.

All investigators (listed above) are required to comply with the researcher requirements outlined at: http://www.irb.vt.edu/pages/responsibilities.htm
(Please review responsibilities before the commencement of your research.)

PROTOCOL INFORMATION:
Approved As: Expedited, under 45 CFR 46.110 category(ies) 7
Protocol Approval Date: May 1, 2017
Protocol Expiration Date: April 30, 2018
Continuing Review Due Date*: April 16, 2018

*Date a Continuing Review application is due to the IRB office if human subject activities covered under this protocol, including data analysis, are to continue beyond the Protocol Expiration Date.

FEDERALLY FUNDED RESEARCH REQUIREMENTS:
Per federal regulations, 45 CFR 46.103(f), the IRB is required to compare all federally funded grant proposals/work statements to the IRB protocol(s) which cover the human research activities included in the proposal/ work statement before funds are released. Note that this requirement does not apply to Exempt and Interim IRB protocols, or grants for which VT is not the primary awardee.

The table on the following page indicates whether grant proposals are related to this IRB protocol, and which of the listed proposals, if any, have been compared to this IRB protocol, if required.
<table>
<thead>
<tr>
<th>Date*</th>
<th>OSP Number</th>
<th>Sponsor</th>
<th>Grant Comparison Conducted?</th>
</tr>
</thead>
</table>
| 04/26/2017 | PET47EVN  | University of California, Santa Barbara  
(Title: View management and user interface optimization for wide-area mobile augmented real) | Not required (VT is not primary inst.) |

* Date this proposal number was compared, assessed as not requiring comparison, or comparison information was revised.

If this IRB protocol is to cover any other grant proposals, please contact the IRB office (irbadmin@vt.edu) immediately.
A.5 Study of Point Marking in Outdoor AR

A.5.1 User Consent Form
Title of Project: **View management and user interface optimization for wide-area mobile augmented reality**

**Investigator(s):**

<table>
<thead>
<tr>
<th>Name</th>
<th>E-mail / Phone number</th>
</tr>
</thead>
<tbody>
<tr>
<td>Wallace Lages</td>
<td><a href="mailto:wlages@vt.edu">wlages@vt.edu</a> (540-750-9380)</td>
</tr>
<tr>
<td>Douglas Bowman</td>
<td><a href="mailto:dbowman@vt.edu">dbowman@vt.edu</a></td>
</tr>
<tr>
<td>Yuan Li</td>
<td><a href="mailto:yli92@vt.edu">yli92@vt.edu</a></td>
</tr>
<tr>
<td>Lee Lisle</td>
<td><a href="mailto:lllisle@vt.edu">lllisle@vt.edu</a></td>
</tr>
</tbody>
</table>

**I. Purpose of this Research Project**

This research project is intended to compare a variety of methods for interacting with virtual reality (VR) and augmented reality (AR) systems. This research will help us understand the best ways to interact with data and information in VR and AR, and produce design guidelines for VR/AR user interface design. The results of this study may appear in future publications and presentations building upon this research, though all results will be anonymous. Study participants should be over the age of 18 and have normal (corrected or uncorrected) eyesight.

**II. Procedures**

The study will take place in the Perform Studio (Moss Arts Center room 159), the Cube (Moss Arts Center), or a nearby outdoor area, and will take approximately 60 minutes for each participant.

First, you will be provided with written or verbal instructions for the experiment, and familiarized with the lab and the equipment you will be using. You will wear either a virtual reality (VR) headset such as the HTC Vive or an augmented reality (AR) headset such as the Microsoft Hololens. Using these devices, you will then complete a series of interaction tasks in VR or AR, using one or more 3D interaction techniques. Tasks will involve physical movements including looking around the environment, walking to targets, pointing to objects, and manipulating virtual objects. Short (>5 minutes) usability questionnaires will be given after each block of tasks. Breaks will be given after each usability questionnaire. After all tasks are completed, you will be interviewed about your experience.

**III. Risks**

Using VR and AR technology can produce symptoms of sickness or discomfort in some users. These symptoms are usually mild, and may include dizziness, nausea, eye strain, headache,
or disorientation. During tasks involving physical movement, there is also some risk that you could collide with obstacles in the physical environment, or contact the physical cables connecting the display to the computer.

You will be given the option to take a break or quit the experiment at any time. To mitigate the risk of sickness and discomfort, we will adjust the display properly, keep task sessions short, provide frequent breaks, and ask you after each set of tasks how you are feeling. To mitigate the risk of physical obstacles and cabling, we will clear the area of obstacles, show you where the boundaries of the space are, display a virtual wall when you near a physical boundary, hold the cables to keep them out of your way, and warn you if you are nearing an obstacle.

**IV. Benefits**

The study will improve our understanding of 3D interaction design for VR and AR, so that effective user interfaces can be designed for real-world VR and AR applications. Study participants will benefit from exposure to state-of-the-art VR and AR technologies and techniques.

No promise or guarantee of benefits has been made to encourage you to participate.

**V. Extent of Anonymity and Confidentiality**

All data collected during this study will be done so anonymously. No names, contact information, or any other identifying information will be attached to your responses to an investigator’s questions or to your results from the provided tasks. At no time will the researchers release identifiable results of the study to anyone other than individuals working on the project without your written consent. However, you may be asked to perform tasks in a public space and bystanders could recognize you.

The Virginia Tech (VT) Institutional Review Board (IRB) may view the study’s data for auditing purposes. The IRB is responsible for the oversight of the protection of human subjects involved in research.

**VI. Compensation**

Participating in this study will provide you with no compensation, monetarily or otherwise. All subjects in the study do so of their own free will and a desire to further computing technologies.

**VII. Freedom to Withdraw**

It is important for you to know that you are free to withdraw from this study at any time without penalty. You are free not to answer any questions that you choose or respond to what is being asked of you without penalty.

Please note that there may be circumstances under which the investigator may determine that a subject should not continue as a subject.
VIII. Questions or Concerns

Should you have any questions about this study, you may contact one of the research investigators whose contact information is included at the beginning of this document.

Should you have any questions or concerns about the study’s conduct or your rights as a research participant, or need to report a research-related injury or event, you may contact the Virginia Tech Institutional Review Board at irb@vt.edu or (540) 231-3732.

IX. Subject’s Consent

I have read the Consent Form and conditions of this project. I have had all my questions answered. I hereby acknowledge the above and give my voluntary consent:

_______________________________________________ Date__________
Subject signature

_______________________________________________
Subject printed name
A.5.2 Questionnaire

Modified SUS Questionnaire
Usability Questionnaire
Participant #: 
Technique:
Please circle the appropriate response for each question:

1. I think that I would like to use this technique frequently.
   
   Strongly Disagree  Disagree  Neutral  Agree  Strongly Agree

2. I found the technique unnecessarily complex.
   
   Strongly Disagree  Disagree  Neutral  Agree  Strongly Agree

3. I thought the technique was easy to use.
   
   Strongly Disagree  Disagree  Neutral  Agree  Strongly Agree

4. I think that I would need the support of a technical person to be able to use this technique.
   
   Strongly Disagree  Disagree  Neutral  Agree  Strongly Agree

5. I thought there was too much inconsistency in this technique.
   
   Strongly Disagree  Disagree  Neutral  Agree  Strongly Agree

6. I would imagine that most people would learn to use this technique very quickly.
   
   Strongly Disagree  Disagree  Neutral  Agree  Strongly Agree

7. I found the technique very cumbersome to use.
   
   Strongly Disagree  Disagree  Neutral  Agree  Strongly Agree

8. I felt very confident using the technique.
   
   Strongly Disagree  Disagree  Neutral  Agree  Strongly Agree
9. I needed to learn a lot of things before I could get going with this technique.

   Strongly Disagree   Disagree   Neutral   Agree   Strongly Agree

10. I was physically comfortable using this technique.

   Strongly Disagree   Disagree   Neutral   Agree   Strongly Agree
Interview Questions
View management and user interface optimization for wide-area mobile augmented reality

Interview topics/questions:
1. Which of the interaction techniques that you tried today did you prefer? Why?
2. Comment on the following aspects of the interaction techniques:
   a. Ease of learning
   b. Ease of use
   c. Naturalness
   d. Comfort
   e. Speed
   f. Accuracy
   g. Fun
3. Do you think these interfaces would be usable in environments with these characteristics? Why or why not?
   a. Dark
   b. Bright
   c. Outdoors
   d. Noisy
   e. Rainy/damp
A.5.3 IRB Approval Letter
MEMORANDUM

DATE: February 16, 2018
TO: Douglas Andrew Bowman, Wallace Santos Lages, Yuan Li, Lee Lisle
FROM: Virginia Tech Institutional Review Board (FWA00000572, expires January 29, 2021)

PROTOCOL TITLE: View management and user interface optimization for wide-area mobile augmented reality

IRB NUMBER: 17-431

Effective February 16, 2018, the Virginia Tech Institution Review Board (IRB) approved the Amendment request for the above-mentioned research protocol.

This approval provides permission to begin the human subject activities outlined in the IRB-approved protocol and supporting documents.

Plans to deviate from the approved protocol and/or supporting documents must be submitted to the IRB as an amendment request and approved by the IRB prior to the implementation of any changes, regardless of how minor, except where necessary to eliminate apparent immediate hazards to the subjects. Report within 5 business days to the IRB any injuries or other unanticipated or adverse events involving risks or harms to human research subjects or others.

All investigators (listed above) are required to comply with the researcher requirements outlined at: http://www.irb.vt.edu/pages/responsibilities.htm

(Please review responsibilities before the commencement of your research.)

PROTOCOL INFORMATION:

Approved As: Expedited, under 45 CFR 46.110 category(ies) 7
Protocol Approval Date: May 1, 2017
Protocol Expiration Date: April 30, 2018
Continuing Review Due Date*: April 16, 2018

*Date a Continuing Review application is due to the IRB office if human subject activities covered under this protocol, including data analysis, are to continue beyond the Protocol Expiration Date.

FEDERALLY FUNDED RESEARCH REQUIREMENTS:

Per federal regulations, 45 CFR 46.103(f), the IRB is required to compare all federally funded grant proposals/work statements to the IRB protocol(s) which cover the human research activities included in the proposal / work statement before funds are released. Note that this requirement does not apply to Exempt and Interim IRB protocols, or grants for which VT is not the primary awardee.

The table on the following page indicates whether grant proposals are related to this IRB protocol, and which of the listed proposals, if any, have been compared to this IRB protocol, if required.
<table>
<thead>
<tr>
<th>Date*</th>
<th>OSP Number</th>
<th>Sponsor</th>
<th>Grant Comparison Conducted?</th>
</tr>
</thead>
<tbody>
<tr>
<td>04/26/2017</td>
<td>PET47EVN</td>
<td>University of California, Santa Barbara (Title: View management and user interface optimization for wide-area mobile augmented real)</td>
<td>Not required (VT is not primary inst.)</td>
</tr>
</tbody>
</table>

* Date this proposal number was compared, assessed as not requiring comparison, or comparison information was revised.

If this IRB protocol is to cover any other grant proposals, please contact the IRB office (irbadmin@vt.edu) immediately.
A.6 Study of ARCritique

A.6.1 User Consent Form
Title of research study: IRB-21-003 Remote Critique Sessions through Collaborative Augmented Reality

Principal Investigator: Doug A. Bowman dbowman@vt.edu (540)231-2058

Other study contact(s):
- David Hicks hicks@vt.edu (540)231-8332
- Wallace S. Lages wlages@vt.edu (540)231-5547
- Sang Won Lee sangwonlee@vt.edu (540)231-4857
- Yuan Li yli92@vt.edu (208)807-6797
- Sang Won Lee akshay@vt.edu (540)231-2108

Key Information: The following is a short summary of this study to help you decide whether or not to be a part of this study. More detailed information is listed later on in this form.

We invite you to participate in this Augmented Reality (AR) study to help us testing a collaborative application for remote design critique about physical artifacts. This work is hoping to help design researchers and education, make AR more usable and we deeply appreciate your participation.

Detailed Information: The following is more detailed information about this study in addition to the information listed above.

Who can I talk to?
If you have questions, concerns, or complaints, or think the research has hurt you, talk to the research team at yli92@vt.edu / (208) 807-6797.

This research has been reviewed and approved by the Virginia Tech Institutional Review Board (IRB). You may communicate with them at 540-231-3732 or irb@vt.edu if:

- You have questions about your rights as a research subject
- Your questions, concerns, or complaints are not being answered by the research team
- You cannot reach the research team
- You want to talk to someone besides the research team to provide feedback about this research

How many people will be studied?
We plan to include about 12 people in this research study.

What happens if I say yes, I want to be in this research?
After initial recruitment, the investigators will first send an email to the participants with a link to the consent form. After the participants agree and consent, they will start with a pre-questionnaire to gather demographic information, such as age, gender, etc. (see the supplemental material for details).
Consent to Take Part in a Research Study

In one experiment session with consenting users, the experimenter will hold a Zoom meeting for verbal communication throughout the study. The experimenter begins the study by walking the participants through the application, providing instructions on how to use the phone to scan, share, view, and communicate around the virtual model. At the same time, the experimenter will also remind the participants to clear their surroundings for their safeties during the study.

Once the participants verbally confirm that they are familiar with the application, they move on to the critique session. The experimenter will record the Zoom meeting and the ARCritique application.

When the student and the instructor finish the critique session, the student will be asked to fill out a post-study questionnaire and the instructor will be given a shorter reflection questionnaire (as we anticipate we will recruit more students than instructors). After the questionnaires, at the end of their participation, a short interview (no more than 30 minutes) will be carried out to ask for more detailed feedback.

To thank you for your participation, we will give out $15 electronic gift card at the end of the study.

What happens if I say yes, but I change my mind later?
You can leave the research at any time, for any reason, and it will not be held against you. If you decide to leave the research, contact the investigator so that the investigator can terminate current experiment session. Any data collected by far will be disregarded and will not be used for further study. There will be no follow-up procedures afterwards.

Withdrawing from the experiment does not affect receiving $15 gift card compensation.

Is there any way being in this study could be bad for me? (Detailed Risks)
Using AR technology can produce symptoms of sickness or discomfort in some users. These symptoms are usually mild, and may include dizziness, nausea, eye strain, headache, or disorientation. During tasks involving physical movement, there is also some risk that you could collide with obstacles in the physical environment.

To mitigate the risk of physical obstacles and cabling, we will remind you to clear the area of obstacles during the study session.

What happens to the information collected for the research?
We will make every effort to limit the use and disclosure of your personal information, including research study and medical records, only to people who have a need to review this information. We cannot promise complete confidentiality. Organizations that may inspect and copy your information include the IRB, Human Research Protection Program, and other authorized representatives of Virginia Tech.

All data collected during this study will be done so anonymously and stored in a password protected computer. No names, contact information, or any other identifying information will be attached to your responses to an investigator’s questions or to your results from the provided tasks.
Consent to Take Part in a Research Study

At no time will the researchers release identifiable results of the study to anyone other than individuals working on the project without your written consent.

If identifiers are removed from your private information or samples that are collected during this research, that information or those samples could be used for future research studies or distributed to another investigator for future research studies without your additional informed consent.

The results of this research study may be presented in summary form at conferences, in presentations, reports to the sponsor, academic papers, and as part of a thesis/dissertation.

**Can I be removed from the research without my OK?**

The person in charge of the research study or the sponsor can remove you from the research study without your approval. Possible reasons for removal include technical issues.
## Consent to Take Part in a Research Study

**Signature Block for Capable Adult** *(contact HRPP for questions about adults not capable of providing consent)*

Your signature documents your permission to take part in this research. We will provide you with a signed copy of this form for your records.

<table>
<thead>
<tr>
<th>Signature of subject</th>
<th>Date</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Printed name of subject</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Signature of person obtaining consent</th>
<th>Date</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Printed name of person obtaining consent</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
</tr>
</tbody>
</table>
A.6.2 Questionnaires

Pre-Study Questionnaire
Q1 Thank you for your interest in participating in our study! Please read through the following consent document before proceeding. If you agree and consent, we can move to the next section.
You are invited to participate in a research study. This form includes information about the study and contact information if you have any questions. “I am a graduate student at Virginia Tech, and I am conducting this research as part of my PhD dissertation work.” Ø WHAT SHOULD I KNOW? If you decide to participate in this study, after initial recruitment, the investigators will first send an email to the participants with a link to the consent form. After the participants agree and consent, they will start with a pre-questionnaire to gather demographic information, such as age, gender, etc. (see the supplemental material for details). In one experiment session with consenting users, the experimenter will hold a Zoom meeting for verbal communication throughout the study. The experimenter begins the study by walking the participants through the application, providing instructions on how to use the phone to scan, share, view, and communicate around the virtual model. At the same time, the experimenter will also remind the participants to clear their surroundings for their safety during the study. Once the participants verbally confirm that they are familiar with the application, they move on to the critique session. The experimenter will record the Zoom meeting and the ARCritique application. When the student and the instructor finish the critique session, the student will be asked to fill out a post-study questionnaire and the instructor will be given a shorter reflection questionnaire (as we anticipate we will recruit more students than instructors). After the questionnaires, at the end of their participation, a short interview (no more than 30 minutes) will be carried out to ask for more detailed feedback. To thank you for your participation, we will give out a $30.00 Amazon electronic gift card at the end of the study. The study should take approximately 90 minutes of your time. The risk associated with this study is minor discomfort. Using AR technology can produce symptoms of sickness or discomfort in some users. These symptoms are usually mild, and may include dizziness, nausea, eye strain, headache, or disorientation. During tasks involving physical movement, there is also some risk that you could collide with obstacles in the physical environment. To mitigate the risk of physical obstacles and cabling, we will remind you to clear the area of obstacles during the study session. You can choose whether to be in this study or not. If you volunteer to be in this study, you may withdraw at any time without consequences of any kind. You may also refuse to answer any questions you don’t want to answer and remain in the study. The investigator may withdraw you from this research if circumstances arise which warrant doing so. Ø CONFIDENTIALITY We will do our best to protect the confidentiality of the information we gather from you, but we cannot guarantee 100% confidentiality. Any data collected during this research study will be kept confidential by the researchers. Your interview will be audio-recorded using a digital recorder and then transcribed. The researchers will code the transcripts using a pseudonym (false name). The recordings will be uploaded to a secure password-protected computer in the researcher’s office. The researchers will maintain a list that includes a key to the code. The master key and the recordings will be stored for 3 years after the study has been completed and then destroyed. Ø WHO CAN I TALK TO? If you have any questions or concerns about the research, please feel free to contact Yuan Li at 208-807-6797 (yli92@vt.edu). You are not waiving any legal claims, rights or remedies because of
your participation in this research study. If you have questions regarding your rights as a research participant, contact the Virginia Tech HRPP Office at 540-231-3732 (irb@vt.edu). Please print out a copy of this information sheet for your records. If you would like to participate in this study, click yes to begin or no to exit.

Q24 Would you like to participate in this study?

- Yes (1)
- No (2)

End of Block: Consent Block

Start of Block: Block 3

Q22 Additional COVID-19 Information Related to Consent to Participate in In-Person Research at Virginia Tech

Protocol #21-003

Principal Investigator Name and Contact Information: Doug A. Bowman, dbowman@vt.edu, (540)231-2058

You have agreed to participate in a research study at Virginia Tech. The research study involves in-person contact or procedures. Here are some things you should know about in-person research while COVID-19 remains a risk:

Risks related to COVID-19: If you choose to participate in this study, the risk for COVID-19 are about the same as those posed by similar activities while the virus is still spreading in your community. Similar activities could include grocery shopping, having your car repaired, or getting a haircut. In addition, participation might increase risk to your family, the community, and the research team. You should not participate if you have any conditions or risk factors that could make a COVID-19 infection more serious. Risk factors for severe illness include having other medical conditions such as asthma, diabetes, heart problems, or any other illness. Certain populations might also be at increased risk or unknown risk, including people aged 65 and older, people with disabilities, women who are pregnant or breastfeeding, people who are experiencing homelessness, and people who are part of racial and ethnic minority groups. The information on people who need to take extra precautions is being updated regularly. We encourage you to check for the latest information before you decide whether to participate. Please visit https://www.cdc.gov/coronavirus/2019-ncov/need-extra-precautions/index.html for the most up to date information.

What we are doing to reduce risk to you: Each lab or study has developed a process for conducting the research as safely as possible, given current knowledge about COVID-19. This process has been reviewed by the Human Research Protection Program at Virginia Tech. You will be given a sheet with information specific to your study. You should review this information and ask any questions before you agree to participate. We will not conduct the study during times of increasing community spread or if we cannot obtain the necessary disinfecting supplies and equipment to
reduce the risk of exposure. Everyone working on the study has been instructed to stay home if they have any symptoms that could be related to COVID-19. If someone on the research team tests positive for COVID-19 and you have been exposed, someone will notify you. We will maintain a contact tracing log that is separate from your data and other details about your participation, and we will provide this log to the New River Health District (540 267-8240) who will conduct contact tracing in the case of a positive test. We will destroy this log 60 days after your last visit.

What you can do to reduce risk to us and to the community: Do not participate if you have had any symptoms of COVID-19 in the past 14 days or have been in contact with someone who has symptoms. Symptoms include, but are not limited to, cough, shortness of breath or difficulty breathing, fever, chill, repeated shaking with chills, muscle pain, headache, sore throat, and new loss of taste or smell. Do not participate if you have tested positive for COVID-19 in the past 21 days, even if you have not shown any symptoms. Do not participate if you know you have been exposed to anyone who has tested positive for COVID-19 in the past 21 days. Let us know if you test positive for COVID-19 within the next 14 days. We will provide your contact tracing log to university or public health authorities who will use the tracing log to contact others who may have been exposed during your visit. Wash your hands frequently and observe current guidance on avoiding virus spread from the Centers for Disease Control. Wear a mask or a cloth face covering over your nose and mouth. Depending on the study, another method may be used, such as physical distancing, a face shield, or Plexiglas barrier. For the latest information on COVID-19 please visit: https://www.cdc.gov/coronavirus/2019-ncov/index.html

Q25 By selecting "I consent", you are consenting to proceed in the study

○ I consent (1)

○ I do not consent (2)

Q23 If you consent, please type your name and today's date:

○ Your full name: (1) ________________________________________________

○ Today's date (mm/dd/yyyy): (2) _____________________________________

End of Block: Block 3
Q19 We want to thank you for your interest in our study and want to ensure your safety during these unprecedented times. If you feel ill, have had any symptoms of COVID-19 in the past 14 days, have been in contact with someone known to have tested positive for COVID-19, have been outside of Virginia, or to any COVID-19 hotspot, please contact your researcher and cancel your appointment. Additionally, please contact Virginia Tech Health Center to receive instructions on testing.

We have worked to mitigate risks related to COVID-19. We will ask you to schedule your participation and read the consent forms online. The study will begin after you sign the consent form. The study will be mostly online, with the only exception when the experimenter will send an iPhone to you and take it back after the study is finished. The physical meetings will happen at Virginia Tech campus. In case you don’t have a mask, we will have spare masks that we could provide. You will be asked to complete a COVID-19 screening 24h prior to the meeting. During the meeting, you and the investigator will be maintaining a minimum of 6 feet distance to allow for adequate physical distancing. Equipment handoffs will be completed by placing it at one position by the investigator or the participant. The investigator/participant will then move at least 6-feet away from the table to allow the participant/investigator the physical space needed to work with equipment. Equipment will be disinfected before and after each use. We will enforce that a device is only used by one participant each day to allow for maximal rest time. Along with this, we will thoroughly wipe down each device before and after each use, using disinfecting surface cleaners (spray) and paper/cloth wipes.

We will be keeping contact tracing logs in case there is an exposure, and this information will be provided to the New River Health District in case of potential exposure, who will then contact any other participants or researchers who were exposed. **You should contact the research team in case you discovered that you test positive for COVID-19 within 14 days of your participation.** In the event that the investigator experiences symptoms of COVID-19, all appointments will be canceled until they are cleared from COVID-19. This is to ensure the safety of all people involved in this study. When someone contacts the research team to notify them they have tested positive, the PI will contact the HRPP, who will then notify the Department of Health.

Everyone will need to follow these safety rules to participate in the study. If you have any questions, please reach out to yli92@vt.edu.

End of Block: Block 2

Start of Block: Demographic Block
Q3 Gender
- Male (1)
- Female (2)

Q5 Age

Q9 Are you:
- Right-handed (1)
- Left-handed (2)
- Ambidextrous (3)

Q11 Occupation (if student, indicate graduate or undergraduate):

Q13 If student, major/area of specialization
Q20 If student, which year are you in:

- Freshman (1)
- Sophomore (2)
- Junior (3)
- Senior (4)
- Graduate (5)

Q15 Have you used Zoom for online learning?

- Yes (1)
- No (2)

Q19 If yes, how often do you use Zoom for online learning:

- Never Used (1)
- Once or twice (2)
- 3 to 10 times (3)
- More than 10 times (4)

Q21 Have you used Zoom for design critiques about physical mockups?

- Yes (1)
- No (2)
Q16 If yes, how often do you use Zoom for design critiques about physical mockups:

- Never Used (1)
- Once or twice (2)
- 3 to 10 times (3)
- More than 10 times (4)

Q27 How many times have you tried augmented reality:

- Never used (1)
- Once or twice (2)
- 3 to 10 times (3)
- More than 10 times (4)

End of Block: Demographic Block
A.6. Study of ARCritique

Post-Study Questionnaire
Poststudy Questionnaire

https://virginiatech.qualtrics.com/jfe/form/SV_3PFNwkkwDUkvVgW

Start of Block: Default Question Block

Q1 **ARCritique Session Reflection**
Thank you for helping us evaluate our AR Critique app. Please answer the following questions about your most recent AR Critique session. If there are questions that do not apply or questions you don’t have any thoughts about, feel free to leave them blank.

Q2 Name
________________________________________________________________________

Q3 Date of the Session
________________________________________________________________________

Q6 Time of the Session
________________________________________________________________________

Q7 (For student only) What was your experience of scanning the physical mockup and sharing it with your professor? Did you walk around the object, or did you rotate it on a turntable? Was it easy or difficult, and why? Did you learn anything about the best ways to scan?

________________________________________________________________________

Page 1 of 10
Q9 Thinking about the remote AR critique section you just completed, how well do you think you were able to communicate with each other about the physical mockup?

________________________________________________________________
________________________________________________________________
________________________________________________________________
________________________________________________________________
Page Break
Q10 During the AR critique session, how beneficial is it for you to be able to view the object from various perspectives?

________________________________________________________________
________________________________________________________________
________________________________________________________________
________________________________________________________________
________________________________________________________________

Page Break
Q11 In the collaborative AR critique session, how useful is it for you to see where your collaborator is located?

________________________________________________________________
________________________________________________________________
________________________________________________________________
________________________________________________________________
________________________________________________________________
________________________________________________________________
________________________________________________________________
________________________________________________________________

Page Break
Q12 In the collaborative AR critique session, are the referencing tools (virtual ray, drawings) helpful? If so, why?

________________________________________________________________
________________________________________________________________
________________________________________________________________
________________________________________________________________
________________________________________________________________
________________________________________________________________
Q13 Can you describe any examples of times when the AR app made it much easier to communicate? Any times when the AR app presented barriers to effective communication?

________________________________________________________________
________________________________________________________________
________________________________________________________________
________________________________________________________________
________________________________________________________________
________________________________________________________________

Page Break
Q17 Do you feel like the AR Critique session was sufficient for an effective discussion about the physical mockup, or would an in-person critique session still be needed? Why?

________________________________________________________________
________________________________________________________________
________________________________________________________________
________________________________________________________________
________________________________________________________________

Page Break
Q14 Would you say this session was closer to the experience of a Zoom call, or closer to the experience of an in-person design critique? Why?

__________________________________________________________________________________
__________________________________________________________________________________
__________________________________________________________________________________
__________________________________________________________________________________
__________________________________________________________________________________

Page Break
Q15 Did you use (or want to use) any additional methods besides the AR app and verbal communication to help you communicate during the session (e.g., gestures, sketches)? If so, why were they needed?

________________________________________________________________
________________________________________________________________
________________________________________________________________
________________________________________________________________

Q16 Additional Comments?

________________________________________________________________
________________________________________________________________
________________________________________________________________
________________________________________________________________

End of Block: Default Question Block
Interview Questions
Interview questions for students

- Tell us about your previous experiences discussing physical mockups in person during a critique session. What is the purpose of these sessions and how do they support your work/learning?

- Tell us about your experience of using Zoom or other remote tools for discussing physical mockups.

- What do you consider to be the advantages and / or challenges in using zoom during critique sessions when compared with in-person sessions?

- How does the AR Critique experience compare with the in-person session and the zoom sessions?

- Do you think that the AR Critique app addresses some of the challenges of doing critiques of physical mockups remotely?

- Is AR Critique or Zoom closer to what you would experience in person? Why?

- Do you think it’s important to be able to view the model from any viewpoint in AR? Why or why not? How did this feature support or hinder the critique session?

- Do you think it’s important to be able to see where your instructor was and which direction their phone was facing? Why or why not? How did this feature support or hinder the critique session?

- Do you think it’s important to be able to point to particular locations or features on the model? Why or why not? How did this feature support or hinder the critique session?

- Do you think it’s important to be able to draw on the surface of the model? Why or why not? How did this feature support or hinder the critique session?

- Thinking about the whole critique session, how well did AR Critique support your communication about the mockup? Did certain features of the app make your communication easier or harder?

- Do you have any other comments or suggestions about AR Critique or this experience?

- Can you think of any other scenarios in which this app would be useful?
A.6.3 IRB Approval Letter
MEMORANDUM

DATE: April 23, 2021

TO: Douglas Andrew Bowman, Wallace Santos Lages, Yuan Li, Sang Won Lee, David Hicks

FROM: Virginia Tech Institutional Review Board (FWA00000572)

PROTOCOL TITLE: Remote Critique Sessions through Collaborative Augmented Reality

IRB NUMBER: 21-003

Effective April 23, 2021, the Virginia Tech Human Research Protection Program (HRPP) determined that this protocol meets the criteria for exemption from IRB review under 45 CFR 46.104(d) category (ies) 2(ii).

Ongoing IRB review and approval by this organization is not required. This determination applies only to the activities described in the IRB submission and does not apply should any changes be made. If changes are made and there are questions about whether these activities impact the exempt determination, please submit an amendment to the HRPP for a determination.

This exempt determination does not apply to any collaborating institution(s). The Virginia Tech HRPP and IRB cannot provide an exemption that overrides the jurisdiction of a local IRB or other institutional mechanism for determining exemptions.

All investigators (listed above) are required to comply with the researcher requirements outlined at: https://secure.research.vt.edu/external/irb/responsibilities.htm

(Please review responsibilities before beginning your research.)

PROTOCOL INFORMATION:

Determined As: Exempt, under 45 CFR 46.104(d) category(ies) 2(ii)
Protocol Determination Date: March 3, 2021

ASSOCIATED FUNDING:

The table on the following page indicates whether grant proposals are related to this protocol, and which of the listed proposals, if any, have been compared to this protocol, if required.
**SPECIAL INSTRUCTIONS:**
This amendment, submitted April 22, 2021, updates research protocol, recruitment materials, and consent forms to increase the study compensation from $15 to $30 in Amazon gift card.

<table>
<thead>
<tr>
<th>Date*</th>
<th>OSP Number</th>
<th>Sponsor</th>
<th>Grant Comparison Conducted?</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

* Date this proposal number was compared, assessed as not requiring comparison, or comparison information was revised.

If this protocol is to cover any other grant proposals, please contact the HRPP office (irb@vt.edu) immediately.