Walk-Centric User Interfaces for Mixed Reality

Wallace Santos Lages

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 and Applications

Doug A. Bowman, Committee Chair
Joe Gabbard
Tobias Höllerer
Chris L. North
Nicholas F. Polys

June 22, 2018
Blacksburg, Virginia

Keywords: Augmented Reality, Virtual Reality, 3D User Interfaces

Copyright © 2018 by Wallace Santos Lages
Walk-Centric User Interfaces for Mixed Reality

Wallace Santos Lages

ABSTRACT

Walking is a natural part of our lives and is also becoming increasingly common in mixed reality. Wireless headsets and improved tracking systems allow us to easily navigate real and virtual environments by walking. In spite of the benefits, walking brings challenges to the design of new systems. In particular, designers must be aware of cognitive and motor requirements so that walking does not negatively impact the main task. Unfortunately, those demands are not yet fully understood. In this dissertation, we present new scientific evidence, interaction designs, and analysis of the role of walking in different mixed reality applications.

We evaluated the difference in performance of users walking vs. manipulating a dataset during visual analysis. This is an important task, since virtual reality is increasingly being used as a way to make sense of progressively complex datasets. Our findings indicate that neither option is absolutely better: the optimal design choice should consider both user’s experience with controllers and user’s inherent spatial ability. Participants with reasonable game experience and low spatial ability performed better using the manipulation technique. However, we found that walking can still enable higher performance for participants with low spatial ability and without significant game experience.

In augmented reality, specifying points in space is an essential step to create content that is registered with the world. However, this task can be challenging when information about the depth or geometry of the target is not available. We evaluated different augmented reality techniques for point marking that do not rely on any model of the environment. We found that triangulation by physically walking between points provides higher accuracy than purely perceptual methods. However, precision may be affected by head pointing tremors. To increase the precision, we designed a new technique that uses multiple samples to obtain a better estimate of the target position. This technique can also be used to mark points while walking. The effectiveness of this approach was demonstrated with a controlled augmented reality simulation and actual outdoor tests.

Moving into the future, augmented reality will eventually replace our mobile devices as the main method of accessing information. Nonetheless, to achieve its full potential, augmented reality interfaces must support the fluid way we move in the world. We investigated the potential of adaptation in achieving this goal. We conceived and implemented an adaptive workspace system, based in the study of the design space and through user contextual studies. Our final design consists in a minimum set of techniques to support mobility and integration with the real world. We also identified a set of key interaction patterns and desirable properties of adaptation-based techniques, which can be used to guide the design of the next-generation walking-centered workspaces.
Walk-Centric User Interfaces for Mixed Reality

Wallace Santos Lages

GENERAL AUDIENCE ABSTRACT

Until recently, walking with virtual and augmented reality headsets was restricted by issues such as excessive weight, cables, tracking limitations, etc. As those limits go away, walking is becoming more common, making the user experience closer to the real world. If well explored, walking can also make some tasks easier and more efficient. Unfortunately, walking reduces our mental and motor performance and its consequences in interface design are not fully understood.

In this dissertation, we present studies of the role of walking in three areas: scientific visualization in virtual reality, marking points in augmented reality, and accessing information in augmented reality. We show that although walking reduces our ability to perform those tasks, careful design can reduce its impact in a meaningful way.
To my wife,
my parents,
and my family.
Acknowledgments

This work would not be possible without the support I received from many people over these four years. They allowed me to focus on my research, navigate the complexities of a new culture, have fun, and grow along the way. Thank you!

I owe a special thanks to my advisor Doug Bowman, for all the enjoyable discussions and for the incentive to keep pushing forward when things seemed difficult. Thanks for being an excellent mentor, an exemplar scientist and person.

I am also very grateful to David Laidlaw. David made himself available to listen and guide me during my whole second year. Even though none of that work became part of this dissertation, it was part of the process that lead me to it.

I also would like to acknowledge:

My committee, Joe Gabbard, Tobias Höllerer, Chris L. North and Nicholas F. Polys for the guidance, encouragement, and hard questions.

The 3DI group members: Bireswar Laha, Felipe Bacim, Mahdi Nabiyuni, Panagiotis Apostolellis, Run Yu, Lawrence Warren, Yuan Li, Javier Tibau, Leonardo Soares, and Lee Lisle; for the all the discussions and piloting of my studies.

Sharon Kinder-Potter, Andrea Kavanaugh, Cal Ribbens, Phyllis Newbill, Tanner Uptegrove, Zach Duer, Melissa Wyers, Lisa Jansen, Holly Williams, and Teresa Hall, for the administrative and technical support that made my life so much easier.

Funding support from the Brazilian Council for Scientific and Technological Development, the National Science Foundation, and from the Immersive Sciences program in the Office of Naval Research.
Attributions

Doug A. Bowman is the Frank J. Maher Professor of Computer Science and Director for the Center for Human-Computer Interaction at Virginia Tech. As my PhD advisor, he was my collaborator and co-author on all papers forming this dissertation.

Tobias Höllerer is a Professor of Computer Science at the University of California, Santa Barbara. He contributed on the writing of the paper forming the Chapter 2 of this dissertation.

Yuan Li is a Master in Computer Science and Research Assistant in the 3D interaction group at Virginia Tech. He was a co-author in the research described in Chapter 2 of this dissertation.

Table of Contents

1 INTRODUCTION 1
  1.1 WALK-CENTRIC USER INTERFACES 2
  1.2 RESEARCH GOALS AND SCOPE 6
  1.3 RESEARCH QUESTIONS 7
  1.4 METHODS AND SCOPE 8
  1.5 MAIN CONTRIBUTIONS 9
  1.6 THE STRUCTURE OF THIS DISSERTATION 10

2 UNDERSTANDING WALKING TRADEOFFS IN VIRTUAL REALITY 11
  2.1 INTRODUCTION 11
  2.2 BACKGROUND AND RELATED WORK 12
  2.3 OVERALL EXPERIMENTAL DESIGN 17
  2.4 PILOT STUDY 21
  2.5 MAIN EXPERIMENT 23
  2.6 CONCLUSION 37

3 MODEL-FREE MARKING IN AUGMENTED REALITY 38
  3.1 INTRODUCTION 38
  3.2 BACKGROUND AND RELATED WORK 39
  3.3 GEOMETRY-BASED MARKING TECHNIQUES 41
  3.4 EXPERIMENT I: PERCEPTUAL VS. GEOMETRIC MARKING 41
  3.5 TECHNIQUES 42
  3.6 APPARATUS 43
  3.7 TASK AND ENVIRONMENT 44
  3.8 PARTICIPANTS AND PROCEDURE 45
  3.9 RESULTS 46
  3.10 DISCUSSION 48
  3.11 EXPERIMENT II: IMPROVING PRECISION WITH MULTI-SAMPLE MARKING 50
  3.12 EXPERIMENT DESIGN 53
  3.13 APPARATUS 54
  3.14 PARTICIPANTS AND PROCEDURE 54
  3.15 RESULTS 55
  3.16 DISCUSSION 57
  3.17 QUALITATIVE USER STUDY 57
  3.18 CONCLUSIONS 59

4 MANAGING WORKSPACES IN AUGMENTED REALITY 60
  4.1 INTRODUCTION 60
  4.2 RELATED WORK 61
  4.3 RESEARCH APPROACH 67
  4.4 DEFINING THE DESIGN SPACE 68
  4.5 AR ADAPTIVE WORKSPACE - VERSION I 75
  4.6 EVALUATING THE AR WORKSPACE 80
  4.7 STUDY I 85
  4.8 AR ADAPTIVE WORKSPACE – VERSION II 100
<table>
<thead>
<tr>
<th>Section</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>4.9</td>
<td>STUDY II</td>
<td>103</td>
</tr>
<tr>
<td>4.10</td>
<td>DISCUSSION</td>
<td>109</td>
</tr>
<tr>
<td>4.11</td>
<td>DESIGNING FOR WALKING</td>
<td>112</td>
</tr>
<tr>
<td>4.12</td>
<td>LIMITATIONS</td>
<td>113</td>
</tr>
<tr>
<td>4.13</td>
<td>CONCLUSION</td>
<td>115</td>
</tr>
<tr>
<td>5</td>
<td>CONCLUSION AND FUTURE WORK</td>
<td>116</td>
</tr>
<tr>
<td>5.1</td>
<td>SUMMARY OF THE CONTRIBUTIONS</td>
<td>116</td>
</tr>
<tr>
<td>5.2</td>
<td>FUTURE WORK</td>
<td>117</td>
</tr>
<tr>
<td>REFERENCES</td>
<td>121</td>
<td></td>
</tr>
<tr>
<td>APPENDIX A: TABLES</td>
<td>132</td>
<td></td>
</tr>
<tr>
<td>APPENDIX B: EXPERIMENTAL DOCUMENTS STUDY 1</td>
<td>133</td>
<td></td>
</tr>
<tr>
<td>APPENDIX C: EXPERIMENTAL DOCUMENTS STUDY 2</td>
<td>154</td>
<td></td>
</tr>
<tr>
<td>APPENDIX D: EXPERIMENTAL DOCUMENTS STUDY 3</td>
<td>157</td>
<td></td>
</tr>
</tbody>
</table>
List of Figures

Figure 2-1: Examples of datasets from the Complex (left) and Simple (right) groups. 18
Figure 2-2: View of the Experimental Environment. 20
Figure 2-3: Task completion time for each participant in the two conditions. Left: Simple dataset, Right: Complex datasets. 28
Figure 2-4: Participants grouped according to high (H) and low (L) scores. 28
Figure 2-5: Time Difference Between Walking and Non-Walking Conditions. 32
Figure 2-6: Time in the Walking Condition. 32
Figure 3-1: Perceptual and Geometric Techniques. 42
Figure 3-2: Experimental apparatus. 44
Figure 3-3: Photo of the experimental site taken from the participant’s point of view. 45
Figure 3-4: Experiment I results. Bars show the standard error at each target distance. 46
Figure 3-5: Increase in variability with distance in Experiment I. 47
Figure 3-6: The VectorCloud technique. 51
Figure 3-7: Top view of the distribution of intersections generated by two viewpoints A and B. 52
Figure 3-8: The VR experimental setup emulating the first experiment. 53
Figure 3-9: Experiment II results. 55
Figure 3-10: Variability increase for original Geometric technique and the VectorCloud. 56
Figure 3-11: Variability on the view plane is also reduced in the VectorCloud technique. 56
Figure 4-1: Diagram of the study methodology. 67
Figure 4-2: Windows adapting to nearby walls. 69
Figure 4-3: Adaptive System Model. 69
Figure 4-4: Adaptation Taxonomy. 70
Figure 4-5: Windows following the user position while staying on the wall surface. 73
Figure 4-6: Wall Editor. 76
Figure 4-7: Configuration Menu. 79
Figure 4-8: Windows used in the first scenario. 81
Figure 4-9: Windows used in the second scenario. 81
Figure 4-10: Paths used in the first study. 82
Figure 4-11: Path used in the second study. 83
Figure 4-12: Playback of a recorded log. 84
Figure 4-13: Diagram of the experimental procedure of Study I. 86
Figure 4-14: Diagram of the analysis process. 87
Figure 4-15: User preference for behaviors. 89
Figure 4-16: Core themes resulting from the axial coding. 91
Figure 4-17: Interaction between the core categories. 97
Figure 4-18: Overview of adjustable behaviors. 102
Figure 4-19: Diagram of the adjustable behaviors with Freeze. 103
Figure 4-20: Diagram of the experimental procedure of Study II. 104
Figure 4-21: Use of the new behaviors. 106
Figure 4-22: Hypothesized relationship between user actions, dual-task costs, and task impacts. 111
List of Tables

Table 2-1: Mean times and standard deviations for the main experiment. 27
Table 2-2. Participant distribution. 29
Table 2-3. Results of the cognitive factors tests. 29
Table 2-4. Significant effects for the time difference. 30
Table 2-5. Significant effects of the regression model for walking time. 31
Table 3-1: Angular drift between Vive and HoloLens. 50
Table 4-1: Equivalence of properties in our taxonomy. 74
Table 4-2: Combinations most tried in scenario 1 88
Table 4-3: Combinations most tried during scenario 2 88
Table 4-4: Most frequent codes for the Auto-Center behavior 90
Table 4-5: Most frequent codes for the Rotation on B behavior 90
Table 4-6: Most frequent codes for the Attraction behavior 90
Table 4-7: Frequency of suggestion codes 98
Table 4-8: Correspondence between suggested changes and goals 101
Table 4-9: Participants mention to the use of each mode for a specific goal 105
Table 4-10: Quotations of the use of each goal code 106
Table 0-1 - Descriptive codes created during the initial analysis 132
List of Terms

WALKING – Visually-guided human locomotion caused by the coordinated movement of the legs, torso, and head. We also use the term to refer to the act of physically moving through space to obtain different points of view or change our surroundings.

AR (AUGMENTED REALITY) – Real time combination of real objects and virtual objects registered in tridimensional space (Azuma 1997). In this dissertation, we consider as AR any interactive blending, whether it is registered or not.

VR (VIRTUAL REALITY) – An approach that uses displays, tracking and other technologies to immerse the user in a synthetic, spatial and interactive environment (LaViola et al. 2017).

MR (MIXED REALITY) – Continuum that encompass all the different ways of experiencing digital and real information, the extremes being fully virtual and fully real (Milgram & Kishino 1994).

FOV (FIELD OF VIEW) – Extent of the world that can be seen simultaneously, in degrees of visual angle.

HMD (HEAD MOUNTED DISPLAY) – Device used to display virtual information that is supported by the head.

DOF (DEGREES OF FREEDOM) – Number of dimensions that can be varied independently. In this dissertation it is used most of the time to quantify the independent motions of a body (LaViola et al. 2017).

IMU (INERTIAL MEASUREMENT UNIT) – Combination of accelerometer, gyroscope, and compass that allows the estimation of angular and position changes.

INTERACTION TECHNIQUE – Method that allows users to accomplish a task by means of the user interface. It specifies how hardware inputs are mapped into actions and outputs are mapped to the output devices (LaViola et al. 2017).

TRACKING – System used to determine the position and orientation of an object as it moves through space.

IMMERSIVE VIRTUAL ENVIRONMENT – A synthetic world which provides a first-person experience, under real-time control of the user (LaViola et al. 2017). May simulate multiple senses such as vision, hearing, smell, etc.

WALK-CENTRIC INTERFACES – Interaction techniques or interfaces that consider aspects of walking in their design.
1 INTRODUCTION

Walking is a fundamental human ability. Walking can get us within the reach of interesting objects or far away from dangerous ones. It can clear us from obstacles and allow us to organize our lives according to different spaces. It is so natural, that most people do not even think about it.

More and more, we also bring computers with us when we walk. First, portable computers and laptops; then smartphones, tablets, and smartwatches. Each of those have drastically changed the way we interact with computers, with each other, and with the world.

Mixed reality (MR) technology has been following a similar path. From initially heavy and tethered devices, both augmented reality (AR) and virtual reality (VR) headsets have become lighter and more portable. Tracking has improved significantly and now provides low latency and high accuracy in larger spaces. Simultaneous location and mapping allows devices to work without the need of external markers or sensors.

What does that mean for walking? First, walking will become more prevalent in VR applications. Walking was quickly recognized by the early VR researchers as a natural and immersive way to explore virtual environments (Chance et al. 1998; Iwata & Fujii 1996; Slater et al. 1995). Natural walking is often easier than using controllers and provides a better sense of presence. In addition, studies have shown that walking can be beneficial for applications requiring spatial understanding (Ruddle & Lessels 2009; Ruddle et al. 2011).

In AR, wearable devices will eventually enable a new class of applications that can be used while walking. Unlike VR, users in AR can still see the real world, which makes it a potential substitute for smartphones and other mobile devices.

Unfortunately, little research has been conducted so far to determine the desirable characteristics of mobile AR interfaces and the implications of walking in terms of possibilities for MR design. Besides motor restrictions, it is known that walking draws on
the same resources used by executive functions and attention (Thomas & Seiffert 2010; Thomas & Seiffert 2011; Yogev-Seligmann et al. 2008b). Self-motion was found to interfere with multiple-object tracking in VR and the real world (Thomas & Seiffert 2010). Even maintaining body posture can be costly enough to impact the performance in memory tasks (Pellecchia 2003; Riley et al. 2012). In particular, walking with AR can be dangerous, as users start facing unstructured and dynamic real-world environments.

1.1 Walk-Centric User Interfaces

In this dissertation, we use the term walking to refer to visually-guided human locomotion caused by the coordinated movement of the legs, torso, and head. We assume a broad view of walking and include all the mechanisms that allow us to perceive the environment, steer away from obstacles, and coordinate our legs. We also use it to refer to the act of physically moving through space, which allows us to get different points of view or change our surroundings.

There are a few different ways in which walking can be considered in the design of MR applications. Walking can be seen as a nuisance that must be accounted for, as a way to improve our spatial cognitive abilities, as an opportunity to design new techniques, or as a specific way of operating in the world. In this dissertation, we use the term walk-centric to refer to interaction techniques or designs that consider walking in any of the above ways. Those are not exclusive lenses and a good design will often involve a few of them.

Consider a task that does not requires walking, being performed during walking conditions. In this case, an option to improve usability would be to modify the interface or technique to mitigate cognitive and motor challenges presented by walking. If walking is the primary task, one might consider a design that avoids interference with walking. Examples in this category are systems that reduce visual load to allow more focus on the environment or systems that use modalities appropriate for walking (such as voice) (Mousavi et al. 1995). Depending on the task, it may be even beneficial to switch to walking instead of staying stationary. Examples include applications that require spatial
navigation abilities or applications that use the space to improve memory and spatial understanding (Robertson et al. 1998; Robertson et al. 2000).

Another possibility is to design walking-based techniques, in which walking is used as a fundamental part of the interaction. In this way, walking is useful as a means to achieve a goal (other than moving). For example, a system that changes the level of detail depending on the user’s heading and position (Ballendat et al. 2010). In this case, the goal task is not related to walking, but walking makes it attainable.

Finally, it is also possible to design interfaces or techniques that perceive and adapt to the changes of context created by walking (i.e., changes in the surroundings and the user’s position in space). Examples include interfaces that conform to new physical environments or that reorient themselves when the user walks around in a room. We will now briefly review the prior work in those areas and then prior work specific to AR.

**Walking as part of interaction**

In some systems, walking has been used to improve performance or user experience. Some examples are systems that support collocated work or those based on tangible/spatial interfaces. Interaction techniques for these systems will often consider aspects that people use naturally in everyday interaction such as gaze, gestures, body position, and body orientation in space. Lundgren et al. (2015) propose a design framework involving four perspectives: stoical, technological, spatial, and temporal. In this framework, the spatial perspective describes design properties that are related to all aspects of space, location, and the physical environment. The spatial perspective has three properties (Proximity, Location and Movement) that are interrelated to the other perspectives. For example, Movement is related to Location, which in turn affects the social perspective.

From the ubiquitous computing field, Ballendat et al. (2010) understand the use of spatial information in interaction in terms of **Proxemics**, the study of the interpersonal distance between people. Knowledge of features of the space can allow devices to react to where users are in the environment, interpreting the directed attention allows them to trigger appropriate actions. Movement can be used to guide continuous actions (such as scaling
an interface with distance) or to trigger discrete actions (as users enter a specific physical zone). The main elements in Proxemic Interaction are: position, identity, movement, and orientation.

**Interaction while walking**

While walking, users may need to look away from the path they are walking. In addition, bouncing can make it harder to acquire buttons in the interface. Most studies have focused on evaluating the impact of walking on performance, attention, and workload on cellphones, smartphones, and tablets. Bergstrom-Lehtovirta studied the effect of walking speed on touchscreen accuracy. They found that, although even slow speeds negatively impact accuracy, users reach best performance when walking at around 80% of their preferred walking speed (Bergstrom-Lehtovirta et al. 2011). Furthermore, holding the device with two hands does not help to improve input accuracy (Nicolau & Jorge 2012).

Musić & Murray-Smith (2016) developed models to compensate for the deviation between the intended point and the actual touch location while walking, evaluating with different input techniques and walking speeds. They found out that correction models need to take into account walking speed and that the gait phase alters the tapping distribution. Interestingly, the results suggest that most taps do not occur during the phase with most accuracy.

Popova-Dlugosch et al. (2013) reported an experiment where participants used a tablet in a seated position or while walking through an indoor obstacle course. They found that the completion time while walking was significantly higher, and that the average speed was reduced by 32% when using the tablet in the obstacle course. The authors also measured an increase in workload during the walking condition (measured by NASA-TLX questionnaire). Hayes et al. (2014), also evaluating tablet use while walking and sitting, derived a model to predict target size based in an acceptable error rate. The target sizes for mobile interaction were consistently larger for any error rate, and the sizes during walking were approximately 1/3 larger.
A DAPTING TO WALKING CONDITIONS

Kane et al. propose the term Walking User Interface to group user interfaces designed to compensate for the effects of walking on the usability of mobile devices (Kane et al. 2008). They present, as an example, an adaptive user interface for mobile devices that can dynamically change the size as the user moves. A user evaluation showed that participants using the adaptive design had lower performance than participants which used a static version with larger buttons. However, the proposed design created a tradeoff which required participants to scroll more in the screens with larger buttons. We prefer to use the more general term walk-centric to include other ways to consider walking in interaction design.

I NTERACTION IN AR

Previous work has explored different techniques and input methods for AR. Billinghurst et al. (2001) describe early systems that used tangible interfaces for interaction with spatial data. For example, in Shared Space, users can manipulate virtual objects by using marked cards. When corresponding cards are handled, users can see different animations. Henderson & Feiner (2010) extended the idea by using features of existing objects in the interaction area for opportunistic haptic feedback, thus not requiring a dedicated physical proxy. In addition, the system developed can use properties of the object to aid in gesture recognition. Exploring the opposite line, Issartel et al. (2016) designed a cubic device with pressure sensors, which is tracked by the camera of the AR device. By virtually replacing the interior of the device, they constructed a tangible volume, which can support different interaction metaphors.

Improved tracking algorithms also allowed the use of bare-hand interaction in AR (Lee & Hollerer 2007). One example is Handy AR, a system that uses the human hand as reference instead of 2D markers. By tracking the fingertips, their approach is able to perform the camera estimation and allows simple selection and manipulation. Hand gestures can also be integrated with other input modalities, such as voice and eye input. Höllerer & Turk (2006) present a system where a state machine is used to combine several input methods to improve the process of manipulating virtual objects.
Few studies, however, have explored walking for interaction in AR. The Tinmith-Metro modelling system (Piekarski & Thomas (2004) allows users to create large architectural models by using 2D projection over planes in space. The planes are used to define intersections when sketching but can also be used for manipulation. In particular, they discuss the possibility of attaching planes to the body coordinate system, which would cause objects to be dragged around as the user moves. The most common application of walking in AR is to control player and camera movement in games and other applications. For example, in ARQuake players walk around in the real world shooting virtual monsters (Thomas et al. 2002). The movement of the player in the game is determined by the actual movement in the world, as tracked by a GPS receiver.

1.2 **RESEARCH GOALS AND SCOPE**

The overall motivation of this research is to better understand the role of natural walking in the context of mixed reality and its impact on design. How can we ensure that MR applications will not interfere with the tasks we want to perform in the real world? How can we design to allow useful tasks to be performed while walking? Our goal is to advance research in this area by:

1. Identifying and modelling the tradeoffs of walking during the execution of a cognitively demanding visual analysis task in virtual reality.

2. Presenting the design, implementation, and evaluation of a technique that explores walking to robustly mark distant targets in mobile augmented reality.

3. Presenting the design space, interaction analysis, design, and implementation of mobile augmented reality workspaces.

Walking can influence the design and usability of MR applications in many ways. Similarly, it can be studied and evaluated from different perspectives. This research will be directed to understand the tradeoffs of applications and techniques explicitly designed for walking scenarios. For this reason, we did not focus on evaluating interfaces designed for non-walking conditions, while walking. We also looked into issues related to motor control and comfort but did not focus on fatigue caused by extended use or device
ergonomics, as those will likely change in the future. Regarding the perception and cognitive aspects, we focused on aspects related to visual processes, such as visual memory and attention, since they are most relevant for MR displays.

1.3 Research Questions

Our goals can be summarized by three research questions and respective conjectures:

**RQ1 - What is the effect of walking on cognitive tasks?**

To be able to make conscious design decisions about use of walking, we need to understand how it compares to the alternatives. In VR, a reasonable alternative to walking is using controller-based object manipulation.

**H1a** – Novice users will consider walking easier and more natural than alternative input methods for camera control

Even if walking is not the most efficient technique, we hypothesize that most novice users will prefer walking due to familiarity. We expect that this preference will change as tasks rely more on speed and as users get more familiar with techniques.

**H1b** – Walking will impair tasks that require similar cognitive resources

We hypothesize that walking will have negligible impact on easy tasks, as long as the environment does not present challenges. However, walking will degrade performance in tasks with high cognitive and attentional demands.

**RQ2 - How can we use walking to improve AR interaction techniques?**

In addition to moving in the world, walking can be used in AR as part of interaction techniques. Marking points is a fundamental task in AR, since it can be used to attach annotations, place virtual objects, etc. However, marking points at unknown distances is hard. We expected that by leveraging wide-area capabilities of current AR devices, we could incorporate walking into a marking technique that would allow us to better estimate distances.
RQ3 - How can we design interfaces that support walking?

Using interfaces while walking is difficult due to jitter and divided attention. In addition, walking in AR causes spatial configuration changes when the user moves from one point to another. Current layout techniques for AR are not suitable for this scenario. We expected that designs that adapt to walking conditions and changes in the physical environment could provide better support for walking scenarios.

1.4 Methods and Scope

Due to their different nature, the questions in this dissertation required different methods, which we applied as appropriate for each topic.

The Impact of Walking on Visual Analysis Tasks

To study the tradeoffs of walking for visual analysis, we conducted a controlled experiment to investigate how walking compares to 3D manipulation in a task that requires precise spatial and memory abilities. The task required participants to perform a visual counting task in two separate conditions. In one of them, participants were allowed to walk. In the other, they would stay still and use a 3D controller. We then used regression models to compare the performance in the two conditions. Controlled studies (or laboratory studies) are adequate to test hypotheses and collect precise measurements at the expense of ecological validity (Schmuckler 2001). In our case, we investigated how performance varied between the conditions.

Design and Evaluation of a Walking Based Marking Technique

For this part of the research, we used interaction design methods to find ways to solve the problem and evaluated our designs using controlled experiments. We designed and evaluated several techniques for marking objects at distance. First, we conducted a baseline study to compare a stationary and a walk-based technique. Then we refined the walk-based technique to take into account head tremors during walking. This design was then compared to the initial walking technique. Although reaching the final solution involved a design process, the problem itself was well specified and had clear metrics for evaluation.
DESIGN AND GUIDELINES FOR WALKING-BASED AR WORKSPACES

To study workspaces for AR, we adopted an exploratory qualitative approach. We designed and evaluated an adaptive interface layout for AR that can recognize the surrounding physical space and adapt to it. Next, we conducted a contextual analysis aiming to improve ecological validity instead of controlling all the experimental variables. This decision allowed us to identify the main problems and opportunities in the walking-based AR interfaces, and frame hypotheses that can be verified later through controlled comparisons. We also learned how users appropriated the technology, leading to unexpected insights.

1.5 MAIN CONTRIBUTIONS

With this research, we contribute towards a better design of walk-centric MR user interfaces in the following ways:

UNDERSTANDING THE IMPACT OF WALKING IN VISUALLY DEMANDING TASKS

We provide evidence that users show performance differences between walking-based and controller-based interfaces for visual analysis of 3D datasets. We further show that the alternative with highest performance depends on the spatial ability and game experience of each user. Users with previous game experience but low spatial ability perform better with the controller-based interface. However, we learned that walking offers higher performance for users with high spatial ability or for those without significant game experience.

TECHNIQUE DESIGNS FOR MODEL-FREE POINT SELECTION AT A DISTANCE

We present the design and evaluation of AR techniques to specify distant points in space, without knowledge about the geometry or distance of those points. We first show that geometric methods provide higher accuracy than pure perceptual methods, but that pointing precision is affected by head tremor. We then provide a technique design that mitigates this issue by taking multiple samples to obtain a better estimate of the target position. We demonstrate the effectiveness of this approach in a controlled AR simulation, as well as in actual outdoor AR.
Technique Design and Characterization of Adaptive Workspace Layouts

We present the design and evaluation of adaptive workspaces for walking. To organize the design space, we first introduce a taxonomy for adaptation in AR interfaces. Next, we applied the taxonomy to derive a set of behaviors that can support the use of AR interfaces while walking. Through interactive design and contextual studies, we identify a set of key interaction patterns and desirable properties of adaptation-based techniques. Using the contextual studies, we also show how a minimum set of behaviors can support different walking scenarios.

1.6 The Structure of this Dissertation

This dissertation is organized in the following way. In Chapter 2, we describe our research to understand the effect of walking in visual tasks. To this end, we evaluated the performance of 3D manipulation compared to walking viewpoint control. In Chapter 3, we look into how walking can be used as part of interaction techniques. For the task of marking points in AR, we compared our new technique against other baseline methods. In Chapter 4, we discuss the design of AR workspaces in walking scenarios. Chapter 5 provides a summary of this work and future directions for walk-centric user interfaces.
2 UNDERSTANDING WALKING TRADEOFFS IN VIRTUAL REALITY

Here we describe in detail our study to uncover the effect of walking in complex visual-spatial tasks. We consider walking as the act of physically moving through space, with the goal of acquiring different points of view. Here, this is accomplished by moving the legs, torso, and head in a 6 DOF tracked area. Physical walking is well established as a simple and natural method for controlling the viewpoint and exploring data in many technological contexts. With large display walls, the user can walk closer to see information in more detail or step back to get an overview of the content. In virtual reality (VR), tracking systems enable users to walk to explore virtual worlds. However, since tracked space is always limited, designers often complement walking with other travel techniques. In mobile augmented reality (AR), physical locomotion is the only mode that will maintain the registration between AR constructs and the world. In addition, since it requires little to no training, walking is always welcome and sometimes preferred to controller-based navigation (Ball et al. 2007). Finally, walking can offer additional benefits, for example, it can improve distance estimation and the spatial sense of location (Ruddle et al. 2011).

2.1 INTRODUCTION

Although it seems effortless, it is known that walking requires attentional resources (Yoge-Seligmann et al. 2008b; Lajoie et al. 1993). Moving in space also interferes with tasks such as multiple-object tracking in VR and in the real world (Thomas & Seiffert 2010). Even maintaining body posture can be costly enough to impact performance in memory tasks (Pellecchia 2003; Riley et al. 2012).

In VR, walking is considered the most natural technique for traveling from one point to another (LaViola et al. 2017). However, in many applications, traveling also supports a different primary task, such as finding a correlation between disparate datasets (Donalek
et al. 2014) or building a neurosurgery plan around interweaving blood vessels (Kersten-Oertel et al. 2014). If walking demands cognitive load, limiting the resources available to the user, it may not always be the best interaction choice. A deeper understanding of the issues involved in walking-based interactions can guide designers in the selection between possible alternative techniques.

In this work, we contribute to a better understanding of the tradeoffs between walking and using controller-based interfaces in VR. While previous work exists regarding the benefits of walking in 3D navigation and also in the use of large 2D displays, little is known about its effect in visually demanding 3D tasks. This category of tasks requires users to understand complex spatial structures, and bears high relevance to scientific visualization and visual analytics in VR. With this in mind, we conducted a study to compare the accuracy of participants in two possible scenarios for an immersive visualization application: walking around the dataset vs. using a 3D interaction technique to manipulate it. Our results indicate that neither technique consistently outperforms the other. Instead, relative performance can be predicted by individual differences in spatial ability and game experience.

In summary, the contributions of this work consist of (i) our report on a study directly comparing walking and 3D interaction for a visual analysis task, (ii) a model for user accuracy including game experience and spatial ability, and (iii) our discussion of the implications of these findings in VR interface design.

2.2 Background and Related Work

Spatial ability and Game Experience

Our capacity to perceive, store, and transform spatial data (spatial ability), is generally understood as composed by different factors (Velez et al. 2005; Carroll 1974). Those factors have been assessed using psychometric tests (Carroll 1974; Hogan 2012; Kirby & Boulter 1999) and also through real-world tasks like navigating in a maze (Moffat et al. 1998), remembering the location of objects in space (Piper et al. 2011), and identifying objects from projections (Kirby & Boulter 1999). However, due to the complex and still ill-understood mechanisms involved, spatial ability factors resist a definitive
categorization. Carroll (1974), for example, defines the spatial orientation factor (SO) as the ability to rotate a representation mentally, and spatial visualization (Vz) as the ability to perform the same process in combination with serial operations. On the other hand, evidence has shown that although apparently symmetrical, the ability to rotate an object mentally and to imagine oneself looking from a different perspective, might involve two different spatial abilities (Kozhevnikov & Hegarty 2001).

People demonstrate different performance in spatial ability tasks (Mazman & Altun 2013; Wolbers & Hegarty 2010). This difference can be attributed to any of the mental process involved in the acquisition of visual cues, forming spatial representations, or manipulating them (Wolbers & Hegarty 2010). One question of interest is if differences in spatial abilities can be attributed to gender. Billen (2001) found no difference on mental rotation accuracy between males and females. However, results indicated that females were faster. Piper et al. (2011) found that males were better in a spatial memory test. Hogan (2012) results using ETS psychometric tests were inconclusive, while Mazman & Altun (2013) found that males were more accurate in a perspective taking test, but found no difference in reaction times.

Linn & Petersen (1985), after a meta-analysis of 81 studies, found that gender differences tend to favor males in both mental rotation and in spatial perception, but no difference was found in spatial visualization. Another meta-study focusing on tasks such as pointing, wayfinding, sketching, and distance estimations, uncovered mixed results (Coluccia & Louse 2004). The author’s analysis suggests that gender differences in orientation only appear when the task creates a high load of the visuo-spatial working memory (VSWM). Thus, males would perform better in studies with higher demand due to a larger VSWM. Wang & Carr (2014) reached a similar conclusion, and observed that males perform better in tasks that require holistic strategies, and that this could explain better mental rotation results reported in the literature (Levine et al. 2016). Since a more extensive discussion would fall away from the focus of this text, we refer the interested reader to the meta-studies mentioned above.

Another way to look into individual differences in spatial ability is to see whether they are correlated with experience performing similar tasks. Smith & Du’Mont (2009)
examined the effect of previous game experience on spatial navigation in virtual environments. Evidence was found that perceived gaming skill and how far participants progress in a First-Person Shooter games were correlated with higher performance on the navigation task. Richardson et al. (2011) also investigated the effect of game experience in the navigation performance, also comparing virtual and real environments. Game experience was positively correlated with performance in desktop and VR, but not in the virtual world. The authors speculate that the use of a joystick in the VR condition, instead of real walking, might have caused that difference. In addition, the lack of correlation in the real environments might be related to the use of different perceptual and cognitive abilities. Some evidence points to the understanding that higher visuo-spatial skills of gamers may be come from the practice (Murias et al. 2016; Basak et al. 2008).

**Walking and Data Visualization**

One of the first studies looking to contrast walking and other interaction modalities was performed by Chance et al. (1998). They compared three locomotion modes: walking, using a joystick to rotate and translate, and using a joystick to translate while rotating in place. The task required participants (n=22) to walk through a maze and indicate the direction of objects found along the way when they reached the end. They found that participants in the walking condition demonstrated better direction accuracy than the ones who only used the joystick. Peck et al. (2011) compared redirected walking with walking-in-place and a joystick interface. Participants in the real walking condition traveled shorter distances and were more accurate when pointing directions. The benefits of walking in spatial perception and navigation tasks have also been confirmed by later studies (Ruddle & Lessels 2009; Ruddle et al. 2011; Frissen et al. 2011).

An important question, though, is whether walking can also produce benefits for tasks that do not directly involve navigation. Rädle et al. (2013) compared the effect of egocentric body movements in a task involving visual search. The study asked 24 participants to find corresponding pairs among back facing cards by moving toward one card at a time to reveal symbols in them. The authors compared spatial memory and completion time in two conditions: sitting at a desk and moving in front of the display wall with a tablet. They found that walking decreased the task completion time by 34%
and found no difference in spatial memory right after the experiment. However, when testing recall after 15 minutes, they did find an improvement in spatial memory for the walking condition. The results seem to indicate that physical locomotion helped in the recall of item location.

Liu et al. (2014) studied the effect of physical navigation on a classification task. The task required participants to sort labeled disks into different containers on a display. In the first experiment, 12 participants performed the task by walking along a large display wall and then using a pan-and-zoom navigation technique with a mouse and standard monitor. Among other results, they found that physical walking was faster for both easy and hard tasks with small labels. The opposite was true for easy and hard tasks with large labels. However, in the analysis, the authors point out that task difficulty changed the average distance between items and containers, which may have confounded the results. In addition, since two different platforms were used, it is hard to isolate the effect of walking.

Later, Jakobsen & Hornbæk (2015) repeated the classification task conducted by Liu et al., but this time using the same hardware platform for both conditions. Similar to the original experiment, they labeled the disks and containers on the display wall and asked participants to sort them by labels. In the Move condition, participants could move freely in front of the display. In the NoMove condition, participants were contained in a predefined area of 40 cm in front of the display and used a gyroscopic mouse to pan and zoom. The authors did find that although participants (n=10) were equally fast in tasks which did not require navigation, they were 15% slower when virtual navigation was necessary to read the labels. In both this and the original study by Liu et al., however, the task was not cognitively bounded. The completion time reflects a mixture of the time spent on search, navigation and manipulation. In our case, we used a reasonably difficult task, so that results do depend on how fast participants use the interface.

Büschel et al. (2017) evaluated user strategies in 3D data visualization using a tablet in two conditions: a) with the view spatially mapped to the device position and, b) using the touch surface to control an orbiting camera. They found that participants (n=18) adopted similar strategies in both conditions by moving the camera to appropriate viewpoints.
Participants deemed spatial interaction better in terms of subjective feeling of camera control, ability to accomplish the tasks and mental demand. There were no significant differences in error rates and the touch interface was faster for navigation tasks. Unfortunately, the three tasks employed required little cognitive effort, and were completed in less than 2 minutes.

**Motion in Space**

Cognition studies in the real world have revealed that vision has a deep connection to motor control and spatial processing. Kerr et al. (1985) report an experiment where spatial memory tasks were disrupted just by asking participants to maintain a difficult standing posture. Participants were asked to perform a memory task while sitting or while standing with the heel of the front foot directly ahead the toes of the back foot. Half of the participants were assigned to spatial memory tasks, while the other half performed non-spatial memory tasks. The authors found that the concurrent balance requirement reduced recall for spatial tasks but not for the non-spatial tasks.

Thomas & Seiffert (2010) studied the effect of self-motion in multiple-object tracking. The task required participants to visually track one or three moving balls in a virtual environment. The authors found that performance was impaired whenever participants moved in space. The effect was present even when participants were moved passively in a wheelchair. Although self-motion interfered with object tracking, no interference was found when participants performed a difficult non-spatial tracking task. They conclude that self-motion uses cognitive resources that could otherwise be allocated to the visual tracking task.

Overall, the literature indicates that walking can be beneficial in search and classification tasks on 2D displays. It also seems to agree that walking can benefit tasks involving spatial abilities such as 3D spatial perception and navigation. An open question is whether these effects can also be observed in other types of tasks. In addition, while some studies have investigated and quantified the interference caused by physical locomotion on other tasks, they have not examined the performance tradeoffs involving the use of walking as an interaction technique. The role of spatial ability and user experience also needs further study.
We were intrigued by the trade-offs between the benefits of walking and the cognitive load induced by walking found in prior studies. An earlier unpublished study run in our lab had also suggested that the choice of walking vs. manipulating a scientific dataset during visual analysis may have impacted users’ ability to spatially analyze the data. Therefore, we decided to study how the design choice between walking and manipulation affects the performance in a 3D visualization task. In addition, we decided to investigate the role of individual differences in the results.

2.3 Overall Experimental Design

To investigate the possible tradeoffs of walking, we directly compared the performance of participants in two conditions. In the Walking condition, walking was the only way to see different parts of the dataset; in the Non-Walking condition, participants would stand still and obtain different views by using a 3D tracked controller to manipulate the dataset.

Dataset and Task

In our experiments, we selected a visualization task that required participants to analyze complex tridimensional models. We were inspired by scientists and practitioners that study 3D datasets derived from Computed Tomography (CT) or Magnetic Resonance Imaging (MRI) of biological structures. The models consisted of a central spherical shape surrounded by tubular branches; a structure similar to one commonly found in the respiratory system of beetles.

The task consisted of verbally reporting if a given model had a specific number of branches connected to the central spherical surface. Instead of using data from a real CT-Scan, we designed a procedural method to generate different versions of the dataset given parameters such as total number of branches, number of connected branches, and number of branch layers. In this way, we could ensure that: 1) there was no ambiguity in the model that would require domain expertise; 2) there was no noise resulting from imperfect segmentation of the original volumetric data; and 3) that we could easily generate datasets with different levels of difficulty.

For the experiments, we generated two different groups of datasets: datasets in the Simple group had 16 branches, of which 4, 5, or 6 could be connected. Datasets in the Complex
group had 30 branches, from which 11, 12, or 13 could be connected to the central sphere. For the Simple datasets, participants were asked if the model had 5 connected branches or not. For the Complex datasets, they were asked if the model had 12 connected branches or not. Each group contained an equal number of datasets whose correct answers were yes and no.

The extra (non-connected) branches in each model ensured that participants would need to change the point of view to avoid occlusion. Branch connections were made in the last step of the generation process so that the number of connections did not change the overall look of the datasets. The synthetic model structure was evaluated by a domain expert to ensure that the model presented similar topology and challenges as the real one. Figure 2-1 shows an example of each type.

![Figure 2-1: Examples of datasets from the Complex (left) and Simple (right) groups.](image)

The final parameters of each dataset were found by trial and error. For the Simple dataset, the goal was to make sure that the task could not be solved by *subtizing*\(^1\) and yet did not present a very hard challenge. For the Complex dataset, we ensured that models were difficult enough to be cognitively bounded (i.e., that the time needed to answer the question correctly depended on cognitive limits rather than perceptual or motor performance). In this way, results would not depend on how fast one could walk around or rotate the dataset, but only on cognitive load.

---

\(^1\) Process which allows humans to quickly determine the numerosity of a small set of objects. When a small number of objects is small (less than 6), their number can be determined at a glance (Mandler & Shebo 1982; Kaufman et al. 1949).
ENVIRONMENT AND APPARATUS

The experiments were conducted in a closed room, free of distractions. We used a consumer version HTC Vive head-mounted display (HMD) and the bundled controller. The HMD has two screens, which were adjusted to a typical interpupillary distance (IPD) of 67mm (we did not adjust the IPD individually for each participant due to mechanical problems with the HMD). Each screen has a resolution of 1080x1200 pixels. The horizontal field of view is 110 degrees. The HMD and controller were tracked with six degrees-of-freedom by the hybrid inertial-optical Lighthouse system. To provide freedom to the participants, the HMD was driven in both conditions by an MSI VR One backpack running Windows 10. The backpack is an integrated PC with an Intel quad-core processor, 16Gb of RAM and an nVidia GTX 1060 graphics card. Participants used the trigger button on the controller to switch between datasets and the touchpad button to grab the dataset for manipulation (grabbing was disabled during the walking condition). The software was written in Unity3D and used the SteamVR plugin. The experimenter observed the participants and monitored task execution using a remote desktop access software (TeamViewer).

We wanted to ensure that participants would walk around the dataset in the Walking condition, but we did not want them to have an unfair advantage by going inside the dataset. For this reason, we established maximum and minimum distances that users could be from the dataset. Participants in the Walking condition were free to walk or move in any way desired as long as they stayed within a delimited area. The inner radius of the walking area was 0.6 meters and the outer radius 1.5 meters (Figure 2-2, in green). If participants moved their heads outside this area, the dataset disappeared rapidly to prevent them from obtaining extra information. Participants in the Non-Walking condition were asked to choose their preferred distance inside the same area. Note that head tracking was still enabled in the Non-Walking condition, but participants were asked to stand in one location.

Besides the limits of the allowed area, the experimental environment only showed the limits of the tracking area and a blue/brown skybox. To improve visualization, the model
had diffuse and direct lighting with self-shadows. However, no shadows were projected on the floor.

Figure 2-2: View of the Experimental Environment. The models appeared in the center of the red area. In the Walking condition, participants could move freely inside the green area. The dataset would disappear if users moved outside this area.

**Metrics**

The time/accuracy tradeoff is common in many experimental designs. If not controlled, it makes analysis and interpretation potentially complex. To avoid this tradeoff (so that our results would not be biased by individual preferences), we adopted the following procedure. Participants were asked to achieve a pre-determined number of correct answers (3 in the first experiment and 4 in the second experiment), and we measured the time they took to achieve this goal. In this way, a participant that was extra careful and spent more time per task would not be penalized when compared to another that was not as careful, but had more errors. This is similar to the metric used in many psychometric tests, and it was not necessary to negatively weight the wrong answers (since we considered the total time spent). Participants used a button on the handheld controller to control the pace of the experiment. The first press would make the dataset appear (and start the timer) and a second press would make it disappear (consequently stopping the timer).
2.4 **Pilot Study**

We ran a pilot study to select a *Non-Walking* technique with good performance to compare with walking. We evaluated two techniques: *Rotate* and *Grab* (we also had a Walking condition in this study, but its purpose was only to pilot the procedure for this condition; we do not consider it further in this section).

**Rotate Technique**

The Rotate technique followed a manipulation metaphor. Participants could rotate the dataset by using the tracked controller as a tangible proxy for the dataset. The controller orientation was mapped to dataset orientation, allowing direct and integral control of all 3 rotational degrees of freedom. The dataset position was fixed in the center of the environment, so participants were free to hold the controller in a comfortable position. We expected that this technique would allow participants to complement the visual information with motor and kinesthetic information from the controller orientation. Clutching was not allowed.

**Grab Technique**

The *Grab* technique used a direct manipulation metaphor. Participants could use the controller to directly grab and rotate the dataset around its center. As before, the dataset position was fixed in the center of the environment and only rotated around its center. To rotate, participants could reach to any point in space (not only the points with branches) and grab the dataset by pressing the trigger button. The dataset would then follow the angular rotation of the hand around the center of the model while the button was pressed. Upon release, the dataset would stay at the new orientation. As with the previous technique, Grab allowed an integral control of all 3 rotational degrees of freedom.

Our hypothesis was that *Rotate* technique would offer better performance than *Grab* since it was more intuitive, did not require clutching and supported precision two-handed manipulation. An earlier study by Hinckley et al. (1997) supported this hypothesis; where they found that directly rotating an object via a tracked proxy was superior to a mouse-based Virtual Sphere manipulation technique.
**Participants**

We recruited 12 university students (9 females) from 19 to 32 (M=24.17, SD=3.85) years old. All the participants were screened for stereo blindness using a random dot stereogram and had corrected or normal vision. Only four participants had used VR more than twice. The experiment was approved by the Virginia Tech Institutional Review Board and all participants gave written informed consent.

**Procedure**

Prior to the study participants were asked to fill out a background questionnaire containing questions about age, occupation, handedness, eye correction, VR experience, and computer experience. After that, they were instructed about the experimental procedure, techniques, and experimental environment (including the physical area where the experiment took place). The experimenter then helped the participants to wear the backpack and headset. They were then asked to walk around the experiment area. After this, the experimenter would ask participants to look straight ahead and would use the height of the head, as given by the tracking system, to adjust the height where the models would appear. This ensured that the connection points on the top of the models were always visible and below eye level. Before each condition, the experimenter ran a few practice trials until participants reported being confident with the techniques, the two dataset shapes, and demonstrated to correctly understand the procedure. Between each condition, participants were given the opportunity to stop and rest for a few minutes. In both conditions, the participants were asked to choose a spot within the 0.9 m green band around the center and to stand at that location for the entire set of trials. For each technique, participants were asked to obtain three correct responses with Simple datasets and three correct responses with Complex datasets. The answer was given verbally (yes/no). We did not use a time limit, but asked participants to be as accurate and fast as possible. The presentation order of the two techniques was counterbalanced. In the end, they answered a post-experiment questionnaire, asking about their experience with the techniques and preferences. The experimental material for this experiment can be found in Appendix B.
RESULTS

Contrary to our expectation, participants using the *Rotate* technique took longer to complete the task with Complex datasets (M = 580s, SD = 289) than those using the controller with the *Grab* technique (M = 360s, SD = 223). Based on a Wilcoxon Signed Rank Test, the difference was significant for the Complex dataset (V=68, p=0.021), but not for the Simple dataset (V=25, p=0.301) (Wilcoxon Signed Rank Test).

Eleven participants rated the *Grab* technique as being “easy”, while eight said the same for the *Rotate* technique. Most participants deemed the *Grab* technique to be more accurate and the *Rotate* technique faster. We noticed that most successful strategies consisted of scanning trajectories that lay on the surface of a sphere. The *Grab* technique allows users to directly “draw” those trajectories. Thus, it provides a better mapping than controlling a surface point based on rotation. The trajectory is also more stable, since the one-to-one mapping in *Rotate* transforms small angular errors into larger displacements (proportional to the radius of the dataset). Additionally, *Grab* allows “parking” when the biomechanical limits of the hand are reached, which is not possible with *Rotate*. Using the latter required participants to keep their hands steady while investigating the dataset. Based on completion time, participant feedback, and the considerations above, we decided to use the *Grab* technique in the *Non-Walking* condition of our main experiment.

2.5 MAIN EXPERIMENT

We ran a within-subjects design comparing walking and non-walking conditions. We used the same dataset and tasks used in the pilot study (section 3.1).

WALKING CONDITION

In this condition, the dataset was fixed in the center of the environment, requiring participants to walk around the dataset to obtain the desired viewpoints. However, the dataset was only visible within a green area defined on floor (Figure 2-2), preventing participants from walking inside the dataset or moving too far away. The controller was only used to move from one trial to the next.
**Non-Walking Condition**

In the *Non-Walking* condition, the participant would stand still at a chosen spot within the green area and use the controller to rotate the dataset. For this task, we selected the *Grab* technique, since it produced higher performance than the *Rotate* technique in pilot study. As before, participants could reach any point in space (not only the points with branches) and grab the dataset by pressing the trigger button. The dataset would then follow the angular rotation of the hand around the center of the model.

As before, we measured the performance for both simple and complex models. We also made sure that both conditions could be easily completed with the simple models. To complete successfully the tasks, participants needed to view the dataset from different sides. However, some viewpoints could only be obtained in the Non-Walking condition (e.g., a view perpendicular to the top). We attempted to minimize this disadvantage by adjusting the height of the model based on the height of each participant, so that all connection points could be seen.

We hypothesized that participants in the *Walking* condition would outperform those in the *Non-Walking* condition. We also expected that users with high spatial ability and more experience with games would perform the task more quickly.

**Spatial Ability and Game Experience Measures**

We collected data about two other factors that might influence the performance of the participants: spatial ability and game experience. Spatial ability can involve different skills, which can be measured by cognitive tests. Velez et al. (2005) found a positive correlation between accuracy in a visualization task and scores on the ETS (Ekstrom et al. 1976) Cube Comparison (S-2) and Paper Folding (VZ-2) Tests. The S-2 is a test where participants are asked to determine if two cubes are different, based on two different views. It has two sections with 21 questions. The VZ-2 test asks participants to select the correct result after a paper is folded five times and then punched. It consists of two parts with 10 questions. The total score was computed as the number of correct answers minus the number of wrong answers. We administered the first part of S-2 and the full VZ-2 test. The S-2 test was included as a reference, since mental rotation was used as a measure of spatial ability in a variety of earlier studies of interaction in virtual environments.
reality (Laha et al. 2012; Moffat et al. 1998; Parsons et al. 2013; Astur et al. 2016; Jang et al. 2016). According to Carroll (1974), both tests involve measuring factors related to the capacity of the visual short term memory (STM). However, the VZ-2 also captures aspects of the central executive involved in performing serial operations, which we noticed was required by our task.

Since one of the conditions required the use of a handheld input device, prior experience could also be relevant. Previous research indicates that game players have a variety of skills such as higher efficiency when switching between tasks (Shawn Green et al. 2012), enhanced attention (Dye et al. 2009), better memory (Boot et al. 2008), and better coordination (Griffith et al. 1983). Prior game experience was self-reported in the background questionnaire. Participants were asked to rate their experience on a 7-point scale ranging from 1 (Never) to 7 (Everyday) to the question: “How often do you play videogames (on computers or consoles like Xbox, Nintendo, etc.)?"

We expected that together, both the visualization score and the game experience variable would capture a large part of the individual differences important for our task. Among them, possible differences in spatial ability due to gender, handedness, and also indirect aspects of playing video game like improved attention and coordination. In addition, to reduce the error variance associated with individual differences not accounted by those covariates, we chose a within-subjects design. The presentation order was counterbalanced to prevent carry-over effects.

**Participants**

We recruited 37 university students. Five participants were not able to complete the tasks in both conditions in the allotted time and were excluded. The remaining participants’ ages ranged from 18 to 44 (M=26.22, SD=6.56) years old (6 females). Two-thirds had limited experience with VR (fewer than three prior experiences). All the participants were screened for stereo blindness using a random dot stereogram and had corrected or normal vision. The experiment was approved by the Virginia Tech Institutional Review Board and all participants gave written informed consent.
Prior to the experiment, participants were asked to complete the Paper Folding Test (ETS-VZ2), the Cube Rotation Test (ETS-S2), and a background questionnaire with demographic information (age, occupation, handedness, eye correction, VR experience, computer experience) and a specific question about gaming experience. After that, they were instructed about the experimental procedure, techniques, and experimental environment (including the physical area where the experiment took place). To minimize effects of learning, we also briefed the participants on common ways to deal with the complexity of the dataset, as observed from the first experiment. They were: looking for useful reference points to start and stop counting, dividing the dataset into parts, and counting groups of branches together. For each technique, participants were asked to obtain 4 correct responses with Simple datasets and 4 correct responses with Complex datasets. The answer was given verbally (yes/no). Participants were instructed to press the trigger button to stop the timer before answering. To avoid extending the experiment for too long, they were told there would be a time limit to complete the task: 5 minutes for the simple datasets and 15 minutes for the complex ones. The experimenter then helped the participants to wear the backpack and the headset. They were asked to walk around the experiment area to familiarize themselves with the space. After this, the experimenter would ask participants to look straight ahead and would use the height of the head, as given by the tracking system, to adjust the height where the models would appear. This ensured that the connection points on the top of the models were always below eye level. Before each condition the experimenter ran at least two trials with each dataset and checked if the participants understood the task and techniques. If they had not demonstrated a correct understanding of the procedure, one or two examples were used for training. The goal of the training was to ensure participants understood the task and were comfortable with the technology. It also helped to level out the participants prior to the beginning of the measured tasks. In the Grab condition, the participants were asked to choose a spot within the 0.9m green band around the center and to stand at that location for the entire set of trials. Between each condition, participants were given the opportunity to stop and rest for a few minutes. The presentation order of the two techniques was counterbalanced. In the end, they answered a post-experiment
questionnaire, asking about their experience with the techniques and preferences. The experimental material for this experiment can be found in Appendix B.

RESULTS
Before fitting a regression model including the measured covariates, we performed a preliminary analysis of the results. Table 2-1 lists the descriptive statistics of task completion times. The mean completion time for the Simple dataset was significantly lower than the Complex one, as designed (F (1,90) = 242, p < 0.0001) and the average completion time for both conditions was similar. A visual analysis, however, revealed that the fastest condition was not the same for every user (Figure 2-3). This suggested that our sample was composed of subpopulations for which the relative performance in each condition varied. A within-subjects ANOVA for the two conditions, as expected, did not reveal a significant difference between the overall means of the Walking and Non-Walking conditions for each dataset.

Table 2-1: Mean times and standard deviations for the main experiment.
There was no significant difference between the two conditions

<table>
<thead>
<tr>
<th></th>
<th>Walking (s) (SD)</th>
<th>Non-Walking (s) (SD)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Simple</td>
<td>141 (52)</td>
<td>151 (101)</td>
</tr>
<tr>
<td>Complex</td>
<td>623 (265)</td>
<td>714 (254)</td>
</tr>
<tr>
<td>Total</td>
<td>382 (308)</td>
<td>433 (342)</td>
</tr>
</tbody>
</table>

Following our initial hypothesis, we looked into whether prior experience and spatial ability were confounding the effect of the experimental conditions. Figure 2-4 shows four subsets of the original data for the complex dataset, separated according to whether participants had values higher or lower than the mean in game experience and in the spatial visualization test score (VZ-2).

Participants with high game experience and low spatial ability had better performance using the Grab technique (Figure 2-4, bottom left). Participants with low game experience had better performance with Walking (Figure 2-4, top). Participants with high game experience and high spatial ability had no clear trend. A similar result appeared when considering the S-2 score in place of VZ-2. In fact, a Spearman’s test indicated a
moderate positive correlation between the two ($\rho=0.51$, $p < 9^{10}$). Table 2-2 lists the distribution of participants according to game experience and spatial ability.

![Graph showing time spent by each participant in each condition for simple and complex conditions]

**Figure 2-3:** Task completion time for each participant in the two conditions. Left: Simple dataset, Right: Complex datasets.

![Graphs showing participants grouped according to high (H) and low (L) scores]

**Figure 2-4:** Participants grouped according to high (H) and low (L) scores. GX: game experience, AB: spatial ability. Dashed lines indicate the mean tendency in the group.
Table 2-2. Participant distribution. Number of participants distributed according to the mean values of spatial ability and game experience (number of females in parenthesis).

<table>
<thead>
<tr>
<th></th>
<th>High Game Experience</th>
<th>Low Game Experience</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>High Ability</td>
<td>11 (1)</td>
<td>7 (1)</td>
<td>18 (2)</td>
</tr>
<tr>
<td>Low Ability</td>
<td>5 (1)</td>
<td>8 (3)</td>
<td>13 (4)</td>
</tr>
<tr>
<td>Total</td>
<td>16 (2)</td>
<td>15 (4)</td>
<td>31 (6)</td>
</tr>
</tbody>
</table>

**Analysis**

To analyze the data for statistical significance, we fitted two robust linear regression models: one on the time differences between *Walking* and *Non-Walking* conditions and another one on walking time. We used a robust regression to avoid arbitrarily removing outliers and to avoid compromising the validity of the results by departures from the ordinary least squares assumptions (Wilcox 2011). We used the `lmrob` function from the R Robustbase package (Maechler et al. 2017), with the default MM-regression estimate.

Our model included main effects and interactions for model complexity, game experience, and spatial ability. We checked for multicollinearity in the model by computing variance inflation factors (VIF) with the package olsrr (all factors were below 1.2). We selected the VZ-2 score as a measure of spatial ability, since it was applied in full and provided slightly higher explanatory power. The score was centralized in the median (Table 2-3).

**Table 2-3. Results of the cognitive factors tests. VZ-2 Spatial Visualization Test, S-2 Spatial Orientation Test (N=31).**

<table>
<thead>
<tr>
<th>Test</th>
<th>Min</th>
<th>Median</th>
<th>Max</th>
</tr>
</thead>
<tbody>
<tr>
<td>Spatial Visualization (VZ-2)</td>
<td>5</td>
<td>16</td>
<td>20</td>
</tr>
<tr>
<td>Spatial Orientation (S-2)</td>
<td>0</td>
<td>11</td>
<td>19</td>
</tr>
</tbody>
</table>
We also checked for outliers between the covariates. One observation was excluded based on robust Mahalanobis distance. A Spearman’s rank test failed to reject the assumption of zero correlation between the covariates at a 0.05 significance level ($p = 0.3179, \rho = 0.09$). The full model also had a significantly better fit than incomplete models with only one of the covariates (robust Wald test, $p < 0.05$). The adjusted $R^2$ for the final time difference model was 0.61, and the residual standard error was 105.6. The adjusted $R^2$ for the walking time model was 0.88, and the residual standard error was 76.5. The significant terms of both models are listed in Table 2-4 and Table 2-5, along with the associated coefficients. Of particular interest are the interactions between spatial ability and game experience.

<table>
<thead>
<tr>
<th>Term</th>
<th>$\chi^2$</th>
<th>p</th>
<th>Coef</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>18.5</td>
<td>&lt;0.001</td>
<td>-220.9</td>
</tr>
<tr>
<td>GameExp</td>
<td>39.0</td>
<td>&lt;0.001</td>
<td>54</td>
</tr>
<tr>
<td>Complexity</td>
<td>17.9</td>
<td>&lt;0.001</td>
<td>-312.9</td>
</tr>
<tr>
<td>GameExp:Ability</td>
<td>10.7</td>
<td>0.001</td>
<td>-9.0</td>
</tr>
<tr>
<td>GameExp:Complexity</td>
<td>36.42</td>
<td>&lt;0.001</td>
<td>76.0</td>
</tr>
</tbody>
</table>

Table 2-4. Significant effects for the time difference. Terms of the regression model for the time difference between the two conditions.

On the time difference model, game experience has a significant positive coefficient (effect 2 in Table 2-4), indicating that higher levels of game experience decrease the mean completion time when using the Grab technique. The combination of high game experience and the complex dataset also favors the Grab technique (Table 2-4, effect 5). However, the intercept and complexity have large negative coefficients, implying that Walking is faster overall and faster with the complex dataset across all participants (Table 2-4, effects 1 and 3). In addition, ability and game experience combined have a negative coefficient, indicating that walking is superior to Grab for participants with higher levels of both. Figure 2-5 shows the time differences predicted by the model for
the Complex dataset\(^2\). Points above zero indicate that the Grab technique is faster, while points below indicate that Walking was faster.

Table 2-5. Significant effects of the regression model for walking time.

<table>
<thead>
<tr>
<th>Term</th>
<th>(\chi^2)</th>
<th>p</th>
<th>Coef</th>
</tr>
</thead>
<tbody>
<tr>
<td>1  Intercept</td>
<td>79.4</td>
<td>&lt;0.001</td>
<td>274.4</td>
</tr>
<tr>
<td>2  Complexity</td>
<td>14.0</td>
<td>0.0002</td>
<td>171.8</td>
</tr>
<tr>
<td>3  GameExp:Complexity</td>
<td>19.1</td>
<td>&lt;0.001</td>
<td>30.43</td>
</tr>
<tr>
<td>4  Complexity:Ability</td>
<td>11.3</td>
<td>0.0007</td>
<td>-27.3</td>
</tr>
</tbody>
</table>

For participants with low game experience, walking is clearly better. However, the difference gets smaller with the increase in spatial ability. For participants with high game experience, the opposite is true: grabbing is advantageous for users with low spatial ability, but the difference reduces with increase in spatial ability. Overall, participants with high spatial ability performed approximately the same with both techniques.

Looking at the absolute time model, we see that time spent during the walking condition increases with complexity (Table 2-5, effects 2 and 3). However, spatial ability decreases the mean time spent walking by 27 seconds for each additional point. The total effect is counterbalanced by game experience which lessens the impact of spatial ability. The different resulting slopes are shown in Figure 2-6.

When asked to choose between the two conditions, 64% of the participants preferred Walking. We also asked participants if they thought walking helped or interfered with the task. Approximately 80% stated that walking was helpful. When we compared their preference with the performance in both conditions we found that in only 61% percent of the cases did the condition with the highest performance match their preferences. We also asked participants to rate the two conditions regarding speed, comfort and on how easy it was to obtain the desired viewpoints. There was no clear preference among the participants regarding any of those characteristics.

\(^2\) The plots were done using the R package visreg and include Wald 95% conditional confidence intervals. Partial residuals were also plotted to aid assessment of variability and impact of any outliers (Breheny & Burchett 2017).
Figure 2-5: Time Difference Between Walking and Non-Walking Conditions. Participants with low spatial ability and game experience performed better walking. Participants with low spatial ability and high game experience performed better with the controller. Users with high spatial ability tend to perform equally on both.

Figure 2-6: Time in the Walking Condition. Spatial ability improved the time of all participants, regardless of game experience.
**DISCUSSION**

In this experiment, we directly compared the performance of walking and non-walking techniques in a complex visual analysis task. Among the results, we highlight a few key findings:

**The relative performance of interfaces based on walking and manipulation depend on individual differences.** We found that the differences between *Walking* and *Non-Walking* conditions can be very large (up to 40%) and that they can be explained by specific combinations of spatial ability and game experience. In some applications, looking at the behavior of specific groups of users instead of the whole population might give a more accurate picture of interface usability.

**Walking provides significant performance advantages for users with mid/low game experience.** Not everyone is acquainted with input devices or possesses the ability to easily control multiple degrees of freedom. For users who did not report practice with games, interfaces based on walking provided better performance.

**Spatial ability is important to achieve high performance in walking-based interfaces.** Spatial ability was positively correlated with faster times in the *Walking* condition. With high spatial ability, participants walking and grabbing had similar performance. But for users with low spatial ability and high game experience, manipulation was faster than walking.

These findings seem to indicate that, for cognitively difficult tasks, the lack of some abilities interferes with both *Walking* and *Non-Walking* performance. For the *Walking* condition, the most important factor is spatial ability, while in the *Non-Walking* condition, game experience is more relevant. However, since users have different abilities, we see individual tradeoffs for each technique.

**Walking and manipulation abilities**

The positive effect of spatial ability on the completion time for the *Walking* condition (Figure 2-6) likely reflects a general characteristic of the task. In our experiment, the ability to remember the starting location and the parts of the dataset that had already been
counted was clearly important. Although there was no need to explicitly perform mental rotations (as in the S-2 test), higher spatial ability might help to integrate several partial viewpoints (Kozhevnikov & Hegarty 2001). As the participant’s relative position changed, it was critical to correctly differentiate whether a specific branch was a new one or the same one slightly rotated. We also noticed during the interviews, that the ability to unfold the sphere surface, perform serial operations (such as divide the model into sub-areas), and adapt the strategy to each model, was beneficial to completion time. These abilities, in turn, are very similar to the ones measured by the spatial visualization test (VZ-2).

Interestingly, the impact of spatial ability was not equal, and had a more pronounced effect in the Walking condition. One might think that the opposite should be true, since walking happens with no apparent effort, and since high spatial ability seems more useful when the user is facing a harder situation. This apparent incongruence was found in other studies and could be explained by positing that the act of walking consumes finite resources shared by whatever mechanism is responsible for updating a person’s position in space (Thomas & Seiffert 2010).

On the other hand, the benefit of higher game experience in the Non-Walking condition seems clear: when users need to spend extra effort to operate the interface, they consume mental resources that could have been used to perform the task. For example, in the Rotate technique used in the pilot study, the tangible representation was spatially disjoint from the visual representation. Some participants found that this was not intuitive and that it was harder than manipulating a real object. Participants with more game experience have been trained to operate non-trivial combinations of input devices and interaction techniques, and therefore have more cognitive capacity available for the spatial task.

We also found that game experience did not significantly affect the mean time in the Walking condition. The lack of any evident impact on walking time indicates that the game experience variable is measuring skills that are distinct from those captured by the spatial ability tests. Among the skills frequently correlated with game experience (task switching (Shawn Green et al. 2012), enhanced attention (Dye et al. 2009), better
working and short term memory (Boot et al. 2008), and better coordination (Griffith et al. 1983), hand-eye coordination seems to be only the one not directly related to the central executive and more likely to be developed in players of fast-paced games (Morris & Jones 1990). Although we did not discriminate the participants by game genre and playing history, the question seemed to be enough to distinguish between players and non-players. In any case, the participants in the “player” group demonstrated superior performance in the non-walking condition.

Our conclusion is that the mechanisms that influence performance in each condition are distinct, even though they work in concert due to the individual capacities available for each type of interaction. That is not to say that they do not interact with each other, but rather that they are qualitatively distinct. The manipulation interface is more difficult to learn, resulting in cognitive resources being divided between the interface and the task, but users with training can largely avoid this cognitive cost. Walking is an interface that almost everyone can use easily, but it still introduces a penalty on spatial resources that results in lower performance by users with low spatial ability.

Choosing the appropriate technique

How should designers approach the tradeoffs between walking-based and manipulation-based interfaces? An easy alternative would be to offer both options and let users choose which technique to use. We found, however, that personal preferences only matched the highest performing technique in 61% of the cases. Carefully characterizing the target user base and considering usage scenarios might be a better option. One way to identify the type of users would be to ask about gaming experience and/or apply spatial ability tests. If there is enough homogeneity, it may be possible to choose the method that offers higher average performance. If users can be trained, have high spatial ability, or are going to use the interface for a long time, it might not be worthwhile to go to the effort of providing a large tracked space for walking.

The decision also depends on the task difficulty. In our study, we were careful to separate the problems into light and heavy cognitive load. Although prior work has shown benefits of physical navigation, those experiments did not typically have very high cognitive load.
In those cases, the performance is likely to depend on mobility and dexterity, factors that are less relevant when solving complex problems. The effect of increasing the task complexity is similar to reducing spatial ability: the cost of walking will appear sooner and will be more relevant.

LIMITATIONS

The task we studied was relatively straightforward, although cognitively demanding. In our case, the difficulty was mainly caused by the need to mentally keep track of the areas already visited in a complex structure. In real visualization scenarios, similar requirements are likely to appear with the addition of other concurrent demands, such as dealing with the software interface or crafting more sophisticated reasoning about the data. Most useful applications will require more than counting from one to thirteen. It would be interesting to see how more complex tasks and concurrent demands interact with spatial ability and walking.

Our experiment also provided limited explanatory power of why the difference between the two conditions exists. We cannot be certain, for example, if the difference we observed was due to the act of moving the legs, due to spatial updating, or due to the increased freedom afforded by the tracking. It would be possible to investigate the underlying mechanism by exchanging some of the ecological validity for additional experimental control.

In addition, it is also reasonable to imagine that walking itself can be challenging in some situations. A user performing a similar task at a construction site would have a much lower walking performance. Our study used a reasonably safe environment, with no distractors beyond the dataset. We expect that in complex tasks like the examples above, walking performance will be further degraded by the extra overhead on spatial cognitive processing. However further studies are needed to examine this hypothesis.

Finally, we have investigated only two of many parameters that might affect performance in visually demanding tasks. Although the model including spatial ability and game experience was able to explain much of the variance, more sophisticated models might be able to better predict the actual performance. For example, spatial memory could explain
a higher percentage of the performance variation than VZ-2 test score we employed. We could also have included other covariates such as gender and handedness to further reduce variance. However, additional variables need careful consideration since they would increase the complexity of the model and increase the risk of multicollinearity. Similarly, our assessment of game experience could be improved by applying more practical tests or by using an improved questionnaire gathering details about the genres of games and playing history.

2.6 Conclusion

As walking becomes more common in AR and VR interfaces, it is increasingly important to understand the tradeoffs involved in physical locomotion. In particular, it is critical to know when it makes sense to choose walking over a more sophisticated interaction technique. In this work, we have presented the results of a study evaluating the performance of a complex visual analysis task in two conditions: walking and manipulating. Our analysis revealed that the relative performance on this task is not consistent across individuals and depend on game experience and spatial ability. Our study further shows that walking can enable higher relative performance for users with low spatial ability and low game experience. However, users with high game experience can perform better with non-walking interfaces, especially if they lack spatial ability. We discussed our results in the light of previous findings and argued that optimal design decisions for this tradeoff should consider the role of training and individual differences.
3 MODEL-FREE MARKING IN AUGMENTED REALITY

In this chapter, we describe our design and evaluation of a walking-based technique to mark points in AR. Walking is explored as a fundamental part of the interaction, in this case, to acquire different viewpoints used for marking. We also introduce and evaluate a method to reduce jitter caused by head movement.

3.1 INTRODUCTION

Being able to mark 3D points is an essential task in content creation applications for Augmented Reality (AR). Points in the real world can be used to attach annotations, place virtual objects, delimit areas and arrange user interface elements. For example, firefighters can use AR to mark secure rescue paths in an emergency situation or to indicate the locations of civilians. In a similar way, architects can measure distances or areas for new buildings using onsite physical landmarks. The point marking task in these examples is conceptually different from a selection task in 3D user interface (LaViola et al. 2017), in the sense that the point may not exist as a defined object of which the system has a priori knowledge. In fact, the target location can be a point or feature on a large object or even a point in empty space. Assuming that the user has a visual representation (marker) that he wants to place in a specific location, the most intuitive approach is to directly indicate the 3D position using the hands or a tracked device. When the point is out of reach, an easy alternative would be to manipulate the marker until it reaches the desired position, but this is only possible if the user can accurately perceive the position of the marker.

When the AR device has geometry reconstruction capabilities, then marking locations on surfaces can be as simple as a ray cast from the user reference frame to the target position. The target point can then be determined by computing the intersection between the ray and the reconstructed geometry. However, geometric information about the
environment is not always available. Short-distance reconstruction can fail without proper illumination or with surfaces that have extreme reflective indices. Long-distance reconstruction is even trickier since most depth-mapping sensors have limited range. But marking distant outdoor locations can be of critical importance in mobile AR applications. Some systems address this limitation by using other sources such as offline models or aerial photographs (Höllerer et al. 2007; Arth et al. 2015; Arth et al. 2015). However, these solutions are limited to specific places (where such extra information is available) and time (since real-world geometry will eventually change).

In this research, we focus on techniques for marking 3D locations at medium/far distances without any environment information. We consider and evaluate instances of two classes of techniques: those based on human depth perception and those which employ geometric algorithms to compute the desired location. We conducted both controlled and qualitative studies to understand the properties and limitations of such techniques in both ecologically valid outdoor settings using an AR display and by simulating AR in a virtual reality (VR) system. Our analysis of these techniques led to the design and implementation of VectorCloud, a multi-sample geometric approach. Our user study revealed that this technique can be used for fast and accurate marking of targets without the need for any information about the environment.

In the next section, we discuss related work. Then, we describe the task in more detail and discuss the results of the user studies. Additionally, we present alternative designs we have explored to improve other aspects of geometric techniques. We conclude by summarizing our findings and presenting suggestions for future work.

3.2 BACKGROUND AND RELATED WORK

The performance of point marking techniques in AR depend on several factors such as visual perception, hardware capabilities, input devices, interaction techniques, and environmental conditions. Here we review some relevant work and theories.

DEPTH PERCEPTION AND PERCEPTION-BASED MARKING TECHNIQUES

Human ability to estimate egocentric distances varies depending on the distance being estimated and the method used for estimation. Cutting & Vishton (1995) divided the
perceptual space into three areas: personal space (under 2 meters), action space (up to 30 meters), and vista space (beyond 30 meters). Judgment in personal space is very good since several sources of information can be used, in particular retinal disparity, convergence, and accommodation. In the action space, the usefulness of those is greatly reduced. After 30 meters, we must rely heavily on less efficient non-stereoscopically information, such as occlusion, relative size, height in the visual field and atmospheric effects.

AR devices also add additional perceptual challenges. Without environment information, AR devices cannot properly render occlusion relationships with the real world, which would otherwise be a dominant depth cue to resolve distance ambiguities. Depth perception can also be distorted by the accommodation-vergence conflict, which occurs when the eyes converge at a virtual point but must focus at a screen located at a different distance. These and other issues are discussed in depth by Kruijff et al. (2010).

Studies of egocentric distance perception in mixed reality have reached mixed conclusions. Jones et al. (2008), comparing distances from 8 to 10 meters, found a small underestimation in VR and observed no effect in a see-through AR display. Participants indicated the position through blind walking. Few studies have looked into AR depth perception issues in vista space, however the large variety of methods, environments and distances make results difficult to compare. Dey et al. (2010) evaluated depth judgments outdoors, at distances ranging from 70 to 95 meters. Using a handheld device with video see-through, they found that participants’ verbal judgments underestimated the true distance. Swan et al. (2006) investigated indoor depth judgments in see-through AR displays from 5.25 to 44.31 meters. They asked participants to match virtual and physical targets and found that they underestimated distances smaller than 23 meters and overestimated beyond that. Later, a similar experiment found overestimation indoors and overestimation outdoors with targets between 4.83 m and 38.64 m (Livingston et al. 2009). In our experiment, targets extended more into vista space, ranging from 12.5 m to 85.5 m.
3.3 **Geometry-based Marking Techniques**

One of the first mobile applications for outdoor AR content creation was the Tinmith-Metro modelling system (Piekarski & Thomas 2003). It allowed users to sketch large architectural structures (such as buildings) using techniques inspired by CAD applications. Users could create working planes, which could be specified without reference to any geometric information from the environment. For example, a user might create a plane containing the view direction vector and the gravity vector. Once created, users could mark points on the plane (e.g., by intersecting the plane with a ray) and manipulate existing objects along the plane (Piekarski & Thomas 2004). This approach is similar in spirit to our *Geometric* techniques.

Polvi et al. (2016) present SlidAR, a system to mark points in AR. Similar to ours, it was designed to work without depth information and uses Simultaneous Localization and Mapping (SLAM) to compute the camera pose, and eventually the target position. SlidAR first casts a ray through the camera center to define a line going through the target position. Then, the user can move a cursor along that line (from any available viewpoint) until it reaches the desired position. This is an example of what we refer to as *Geometric* techniques. The geometric technique we evaluated is similar, although the second step was replaced by a second raycast, which is faster but does not allow adjustment. SlidAR was evaluated with targets within arm’s reach, while our evaluation focused on medium and far distances. The gesture annotation system presented by Nuernberger et al. (2016) is also conceptually similar to our geometric techniques, by utilizing interactive disambiguation in 3D AR. However, their work focused on image-based scene reconstruction and therefore relied on a pre-defined scene geometry model.

3.4 **Experiment I: Perceptual vs. Geometric Marking**

We designed an experiment to compare two marking techniques: one that relies on perception to mark from a stationary point of view (*Perceptual*) and one that uses two points of view, obtained by walking along a baseline (*Geometric*). We assessed marking accuracy outdoors, using an optical see-through display at distances between 12 and 85 meters. Given the lack of definitive agreement on the effect of AR displays on distance perception, our hypothesis was that participants using the perceptual technique would
follow a similar behavior as observed in real life, which is to underestimate targets at long distances (Norman et al. 2016). Since the Geometric technique does not rely on depth perception, we did not expect it to show any such bias. We also expected that Perceptual would yield good estimates at near/medium distances without incurring the cost of the two steps necessary for the Geometric technique.

3.5 Techniques

The Geometric technique calculates the intersection of two rays: each one cast from a different position and towards the target. This requires knowledge of the 3D baseline distance $b$ between the two points and the angles $\alpha$ from each point (Figure 3-1 B). For a target $T$ at a fixed distance, the accuracy is determined by the length of the perpendicular baseline between the two points and the user’s ability to cast the rays accurately in the direction of the target. Since two rays in 3D space will not necessarily intersect, we consider the marked point to be the point on the first ray which is closest to the second ray, or in our interaction methods that visualize cast rays and thus allow for correction of first ray inaccuracies, least-squares line intersection midpoints.

In the Perceptual technique, only one ray from the user location is needed. The ray defines a 2D position on the fronto-parallel plane from which a cursor can be moved in depth (Figure 3-1 A). In both techniques, the first step requires the user to trace a ray from its current position to the target.

![Figure 3-1: Perceptual and Geometric Techniques. A) Perceptual technique: the target location is determined by the position where the user places the cursor. B) Geometric technique: the target location is determined by the intersection of two rays (view from top).](image)

---

3 Smaller distances are better judged due to size constancy expectations and other cues [5].

4 A plane parallel to the one that divides the body into front and back portions.
The *Perceptual* technique is conceptually easier to understand and does not require users to move. However, it demands the use of a slightly more sophisticated input control, so that users can move the cursor with appropriate speed and in two directions. The *Geometric* technique calls for physical movement and requires users to understand that longer baselines generally result in more accurate point marking. On the other hand, it just requires pressing a button twice to cast the rays.

### 3.6 Apparatus

We used a Microsoft HoloLens optical see-through head-mounted display (HMD) and a wireless Xbox Controller for input. We used a virtual crosshair to help in aiming, rendered as a 3D object positioned 4 meters in front of the user. This distance made convergence easy and caused little eye parallax. We assigned a button on the Xbox controller to cast a ray from the position of the dominant eye (which could be specified in the application settings) through the crosshair. In the Perceptual technique, after the first ray was defined, participants could use the left thumbstick of the controller to move the cursor forward or backwards along the ray (Figure 3-2 A). The cursor was a white cube with blue edges, since white offers the highest amount of contrast in the HoloLens display. In both techniques, pressing the first button for a second time confirmed the marked location.

In the Geometric technique, we used the position and orientation provided by the HoloLens’ built-in tracking algorithms to determine the distance between the baseline points and the vectors to the target. To improve tracking, we provided more real-world features by placing a physical tripod next to one of the baseline points. The target distance was measured from an origin point placed next to the tripod. To reduce the chance of drifting we fixed the origin point using a HoloLens “spatial anchor.” The spatial anchor is adjusted automatically by the device so that its local relative position to the placement point remains fixed over time.

To minimize the effect of the sun, we scheduled the experimental sessions in the times of the day when the participants could stay under the shade of a large tree. If the sunlight around the target region was too bright, we added shades in the front of the HoloLens to
maintain good contrast between the target and AR graphics. It was made from several layers of translucent plastic attached to the headset using Velcro (Figure 3-2 B).

Figure 3-2: Experimental apparatus. A) Xbox controller and mappings used in Experiment I. B) Microsoft HoloLens with shades attached

3.7 Task and Environment

The task consisted of marking six targets. Each participant marked each target eight times, resulting in 48 data points per participant. The target sequence was initially randomized and kept the same for all participants and conditions. As participants completed one marking, they could see the number of the next target in the lower part of the display. The final position of the marker was not displayed, so that users could not learn from previous trials. We also counterbalanced the presentation order of the techniques.

The experiment was conducted in a wide and flat outdoor area on a university campus (Figure 3-3). There were some trees and buildings in the background. We used six lampposts as targets, at distances 12.5m, 26.7m, 40.2m, 55.2, 70.3m, and 85.5m. We asked participants to aim at the tip of each lamppost, which was at 3.84m from the ground. To indicate the baseline extent to be used in the Geometric technique, we picked two points clearly marked on the sidewalk, approximately 1.9m apart. In this way, participants could align themselves easily and precisely.
3.8 PARTICIPANTS AND PROCEDURE

We recruited seven graduate students. One participant was excluded due to technical issues. From the remaining participants, four were females and three males. Ages ranged from 23 to 25 years old and half of them had no prior experience with virtual or augmented reality. All participants were screened for stereo blindness and had corrected or normal vision. The experiment was approved by the (removed for blind review) Institutional Review Board.

Participants were asked to complete a background questionnaire with demographic information. Next, we measured each participant’s interpupillary distance using a pupilometer (Sunwin digital meter) and adjusted the HoloLens to the same value using its portal configuration page. Participants were introduced to the device, task, and techniques indoors and could train with indoor targets until they were confident performing the task. The participants were then taken to the outdoor site, where they completed the 48 trials. Finally, they were asked to answer a brief post-questionnaire about the techniques. The experimental material for this experiment can be found in Appendix C.
3.9 Results

Since in both these techniques the marking point lies on the first ray, the variance on the fronto-parallel plane should be the same. Thus, any systematic differences are due to different marking distances and techniques. Figure 3-4 shows an overview of the results obtained by participants using the Geometric and Perceptual techniques. The horizontal axis represents the true distance and the vertical axis the distances estimated by the two techniques. A perfect result is indicated by the dashed diagonal line. The Perceptual technique showed higher precision, showing only moderate variance change after 25 meters, but the accuracy was low. None of the participants marked a distance beyond 50 meters. The Geometric technique, on the other hand, was more accurate, but the variance increased steadily with distance. We fitted a multiple regression model to compare the variability in both techniques (adjusted r2 = 0.92, standard error = 1.2). Figure 3-5 illustrates how variability increased significantly faster with the Geometric technique (p = 0.045, coefficient difference = 0.032) although it started smaller (p = 0.043, coefficient difference = -1.77).

![Figure 3-4: Experiment I results. Bars show the standard error at each target distance.](image-url)
Interviews with the participants indicated that the majority preferred the *Geometric* technique (85%). Most of them reported that it was hard to judge the distance of the cursor while using the *Perceptual* technique. One participant expressed that not only it was hard to judge the size, but he could not even notice that the cursor was moving at all. At near targets, some tried to overshoot the target with the cursor and then bring it back to have a better sense of the real distance. When asked about issues with the *Geometric* technique, one participant also mentioned the difficulty of locking the crosshair on the target due to head tremor. Participants also indicated that some trees in the background made it difficult to see the crosshair and suggested changes in the crosshair or the experiment location.

![Variability Increase For Each Technique](image)

*Figure 3-5: Increase in variability with distance in Experiment I. Perceptual technique: green, Geometric technique: red. Lines correspond to a multiple regression fit of the points.*
3.10 Discussion

Interpretation of Results and Technique Improvements

Although the Perceptual technique had good accuracy at the first target (12.5 meters) it degraded rapidly, showing a clear pattern of increased underestimation. This result is consistent with distance compression findings in real-world and virtual environments.

However, our results differ from that of Swan et al. (2006), which found a switch to overestimation in augmented reality after 23 meters. Although both experiments used a matching task as measurement, the cursors used and the environment were quite different. While they conducted the experiment in a long corridor, which provided strong perspective cues, we used a large open field. Their cursor plane also filled a large part of the visual field, which was not true for our cube. Their AR setup was fixed on the ground, while in our experiment, it was head-worn. In addition, our physical targets displayed distinct features that might have provided better cues of relative size and height in the visual field. Also, they were not aligned with the target direction, and were viewed from an oblique viewpoint. Finally, Swan et al. modulated the opacity of the cursor to provide an atmospheric perspective cue, and we did not.

We could potentially improve the performance of the Perceptual technique by using a larger cursor, replacing the cursor with a model that has a known real-world size (e.g., a person), offering training, or even compensating for underestimation based on empirical measurements. However, the results from this experiment and prior work indicate that techniques based on perception will likely stay rather inaccurate and vary from person to person (Kuhl et al. 2006).

Since the Geometric technique does not rely on depth perception, the variability observed in Figure 6 must have other origins. As distances grow larger, small angular errors in head orientation due to system noise and head tremors can introduce substantial placement errors. Although it is easy to fixate the eyes on a distant target due to the vestibulo-ocular reflex (Lorente de Nó 1933), 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 the target was quite obvious. The accuracy of the ray-casting
depended on the stability of the participant’s head and the time at which they pressed the button to cast the ray.

Thus, the variability issue in the Geometric technique can be improved by changing the input modality to provide more precise aiming, stabilizing the crosshair by filtering, and/or gathering more samples. Although speed was not our primary design goal, we felt that using complex pointing procedures or trading jitter for latency would detract from the simplicity of the original technique. Continuing the improvement to the above techniques, we decided to redesign the Geometric technique to enhance precision by taking multiple samples.

**HOLOLens accuracy**

We set about more thoroughly understanding the source of the underestimation observed with the Geometric technique in Figure 3-4. The final marking position depends on the measurement accuracy of both angles and the distance between the origins of the two rays (Figure 3-1 B). Both values were obtained directly from the HoloLens, based on its proprietary tracking algorithms. As with any other device, it can present inaccuracies which may be more or less acceptable depending on the intended use. Prior work has addressed the position accuracy of the HoloLens (Vassallo et al. 2017), but we decided to perform some additional analysis on angular accuracy, since it held particular importance for our Geometric technique.

To produce the underestimation observed in experiment I, the average angular error would need to be on the order of 0.25 degrees. To see if HoloLens tracking might be the source of such errors, we attached an HTC Vive controller rigidly to a HoloLens. Both devices were then mounted on a spinning bar stool, rotated 360 degrees over about 7 seconds, and stopped. The relative difference between the two forward vectors was recorded before and after the rotation. This procedure was repeated four times. According to Niehorster et al. (2017) the inertial-optical Lighthouse system can track the controller with an angular error of 0.02 degrees as long as tracking is maintained. The results of our tests showed an average difference of 1.2 degrees between the Vive and HoloLens data (Table 3-1). Although this was not a formal evaluation of the HoloLens accuracy, and we had shielded against larger tracking errors via a much more controlled and careful
tracking setup than in this bar stool experiment, it was sufficient to cause us to suspect that HoloLens tracking error (especially outdoors) may be significant enough to interfere with the evaluation of our techniques. Our hypothesis is that insufficient environment visual features (as in our outdoors experiment) or rapid motion (our spinning test) make the HoloLens rely more on IMU for localization (visual tracking probably runs a lower frequency than the IMU). IMU based tracking is known to suffer from both position and cumulative drifts. Even when carefully corrected, cumulative angular drift can still be around 2.5 degrees (LaValle et al. 2014), a value that can create large errors in long-distance triangulation.

Table 3-1: Angular drift between Vive and HoloLens. Difference in the reference frames in the spinning test. All values are in degrees.

<table>
<thead>
<tr>
<th></th>
<th>Max</th>
<th>Min</th>
<th>Mean</th>
</tr>
</thead>
<tbody>
<tr>
<td>x-axis</td>
<td>-0.144</td>
<td>-0.198</td>
<td>-0.17</td>
</tr>
<tr>
<td>y-axis</td>
<td>0.848</td>
<td>0.780</td>
<td>0.827</td>
</tr>
<tr>
<td>z-axis</td>
<td>0.841</td>
<td>0.773</td>
<td>0.820</td>
</tr>
<tr>
<td>Angle difference</td>
<td>1.21</td>
<td>1.11</td>
<td>1.18</td>
</tr>
</tbody>
</table>

3.11 EXPERIMENT II: IMPROVING PRECISION WITH MULTI-SAMPLE MARKING

In this experiment, we compared the original Geometric technique with VectorCloud, a geometric multi-sampling technique. To avoid errors or variation due to the HoloLens tracking system, we simulated the AR setting of the first experiment in a VR system. With more precise tracking, we could reduce the variance introduced by hardware limitations and evaluate the techniques under more controlled circumstances. Our hypothesis was that in the controlled VR setting VectorCloud would have the same accuracy but be more precise, as compared to the Geometric technique. We also expected VectorCloud to improve user experience by relaxing the requirement for pointing precision.

THE VECTORCLOUD TECHNIQUE

Assuming that the errors in the Geometric technique are 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. The insight behind the technique is the assumption 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 head orientation. Next, we compute all possible intersections between the stored rays (Figure 3-6), in a similar manner as in the original Geometric technique. However, we use the midpoint between the closest points on the two rays as the intersection point.

Since all intersections are tested, this technique allows users to take samples from any number of arbitrary positions in space. Users can also sample continuously as they walk between any two points. When sampling from two discrete locations, ray intersections will always generate useful information, since each intersection will be valid. However, special care is needed in the case of continuous sampling and discrete sampling from more than 2 locations. For the evaluation, we implemented a simple solution that works in both cases: we first 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. More complex solutions, like clustering the rays, will likely be faster and even more accurate. We leave this for future work.

![Figure 3-6: The VectorCloud technique. Users aim at the target from different locations while the rays are recorded. All pairs of rays are then intersected to obtain several sample positions.](image)

Once the intersections are computed, we can use any location measure to determine the estimated position the user was trying to mark. Due to the geometry, the distribution of intersections from two points has a positive skew, with most values concentrated closer to
the user. For this reason, computing the mean will result in a slight overestimation of the target position (Figure 3-7).

We have explored other methods to estimate the location of the target, including using the median or mode, or performing asymmetric trimming before computing the mean. All of these, however, only provide the exact position if the axis used for ordering is exactly aligned with the extremes of the distribution. In our initial tests, we found that noise and smaller baseline differences widely changed the 3D shape of the distribution, preventing a robust determination of the correct axis. In addition, the shape of the distribution is ill-defined when more than two points are used (or during continuous sampling). In the end, we decided to use the mean because we expected the overestimation introduced by this to be small for our purposes.

![Figure 3-7: Top view of the distribution of intersections generated by two viewpoints A and B. The red dot is the actual target, aimed with same angular error form each viewpoint. Due to the long tail, the mean (green dot) falls beyond the target.](image)

From the user’s perspective, VectorCloud follows the same steps as the original Geometric technique, except that user can hold down the controller button to generate more ray samples. Since we do not impose a limit on how many samples the user can gather, we used a second button to finalize the overall marking process.
3.12 EXPERIMENT DESIGN

As explained at the beginning of this section, we conducted experiment II in VR to reduce confounds caused by the AR system. This approach is known as Mixed Reality Simulation [2]. We rendered the crosshair and other AR graphics with a customized shader to prevent occlusion by other geometry and with a transparency value close to a typical experiment day with the shades used in experiment I. We also used a similar environment and targets. Six virtual lampposts were placed at the same distances as their real-world counterparts: 12.5m, 26.7m, 40.2m, 55.2m, 70.3m, and 85.5m. We also picked two locations in the virtual scene corresponding to the user positions in experiment I so that the baseline is the same (1.9m). The VR scene is shown in Figure 3-8 (compare the AR scene in Figure 3-3). 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 point that should be marked (Figure 3-8).

![Figure 3-8: The VR experimental setup emulating the first experiment. The targets and interfaces are visible.](image)

Each participant marked each target eight times, resulting in 48 data points per participant. The target sequence was the same as in the previous experiment. As participants completed one marking, they could see the number of the next target in the lower part of the display. The final position of the marker was not displayed, and we counterbalanced the presentation order of the techniques. To help with aiming, we asked
participants to only use their dominant eye while casting the ray(s). We collected 300 samples from each position.

3.13 Apparatus

We used a consumer version HTC Vive head-mounted display (HMD), and a wireless Xbox controller for input (Figure 3-2 A). We used the right shoulder button to cast rays in both techniques. In the VectorCloud technique, the left shoulder button was used to complete the trial. The Vive has two screens, each with a resolution of 1080x1200 pixels. The total horizontal field of view is 110 degrees. It was tracked with six degrees-of-freedom by the hybrid inertial-optical Lighthouse system. The software used in the experiment was written in Unity3D.

3.14 Participants and Procedure

We recruited six graduate students and four undergraduate students (three female and seven male). Ages ranged from 21 to 39 years old, with the mean being 25.33. Most participants had used VR or AR at least once or twice before, while one participant had no prior experience with the technology. 90% of the participants were right-eye dominant and one participant was left-eye dominant. The experiment was approved by the (removed for blind review) Institutional Review Board.

We followed a similar procedure as in experiment I, including stereo blindness screening and inter-pupillary distance measurement. Because we wanted the participant to aim the target with only the dominant eye, a Porta test [20] was added to the measurement process to identify the user’s dominant eye. Based on the result, we blocked the display in the HMD for the non-dominant eye with a black cloth to minimize fatigue. Participants were trained until they had sufficient confidence for the tasks, including on the use of the Xbox controller. To ensure accurate tracking, we paid special attention during the experiment that the system never lost tracking of the HMD. After participants completed all trials, a post questionnaire was presented to gather qualitative feedback about the techniques. The experimental material for this experiment can be found in Appendix C.
3.15 Results

Figure 3-9 shows the overall results, comparing the original Geometric technique and the VectorCloud technique. The horizontal axis represents the true distance and the vertical axis the distances estimated by the two techniques. A perfect result is indicated by the dashed diagonal line. Both techniques were highly accurate, although the Geometric technique shows a slight overestimation for the last target. No overestimation is seen for VectorCloud and we also see a strongly reduced overall variability.

This improvement is more evident in Figure 3-10. The points correspond to the standard deviation for each target and the line is a fitted polynomial multiple regression (adjusted $r^2 = 0.97$, standard error = 0.82). The variability of both techniques appears to grow with the square of the distance. The difference between the regressed coefficients for the first-degree term is 7.13 ($p < 0.0001$), and for the second-degree term 3.1 ($p < 0.001$). The quadratic model had a significantly better fit than a line ($p < 0.002$).

Figure 3-9: Experiment II results. Bars show the standard error at each target distance.
The standard error of VectorCloud for the last target is more than three times smaller than the Geometric technique, which represents a considerable gain in precision. We also inspected the variance on the fronto-parallel plane and, as expected, this variance was also smaller with VectorCloud. Figure 3-11 shows a comparison between the view plane variability of the Geometric and VectorCloud techniques accumulated for all targets.
3.16 Discussion

The multi-sample approach of VectorCloud improved the marking precision by computing a better estimate of the target position. For both techniques, the standard error increases with the square of the target distance. Given a constant angular error $\epsilon$, the amount of linear error in a plane parallel to the view plane at the target distance is $\tan(\epsilon)d$, leading to an increase in the variance by the square of this factor. Similarly, while developing a model for distal pointing, Kopper et al. [8] found that movement time increased with the square of the task difficulty index (which was modeled as an inverse function of angular size).

Although the magnitude of the pointing errors quickly increase with distance, the reduction in variability due to the multi-sampling of the VectorCloud technique made the actual error much smaller than what would be predicted by a model based on information about human head orientation stability. Skavenski et al. [21] report that head rotations, even while participants try to be as still as possible, can reach angles of 0.75 degrees. At an 85-meter distance, this would correspond to underestimation errors of up to 40 meters (far below the VectorCloud standard error of 5 meters and closer to the Geometric standard error of 20 meters).

Using the mean as a measure of location of the point distribution was satisfactory. For both techniques, the difference between the mean and the veridical distance was well within the variability created by head and system jitter (approximately 22% of standard error).

While the improvement from Geometric to VectorCloud is promising, we still observed several limitations of the technique itself. As it does not provide any feedback on the sampled rays, the user has no way to improve accuracy because she does not know how many samples she needs or how precise she has to be. The user also does not have a fine-grained control of the final ray direction as the computation is not presented to her.

3.17 Qualitative User Study

In addition to the quantitative experiments, we also wanted to get qualitative feedback from users about the marker placement techniques. A third study was conducted as a
proof-of-concept for the AR application of the geometry-based techniques. In the study, we asked participants to perform a series of marking tasks outdoors at targets 5m, 25m and 50m away using the original Geometric technique, VectorCloud from two discrete locations, and VectorCloud while walking. After the tasks, an interview was carried out to gather detailed feedback based on the experience. Our goal was to identify user experience issues with the techniques, evaluate continuous walking, and understand how users might use the techniques in practice. To increase ecological validity, we chose to run the study using the HoloLens in an outdoor environment. Unlike experiments I and II, persistent visual markers were placed in the environment after each task so that users could subjectively evaluate the accuracy of the marking techniques.

Participants and Procedure
We recruited 2 graduate students and 1 undergraduate student (all females, mean age 25.33) from the university. All participants had used VR or AR at least once or twice. All participants were screened for stereo blindness and had corrected or normal vision. The experiment was approved by the Virginia Tech Institutional Review Board.

Participants were welcomed by the experimenter and asked to sign the consent form. Next, they were asked to complete a background questionnaire with demographic information. Then the HoloLens was calibrated to fit each participant’s interpupillary distance. They were then introduced to the device, task, and techniques in the lab with indoor targets until they were confident performing the task. The participants then went outdoors to mark three designated targets of 5m, 25m and 50m away with all three techniques. Then they were encouraged to place other markers at their discretion. They were asked to think aloud throughout the study. Finally, they were taken back to the lab for an interview.

Results and Discussion
Overall, all of the users were happy with the techniques and agreed on the potential applications with comments such as “They’re supposed to be useful in daily life.” Users were excited to express how they could apply the AR marker placement techniques to use at home/work. They unanimously asserted when such applications were available for them, they would be “very willing to” try and use them.
VectorCloud from two discrete locations was the most preferred technique. Two users claimed it to be “more accurate and reliable.” They appeared to become more confident knowing that more samples would be considered in the calculation. They also identified several expected benefits, including: offloading aiming tension, recovering from previous mistakes, progressive refinement aiding confidence in aiming process, and reducing the influence of head tremor.

Only one user expressed a preference for VectorCloud while continuously walking, mainly because physical locomotion was “more useful in real life.” The other participants agreed that continuous walking could make use of more locations along the walking path; however, they were unable to state how this could benefit the task of marker placement. Moreover, the other two participants claimed walking would add extra variance in the aiming process due to physical movement. One further described continuous walking as “distracting” because of the extra attention dedicated to walking. According to this user, continuous marking was not natural or comfortable. We were not surprised to see that users moved pretty slowly when using continuous walking as they were still attempting to produce accurate samples.

3.18 CONCLUSIONS

Point marking in 3D is fundamental yet challenging task in content creation in AR, especially when information about environment geometry is unavailable. In this research, we have presented and evaluated point marking techniques under this constraint from two important classes: perceptual and geometric. Our work confirmed the effect of significant perceptual distance compression, leading to geometric techniques being superior in terms of accuracy. In addition, we showed that involuntary head tremor can reduce the precision of naïve geometric techniques. Our VectorCloud technique addresses the issue of head tremor by exploiting multiple samples to reduce variability and also allows continuous marking while walking.
4 MANAGING WORKSPACES IN AUGMENTED REALITY

In this chapter, we describe our research on interfaces that can adapt to walking conditions and to changes in the spatial configuration of the environment caused by movement from one point to another. We also discuss aspects related to walking as an AR dual-task.

4.1 INTRODUCTION

Wearable augmented reality (AR) devices give us the valuable ability to visualize digital information anywhere. However, the way we manage information in AR still limits its full potential. Imagine you are working on a project using a virtual AR workspace. You place a diagram on the wall, notes and images on the table, and your e-mail client on your right side. Eventually, you leave your office to get a coffee. Suddenly, someone asks you a question about your project. What do you do? The information you need is on your office wall. You could, of course, go back and pick it up. Or you could open it again from a system menu. However, this would quickly become tedious and inefficient to do every time you walk during the day. It would be nice to be able to call the information to your current location and have it arrange itself on the surfaces around you. In other contexts, you might prefer to keep the relevant information always within your sight—for instance, if you are on a field trip or reviewing your appointments while you walk to work.

If AR is to be truly mobile, interfaces should reflect the way we seamlessly move around in the world. This entails not only being able to access information in different places but also being able to make use of it during short or long walks. Unfortunately, the current approaches used for AR information layout are not sufficient to achieve this goal. As the scenario above illustrates, a fixed layout in the world does not allow mobility, and manually moving windows is not practical every time you move. In addition, unlike applications on your mobile phone, the physical environment does matter when you walk
in AR. The new room might not have not a table on which to spread your documents anymore, or it might be smaller. Other common solutions, such as attaching the information to your body or head, also have restrictions. They can, for instance, occlude the real environment and limit your awareness of the surroundings. In addition, this solution negates the benefits of a closer integration between AR and the world.

What if your workspace could follow you around, and when desired, quickly adapt to your new location and task? The main goal of our research was to investigate what such a system might be like. Instead of designing static interfaces (such as a head-up display or world-fixed information displays), our research considered dynamic interfaces that adapt to the user’s movement and to the physical environment. These adaptive behaviors could be triggered by changes in the user’s position and orientation, and also by changes in the physical space. By sensing when the user walks, interfaces could move to a more suitable position for reading, change their layout, or switch the primary interaction technique. By sensing the environment, an adaptive interface could maintain its consistency with the world and take advantage of surfaces such as tables and walls.

We designed a variety of behaviors to provide adaptive window management in mobile AR, implemented those behaviors in a modular system that can combine those individual behaviors, and proposed a final minimal set of useful behaviors that can be easily controlled by the user in a variety of mobile and stationary tasks. We also used this system to capture user experience during contextual studies and improve our general understanding of the design requirements for mobile AR workspaces.

4.2 RELATED WORK

Here we review some of the prior work on AR information displays, 3D workspaces, and walking performance under dual-tasking.

AR INFORMATION DISPLAYS

Feiner et al. (1993) describe an X11 window system for AR that supports three configurations for windows: 1) fixed relative to the HMD; 2) fixed to a sphere the surrounds the user; or 3) fixed to locations and objects. In the HMD-relative
configuration, the windows behave as in a Heads-up display, so that they are always visible, regardless the user head pose. Windows in the spherical surround configuration are projected to a fixed position in a virtual sphere which encloses the user. The sphere moves with the user, dragging and rotating the windows along. Finally, when windows are fixed to locations or objects, they stay in the same position as the object. Since the windows were composited from the X11 system, they did not have depth and interaction between the windows was restricted. One of our behaviors is similar to an initial description of the user-surface layout. Unlike the authors, though, we evaluate our translation mode independently of rotation.

Another early desktop manager was ARWin. Although still based on the X11 manager, ARWin could render the content of each window into a polygon mesh, creating fully 3D windows. The position of each mesh was determined in the world coordinates using markers (Di Verdi et al. 2003). Like ARWin, our system makes use of 3D windows. However, our windows are fully dynamic.

No matter which configuration is used, one can potentially have a large amount of information displayed in a limited field of view. Bell et al. (2001) introduced algorithms for view management, with had the goal of automating the layout of annotations in the viewing plane as the user moves. This allows the system to optimize the layout for given set of constrains, such as maintaining the visibility of real objects, keeping a minimum size or keeping the proximity to their original positions. All the behaviors we explore in this work considered the initial position of the elements. Our goals, though, were different. Our intent was to respect the user preferences for the layout as much as possible. This initial arrangement is 3D and the elements may move inside or outside the view according to this configuration.

Based on these and prior work Müller & Dauenhauer (2016) propose a taxonomy for information annotation in AR. However, for the coordinate dimension, they adopt only two coordinates: a world coordinate system (WCS) and a spectator coordinate system (SCS). All specific references must be described by any of the four different combinations for position and orientation. In this taxonomy, an element with position in the WCS, and rotation in the SCS is a viewpoint oriented billboard. The taxonomy of
Müller & Dauenhauer (2016) is similar to ours, in the sense of considering the separate effect of translations and rotations. However, since they use a single reference, they cannot differentiate between the same pose in different coordinate frames.

Finally, Wither et al. (2009) describe how AR annotations are related to the environment with the concepts of “location complexity” and “location movement”. The location complexity reflects the complexity of the registration to the world. An example of low complexity would be a point anchor, while an annotation with high location complexity could consist of areas or volumes. The second concept, location movement’, measures the freedom the annotation has to move away from the registration point. That include only the distance, not movements within the annotation itself. Considering movements within a coordinate frame can increase the descriptive power of layout annotations. However, only measuring the freedom distance in 3D space does not tell what the movement does. Our adaptation taxonomy tries to improve on this idea by describing how the element moves (by employing user movement, world surfaces, and normal as references).

3D WORKSPACES

Depending on how the information is rendered, users may explore of spatial properties to facilitate information organization and retrieval. The Data Mountain is an early document management technique that takes advantage of this fact. It allows users to place documents over a 3D inclined plane (the mountain) and drag them to create groups and hierarchies (Robertson et al. 1998). Because of the perspective, documents on the top of mountain seem smaller, while documents on the base look comparatively larger. This, along with other cues, such as sound, let users infer the 3D position of each document. In a study, participants were able to accurately and efficiently retrieve stored documents.

The Task Gallery extended this idea to a 3D window manager designed to take advantage of the human spatial perception and memory to enable multitasking in familiar space, an art gallery (Robertson et al. 2000). To this end, the system provided the ability to combine several operations (like opening and placing windows) in a single step. In addition, it introduced spatial navigation, in an attempt to capitalize even more spatial abilities. The Task Gallery operation is conceptually very close to the way we view the
adaption to the environment. In this study we are focused on adaptation to walking, but our work could potentially be used with the same goal. Interestingly, the authors discovered that even in a non-immersive environment, participants did not like to place the documents on the floor and ceiling. During our experiment, no one tried to place windows on horizontal surfaces either.

Information-rich virtual environments (IRVEs) explore the use of 3D space to integrate complex, heterogeneous data into a single integrated information space. Spatial information is visually associated with abstract and symbolic information to allow users to quickly form mental models of complex data (Polys et al. 2011). In this sense, the abstract information (e.g., text) can be seen as an annotation for more concrete entity (e.g., an object). For this reason, the display location of the information becomes relevant and is described as being world-fixed, display-fixed, object-fixed, or user-fixed. A new distinction is made between world-fixed and object-fixed: the first is stationary, while the second moves with the object (Bowman et al. 2003).

Billinghurst et al. (1998) evaluated the performance of users in locating and remembering AR information in three configurations: head stabilized, stabilized around the participant body with trackball interaction, and stabilized around the body with head tracking. In this study, the two last conditions were equivalent to a world coordinate system, since the participants did not walk. The authors found that participants were faster in the body stabilized configurations. The head tracked condition, in particular, enabled higher recall. Participants also reported the head tracking condition to be more natural and intuitive. During our preliminary design sessions, we found that the body-based configuration is, indeed, more convenient. However, during the user studies participants showed a clear preference of not moving their heads too much.

More recently, immersive 3D workspaces have been suggested as a way to improve the understanding of complex non-spatial data. Bacim et al. (2013) evaluated the impact of several factors on the performance of the inspection of 3D undirected graphs for tasks like detecting intersections, path following and connection identification. The results indicated that the use of stereoscopic, immersive mode with head tracking, improved the performance on graph analysis and spatial understanding tasks.
WALKING, PERCEPTION, AND SPATIAL COGNITION

The benefits offered by 3D workspaces are only possible due to our ability to perceive and understand the space, which we then skillfully explore to create meaningful relationships. However, as we try to do more (in our case walking), we eventually reach the limits of our natural abilities.

Walking is a complex task that requires one to be cognizant of the destination, the surrounding environment, and to be able to coordinate the limbs to successfully reach the destination. The interference of cognitive tasks on walking has been well documented. It has been observed in many different dual tasks, such as talking, doing arithmetic, counting, verbal fluency tasks, etc. (Yoge-Seligmann et al. 2008a). In a meta study, Al-Yahya et al. (2011) concluded that these dual-tasks cause changes in walking speed, cadence and stride characteristics.

In general, the mechanism underlying cognitive-motor interference is still not clear. (Leone et al. 2017) points to three possible attentional models proposing: 1) that a capacity-limited process requires resources to be redistributed between the tasks, 2) that critical processes must be carried out sequentially, creating a bottleneck, or 3) that if two tasks share common neuronal populations, the conflict will result in cross-talk. The executive function, thus, plays a relevant role during dual-task conditions. Factors like poor self-awareness, lack of response inhibition, and attentional deficits, can lead to gait disorders and even falls (Yoge-Seligmann et al. 2008a; Hall et al. 2011; Plummer-D’Amato et al. 2012).

Sedighi et al. (2018) compared the effect of using smart glasses, smartphone, and a paper notebook on gait variability. Participants were asked to perform three different cognitive tasks while sitting in a chair or walking in a treadmill. The authors found that the risk of fall was higher during the dual-task, but participants used more adaptable gait strategies with the head-up display, which might help decrease the risk of falling.
Adaptation is a strategy to improve user interaction by optimizing the interface to the way it is being used. Park & Han (2011) describe four categories of adaptation: Adaptable interfaces allow the user to manually customize the interface so that items of interest are more visible and accessible. In adaptive interfaces, however, this role is delegated to the system, which automatically modifies the interface according to a policy (e.g., prioritize last used item or most frequently used one). Both options have advantages and disadvantages. An adaptable interface, for example, is less confusing and easier to remember, but creates an additional effort to manually modify the interface (Findlater & McGrenere 2004; Park et al. 2007; Lavie & Meyer 2010; Mitchell & Shneiderman 1989). The authors also describe mixed approaches, in which the responsibility for determining and carrying out the adaptation is shared between the user and the system. In adaptable interfaces with user support, the system recognizes when adaptation is required, but the user is responsible for determining and carrying out the adaptation. Alternatively, in adaptive interfaces with user control, the system recognizes and carry out the adaptation, but the role of determining the adaptation is shared by both the user and the system. Our approach constitutes a third mixed approach, in which the adaptation is recognized and determined by the user, however it is carried out by the system. In this work we refer to it generally as an adaptive interface.

Bouzit et al. (2016) propose a design space for adaptive menus, the most studied area in interface adaptation. By looking into different visual aspects of user interfaces, they discuss how each visual attribute could be used to construct adaptive 2D interfaces. Position-changing menus, for instance, groups adaptation manifested by changes in position. Other attributes, such as shape and color, could be used to create shape-changing and color-changing menus respectively. In addition, the authors define four stability properties, which are attributes that do not change during adaptation: Spatial stability refers the ability of the adaptive menu to keep its spatial layout constant (position and orientation); physical stability refers to the constancy of size and shape; format stability is the ability to preserve the format of the layout (color, texture); and temporal stability is defined as the capacity of preserving the layout over time (i.e., keeping motion constant). Although not discussed in detail, the authors imply that there is
a stability/adaptation tradeoff: more stability allows the system to be more predictable for the user, at the cost of the ability to adapt. In our work, we try to preserve relative spatial stability: the interface elements change position, but keep their relative position as constant as possible. All other properties are preserved: there are no time discontinuities and no changes in size or format.

4.3 Research Approach

To investigate the use of adaptive AR workspaces for walking, we conducted exploratory research to obtain design insights and generate hypotheses for further exploration. We began by expanding our initial conceptual design into a detailed design space, followed by the implementation of configurable proof-of-concept system. We then conducted a qualitative contextual study, in which participants used our system to explore and generate new insights about adaptation. Based on what we learned, we redesigned the system and recruited another group of participants to a second study (Figure 4-1). Finally, we identify and discuss specific properties that can be used as design guidelines for future systems.
The qualitative studies were semi-structured (Soegaard & Dam 2012), and we applied thematic analysis to identify, analyze, and report the patterns from the data (Braun & Clarke 2006). The data consisted of system logs, interview questions, and observations. The interview questions were adapted or modified when necessary to keep a productive dialogue with the researcher. During the analysis, we adopted an inductive approach in which themes were identified from the data without the use of a-priori categories. After repeated patterns were aggregated into categories, those were interpreted in the light of known principles and theories. Some codes, though, reflected the structure of the study. For example, the specific behaviors implemented in our system. So, our approach was both inductive and deductive, starting from the phenomenon, capturing the data and proceeding to the explanation (Corbin & Strauss 1998; Reichertz 2014).

During the contextual studies, our system was used as a technology probe (Hutchinson et al. 2003) that allowed us to: 1) identify the strengths and limitations of our adaptation approach to walking conditions, 2) investigate how participants would appropriate AR to their needs and interaction styles in real settings, and 3) engage participants in co-creative design to generate new ideas and directions for walking interfaces.

4.4 Defining the Design Space

Our initial vision was to use adaptation as a way to: 1) support the use of AR interfaces while walking; and 2) minimize the effort required to adjust the layout to a new environment. For example, a user watching a game in the living room may decide to have some food in the kitchen. Instead of staying in the living room, the window displaying the game could follow the user around the house and as he stops in the kitchen, enlarge and attach to the nearest wall (Figure 4-2 left). A layout with several windows could work in a similar way, with the additional constraint of considering both the initial layout and the new surrounding physical space to find a good match (Figure 4-2 right). In this concept, the entire adaptation could be defined by a set of behaviors, such as follow, or attract, that when combined would produce the desired outcome.
Figure 4-2: Windows adapting to nearby walls. Left – Window attaching to the nearest wall. Right – Adapting while respecting the workspace layout.

To address the design problem in a more systematic way, we proceeded to expand and detail the elements of our initial vision. Following the work of Zimmermann et al. (2005) on contextual adaptation, we built upon the framework of Brusilovsky et al. (1998) of adaptive hypermedia. We extend it to include information about the context (in our case, the physical space around the user) and the effect of the adaptation on the visual aspects. Brusilovsky et al. (1998) also propose a classification of adaptive systems based on four dimensions: adaptation goals, features to adapt, features used for adaptation, and adaptation method (Figure 4-3).

Figure 4-3: Adaptive System Model. It was extended from Brusilovsky et al. to include the environment and the reflexive term from the system. Green boxes indicate the corresponding dimensions of adaptation.
Although the goal of most adaptive systems is to improve visual-motor performance on the interface component-level (e.g., menus), our adaptation goal was to improve the experience within the context changes caused by walking. Our first choice was to decide which features to adapt. Although several visual aspects could be used, we focused on position-adaptation (Bouzit et al. 2016). Position has the highest generality across applications, since it is a feature common to all spatial content. Below, we discuss in more detail the features used for adaptation and our choice of adaptation method.

**Features used for adaptation**

As sources of adaptation, we considered information that could be easily obtained from sensors available in AR headsets. These include: 1) information about the user position; 2) information about the surrounding environment; and 3) information about the layout itself (Figure 4-4).

![Figure 4-4: Adaptation Taxonomy. We consider three possible triggers to adaptation behaviors: the user, the environment, and the layout itself.](image)
We decomposed each information source into properties and events that could trigger adaptation of the interface, creating a taxonomy to describe the adaptive behaviors. For position and orientation (visual aspects that we are interested in), the taxonomy can be used to describe many position-changing adaptive responses. Below we describe each dimension.

**User Information**
This category encompasses all the data that could be inferred from the user action directly (for example, gestures with the head or the body).

| Body Position | • Ignore user position change – Stay fixed in space even if the user walks away  
• Follow user position – Update element position to match current user position  
• Update orientation – Change the orientation to compensate for the change in position (rotate in the y-axis) |
|---------------|------------------------------------------------------------------------------------------------------------------|
| Body Orientation | • Ignore user turns – Do not translate or rotate to keep the same relative position regarding the user view  
• Rotate around user – Move to keep the absolute coordinates  
• Update orientation – Change the orientation to compensate for a change in yaw |
| Head Pitch | • Ignore up/down head movement - Keep the original height if the user looks up or down  
• Adjust height – Move up or down to follow user’s head direction  
• Change orientation – Rotate around x-axis to compensate for a change in pitch |
| Head Roll | • Ignore head tilt – Keep the original rotation if the user rolls the head  
• Adjust with head tilt – Rotate the y-axis to adjust with the head roll |
### Explicit Commands
- Device Input – Buttons, touchpads, and other device-based input
- Gesture – Discrete or continuous gestures

### Environment Information
This category covers the adaptation of the elements to the physical configuration of the space (for example, to align with a physical table or wall).

<table>
<thead>
<tr>
<th>Surface Position</th>
</tr>
</thead>
<tbody>
<tr>
<td>- Ignore – Allow elements to pass through physical obstacles (no collision)</td>
</tr>
<tr>
<td>- Collide – Prevent elements from going through physical obstacles</td>
</tr>
<tr>
<td>- Attract – Move towards walls</td>
</tr>
<tr>
<td>- Stick - Stay fixed on the wall position once close enough</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Surface Orientation</th>
</tr>
</thead>
<tbody>
<tr>
<td>- Do not align - Keep orientation and disregard physical surface normal</td>
</tr>
<tr>
<td>- Soft align – Align with the surface normal proportionally to the distance</td>
</tr>
<tr>
<td>- Hard align – Align with the physical surface only after a collision</td>
</tr>
</tbody>
</table>

### Layout Information
This category covers the changes initiated by the internal status of the system (for example, to prevent one element from occluding another). This can be seen as the view-management aspect of the interface.

<table>
<thead>
<tr>
<th>Workspace Arrangement</th>
</tr>
</thead>
<tbody>
<tr>
<td>- Relative position - Maintain relative position and distance from other elements</td>
</tr>
<tr>
<td>- Global position – Move the windows to conform to the available physical space</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Field of</th>
</tr>
</thead>
<tbody>
<tr>
<td>- Absolute size – Maintain the absolute size of interface elements</td>
</tr>
</tbody>
</table>
View

- Relative size – Scale the interface size (for example, to account for user’s distance)

The design space of combinations of *user information* adaptations subsume many layout categories discussed in the literature while offering a practical way to describe mixed interactions with the environment as well. Table 4-1 shows how our taxonomy can describe adaptations that behave like specific coordinate frames from previous taxonomies (for example, the spherical surround of Feiner et al. (1993) / user-fixed reference frame). In addition, by selecting other combinations from these and other dimensions, it is possible to generate behaviors that are not easily described by a simple reference frame. For example, [Body Position - Follow] and [Surface Position – Attraction] makes the windows follow the user while staying on the walls. It neither corresponds to a fixed position in the user coordinate frame, nor to a fixed position in the world coordinate frame (Figure 4-5).

![Figure 4-5: Windows following the user position while staying on the wall surface.](image_url)
Table 4-1: Equivalence of properties in our taxonomy. The table maps prior layout taxonomies to our adaptation taxonomy (using only the User Information dimension).

<table>
<thead>
<tr>
<th>Taxonomies from</th>
<th>Body Position</th>
<th>Body Orientation</th>
<th>Head Yaw</th>
<th>Head Pitch</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bowman et al. (2003) / Müller &amp; Dauenhauer (2016)</td>
<td>Follow</td>
<td>Adjust</td>
<td>Adjust</td>
<td>Adjust</td>
</tr>
<tr>
<td>Display-fixed / SCS positioned &amp; SCS oriented</td>
<td>Follow</td>
<td>Adjust</td>
<td>Adjust</td>
<td>Adjust</td>
</tr>
<tr>
<td>User-fixed / SCS positioned &amp; WCS oriented</td>
<td>Follow</td>
<td>Adjust</td>
<td>Ignore</td>
<td>Ignore</td>
</tr>
<tr>
<td>World-fixed / WCS positioned &amp; WCS oriented</td>
<td>Ignore</td>
<td>Ignore</td>
<td>Ignore</td>
<td>Ignore</td>
</tr>
<tr>
<td>WCS positioned &amp; SCS oriented (Do not follow but always face the user)</td>
<td>Ignore</td>
<td>Adjust</td>
<td>Adjust</td>
<td>Adjust</td>
</tr>
<tr>
<td>(Follow but maintain world orientation)</td>
<td>Adjust</td>
<td>Ignore</td>
<td>Ignore</td>
<td>Ignore</td>
</tr>
<tr>
<td>(Do not follow, face the user, but keep upright)</td>
<td>Ignore</td>
<td>Ignore</td>
<td>Adjust</td>
<td>Ignore</td>
</tr>
</tbody>
</table>

**Adaptation Method**

Our interface adaptation strategy requires that interface elements react both to the user’s actions and to the environment. Our implementation approach was inspired by behavioral robotics: each element moves around in space with minimal information about the world and sensing only what is necessary to perform the next move. This approach allows each interface element to be independent and does not rely on global planning. This is a useful feature for real-time implementation in a real device, as it allows elements to behave with incomplete knowledge of the environment.

In this way, interface elements follow a gradient descent over a potential field created by surfaces and other elements in space (Andrews & Hogan 1983; Khatib 1986). This approach allowed us to quickly prototype and test several behaviors. By appropriately changing the relative weight of the attractors that compose the field, elements can move towards the user, along walls, or move smoothly around obstacles. To implement more complex behaviors, the weights can be adjusted automatically. For example, weights can
be set to react to the user speed so that at lower speeds, elements attach to environment and at higher speeds they move along with the user. Weights can also be manipulated indirectly by the user, for example, to create default positions for elements.

A drawback of this approach is that it is susceptible to local minima and can create problems in narrow passages (Koren & Borenstein 1991). An interface element can get stuck in an equipotential field created between two walls, or in front of an obstacle. In addition, several elements moving together can create conflicts as they try to reach a location. As our goal was to develop a proof-of-concept to assess the viability of adaptation, we solved those issues by changing the scene model when necessary. In a more sophisticated system, global optimization could be implemented for less flexible and more robust solutions (Rao & Rao 2009).

4.5 AR ADAPTIVE WORKSPACE - VERSION I

Following the approach discussed in the previous section, we developed a prototype using Unity 2017 and two Microsoft HoloLens optical see-through head-mounted displays. The HoloLens is untethered and can track itself without the need of external systems. In addition, it can reconstruct nearby geometry for adaptation and support hand gestures for interaction, making it a good platform for our application. One HoloLens was worn by the user, while the other was used by the researcher to monitor the operation of the system. We also used an xBox 360 Bluetooth controller for system and behavior control.

IMPLEMENTATION DETAILS

Our adaptation strategy considered three information sources: user movement, environment surfaces, and the relationship between the windows in a given layout. Our general approach involved:

1. Assigning an attraction field to each wall and to the window initial position
2. Restricting the degrees of freedom of the window according to the behavior configuration
3. Assigning weights to the attraction fields according to the behavior configuration
4. Computing the potential field gradient at the window location
5. Setting the speed vector of the window to the negative of the gradient

The user pose was obtained directly from the HoloLens in real time. Since the HoloLens is head-worn, body orientation was estimated using the velocity vector. In addition, the position was filtered by a running average to remove head bouncing.

The environment information could also be obtained from HoloLens, by accessing the spatial mesh data created by the self-localization system. However, the scanned mesh may include people walking nearby, creating “ghosts” that need to be removed during processing. The mapping is also slower compared to average human walking and has a limited range\(^5\). Our solution was to manually indicate the location and extent of the walls and other surfaces of interest by drawing them in an editing mode. These surfaces are then saved as a HoloLens spatial anchor, and can be reused again when the room is recognized by the device. Each rectangular shape is associated with a finite equipotential surface during runtime (Figure 4-6).

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{wall_editor.png}
\caption{Wall Editor.}
\end{figure}

The two HoloLenses were networked using a text-based UDP/IP protocol. The reference frames were synchronized by setting spatial anchors at specific positions on the floor. We used several anchors distributed around, so that the system was not forced to use an anchor which was potentially too far to be reliably detected and tracked.

\footnote{\(^5\) According to the Microsoft documentation, the maximum distance for scanning is 3.1m, with a 70-degree cone (Microsoft 2018).}
We also added support for adjusting the position of the windows using the hands. The user can do the tap-and-hold gesture to grab the window in the center of the display. The window position is then adjusted to follow the relative movements of the hand. When the user opens the hand, the window is released at its current position, reoriented to face the user, and its original position is reset to the new location.

**Behavior Selection**

To reduce our initial design space to the most promising combinations, we conducted an informal formative evaluation (Scriven 1967) with four virtual environments experts. Different combinations and behaviors were tested. As they were implemented, they were evaluated for usability and potential utility to achieve our goal. We also investigated ad-hoc behaviors that could not be easily described by combinations, such as:

- Adapt to head orientation only if walking speed was greater than a specified value
- Adapt to the body position if the distance between the user and the element exceeded a threshold
- Stick to the walls if the distance from the element to a wall is less than a threshold
- Move elements and position them in front of the user if a button is pressed

The first iterations led us to combine parameters that did not work well independently and fix default values for a few behaviors. We decided to:

1. Combine the yaw and pitch in a single rotation parameter and ignore changes in head roll. After experimentation, we saw that users tend to understand yaw and pitch as part of the same integral space. On the other hand, we expected the up axis of elements to always point up so it was fixed to align with the up vector of the world.

2. Always detect collision with physical surfaces. We explored the options of having virtual elements ignore physical walls or to actually collide with them. We decided for the latter, since it would enable users to take advantage of physical surfaces in space and also prevent depth mismatch (seeing one thing inside another).
3. Try to always respect the relative layout of elements. One of the options discussed was regarding how much freedom to give to elements as they adapt to changes (in particular changes in the local physical space). If elements have complete freedom, they can always rearrange themselves optimally in relation to the available space. For example, during adaptation a larger element could go to a large wall and a smaller element to a small one. However, arbitrary rearrangement could generate confusion, since the user could not use spatial memory to quickly locate elements. A compromise was to allow adaptation but preserve the relative position of the elements with respect to each other.

4. Combine some of the behaviors into two new “smart” behaviors. A few ideas could not be implemented by activating basic options together. Auto-centering was a behavior designed to automatically move and align the windows in front of the user during walking, following a body-centric reference frame. The Attraction behavior moved and placed the elements along the wall closest to the user, if the user was reasonably close. If the user moved away, the elements would detach from the wall.

By the end of the process, we reduced the initial set of options to four basic behaviors that demonstrated high potential to support different goals and interaction styles:

- **Follow** – Adaptation to the user position. Maintains the relative positions of the element and the user.

- **Rotation** – Adaptation to the user orientation. Maintains the relative positions and orientations of the element and the user’s head.

- **Attraction** – Adaptation to surfaces. Moves the window to the surface closest to the user and changes the orientation proportionally to the wall distance.

- **Auto-Centering** – Adaptation to user movement while walking. Follows the user as in Follow, but rotates to maintain the relative orientation to the user body.
With these four behaviors, it was possible to use different strategies to access information while walking\(^6\). In addition, any behavior could also be temporarily activated (see next section).

**Behavior Configuration**

To explore different usage strategies, it was necessary to individually disable or enable the four behaviors implemented (follow, rotation, attraction, and auto-centering). We also felt that the ability to temporarily enable a behavior could lead to interesting uses. To achieve both goals, we implemented a configuration menu (Figure 4-7). The menu could be opened or dismissed by pressing the X button on the xBox controller. Once opened, users could air-tap the corresponding option to activate, deactivate, or assign a behavior to the controller’s B button. If a behavior was assigned to this button, it would be active as long as the B button was pressed. Multiple behaviors could be assigned to the button simultaneously. The configuration parameters were used for all the windows.

\[\text{Figure 4-7: Configuration Menu}\]

---

\(^6\) A video demonstrating these behaviors can be found in the support materials.
4.6 **Evaluating the AR Workspace**

In order to evaluate the adaptive workspace system, we designed two scenarios. Our goal was to create a setting that was realistic, and involve participants in tasks that were reasonably complex and believable.

**Scenarios**

In the first scenario, participants were asked to assume the role of an interior designer. They would then meet the owner of an apartment (experimenter) and review some of what was discussed in a previous meeting. In the second scenario, the participants would pretend to be talking to a friend (experimenter) about news and trying to schedule a day to meet at the beach.

The scenarios were designed so that a variety of content could be used. The information for both scenarios was distributed in the following applications: **web browser, list of prices, floor plan, notes, calendar**, and **weather forecast**. Each one consisted of a single static window pane (a texture). Each scenario required the use of three windows, two in conjunction and one isolated.

The interior design scenario required the notes, floorplan and price list (Figure 4-8). The experimenter would ask:

1. **What was decided regarding one of the rooms.**
   Example “Do you remember what we have decided about the kitchen?” This question required users to look at the notes and read the corresponding text.

2. **Which furniture costs a certain amount of money and where it will go in the room.**
   Example: “Which furniture costs $100? Where did we decide to put it?” This question required users to search the furniture in the price list and locate its position on the floorplan provided.

3. **To mark the position of the furniture on a floorplan drawing on a real whiteboard.**
As the last task in this scenario, we handed a physical marker to the participants and asked them to mark an X on a similar floorplan drawn on a whiteboard.

The second scenario required the weather forecast, calendar and webpage (Figure 4-9). The experimenter would ask:

1. **What is the news on a specific topic.**
   Example: “Do you know what’s in the technology news today?” This question required participants to read the corresponding headline on the webpage.

2. **When the weather will be good and the participant will be free on a given day of the week.**
   Example: “Do you think we could go to the beach in a Wednesday? What will the weather look like?” This question required participants to check which days they are free and then if the weather would be good to go to the beach.
The tasks, then, consisted of: 1) listening to and interpreting the experimenter’s question, 2) finding the window(s) with the required information in the layout, 3) retrieving the information, and 4) verbally expressing the answer. In most cases what followed was a natural conversation between the experimenter and the participant.

**Environments**

The study was conducted in two spaces in which the participants could walk alongside the experimenter. In Study I, each scenario was run separately. In the first scenario, we asked participants to walk around tables inside a large lab. The path was approximately 40m in length (Figure 4-10 top). In the second scenario, participants walked around 60m along a corridor (Figure 4-10 bottom). In Study II, both scenarios were merged into a single large scenario that was run entirely inside the lab, totaling around 40m (Figure 4-11).

![Room Path](image)

**Figure 4-10:** Paths used in the first study. A) Path used in the first scenario; B) Path used in the second scenario.
DATA COLLECTION

We collected different types of information during the study, and most of it was converted to text format so that it could be included in the coding process. Interviews were audio recorded and transcribed. The answers from the background questionnaires were converted to text, and the notes taken by the experimenter were attached to the file of each participant. We also collected interaction logs captured by the system, which we used as supplementary source to understand and confirm events.

INTERVIEWS

We conducted interviews after each trial. Interviews lasted between 5 and 15 minutes. We asked questions regarding participant’s choice of behaviors, layout of the windows, how much their expectations were met during the trial, etc. Since the interview happened in the same space, participants could point to the physical locations and objects and explain what happened in the critical moments of the trial. We also asked them to elaborate on differences between the most recent trial and previous trials. After the last trial, we conducted a final interview, and asked participants to judge and rank all the behaviors they tried. We also asked about what they liked, what they did not like and what they would like to see changed or added.
**Logs**

Our system automatically logged additional data. We recorded the trajectory of the participants in space, the configuration selected, and the behavior activations that happened through the joystick. We also recorded the head orientation of the participant and the position of all the windows throughout each trial. During the analysis, we used this log to replay each trial in 3D (Figure 4-12). This review was useful to contextualize the interview data and also generate new observations, which were then added to the participant’s text file.

![Figure 4-12: Playback of a recorded log.](image)

**Observation**

The experimenter used a second HoloLens to monitor what the participant was doing during the study. Besides the position of each window, the experimenter was also able to see the current behavior configuration and button presses from the participant. This was useful to help the participants, to detect problems, and to share a common ground with them.
4.7 Study 1

We recruited 14 university students with corrected or normal vision. Ten participants were males and four females. Eight of them had used AR just once or twice.

**Procedure**

Upon arrival, participants were asked to read and sign the consent form. Next, they were asked to fill out a background questionnaire containing demographic questions, a question about experience with VR, and another about experience with computer games. Next, the experimenter explained the goals of the study, described the scenario, and explained the interaction method. Next, they were introduced to the HoloLens, guided on how to adjust it, and on how to perform the HoloLens calibration procedure. The experimenter then walked the participants through each behavior, activating one at a time and explaining how each one worked. The experimenter also trained the participants to grab and move the windows around to configure the layout. Participants then performed a series of trials with each scenario. Each trial consisted of thinking and deciding which combination of behaviors to try, adjusting the windows, walking and performing the tasks, and joining in a semi-structured interview (Figure 4-13). The participants were instructed to follow the experimenter as he walked along the path at a comfortable pace. If the participant stopped to answer, the experimenter asked the participant to start walking again (if they did not do so after answering). The experimental material for this procedure can be found in Appendix D.
**Coding Method**

Coding of the data was performed in two cycles (Figure 4-14). In the first cycle, we used flexible methods, in order to be sensitive to the views expressed in the transcriptions. For the most part, we employed Open Coding. Open Coding (also known as Initial Coding), is a process in which codes are created by breaking down the data into parts and looking for commonalities and differences (Saldaña 2015). The codes were developed following the principle of “constant comparisons” to ensure consistency (Corbin & Strauss 1998). Some of them were divided into a positive and a negative aspect to facilitate analysis.

We also applied Holistic Coding to identify passages related to our probe design (e.g., behaviors, configurations, etc.) and Structural Coding to identify answers from recurrent interview questions. We applied descriptive codes for other passages (usability of the behaviors, preferences, strategies adopted by the participants, reading, content, safety concerns, etc.). The codes for demographic information were singular and were used to characterize each participant file (Attribute Coding). Some super-ordinate codes were also created to identify passages as belonging to different trials and locations. These codes were used later to facilitate sub-coding. We used the qualitative analysis software AQUAD 7 to perform the coding in approximately 400 lines of text and generated 90 initial Level 1 (L1) codes.
In a second cycle, the L1 codes were grouped into categories. To investigate characteristics of the behaviors, we applied Axial Coding (Figure 4-14). In this method, the fractured data from the first cycle is recombined to retrieve larger categories (Saldaña 2015). We grouped codes along attribute dimensions and applied the C-family coding (Böhm 2004) (causes, contexts, contingencies, consequences, and conditions) to identify possible causal conditions, contexts and consequences of phenomena. We also looked for patterns expressed by expressions such as “if/else”, “because” that could help to understand participant’s recurrent patterns (Pattern Coding). Aggregated code categories are listed in Appendix A, Table 0-1.

**FINDINGS**

The analysis resulted in several categories and themes. We discuss the ones most related to our research problem.

**Use of combinations**

To focus the analysis, we first compared the combination codes in importance. We looked at how frequently each combination code appeared in data from different participants in the first (Table 4-2) and in the second (Table 4-3) scenarios. The most used combinations were [Auto-centering], [Follow-On, Rotation-Button], and [Follow-On, Attraction].

**Figure 4-14: Diagram of the analysis process**

In a second cycle, the L1 codes were grouped into categories. To investigate characteristics of the behaviors, we applied Axial Coding (Figure 4-14). In this method, the fractured data from the first cycle is recombined to retrieve larger categories (Saldaña 2015). We grouped codes along attribute dimensions and applied the C-family coding (Böhm 2004) (causes, contexts, contingencies, consequences, and conditions) to identify possible causal conditions, contexts and consequences of phenomena. We also looked for patterns expressed by expressions such as “if/else”, “because” that could help to understand participant’s recurrent patterns (Pattern Coding). Aggregated code categories are listed in Appendix A, Table 0-1.

**FINDINGS**

The analysis resulted in several categories and themes. We discuss the ones most related to our research problem.

**Use of combinations**

To focus the analysis, we first compared the combination codes in importance. We looked at how frequently each combination code appeared in data from different participants in the first (Table 4-2) and in the second (Table 4-3) scenarios. The most used combinations were [Auto-centering], [Follow-On, Rotation-Button], and [Follow-On, Attraction].
Table 4-2: Combinations most tried in scenario 1

<table>
<thead>
<tr>
<th>Follow</th>
<th>Rotation</th>
<th>Attraction</th>
<th>Auto-Centering</th>
<th>Count</th>
</tr>
</thead>
<tbody>
<tr>
<td>On</td>
<td>Button</td>
<td></td>
<td>On</td>
<td>8</td>
</tr>
<tr>
<td>On</td>
<td></td>
<td>Button</td>
<td>On</td>
<td>7</td>
</tr>
<tr>
<td>On</td>
<td></td>
<td></td>
<td>On</td>
<td>4</td>
</tr>
<tr>
<td>On</td>
<td>Button</td>
<td></td>
<td>On</td>
<td>3</td>
</tr>
<tr>
<td>On</td>
<td>Button</td>
<td></td>
<td>On</td>
<td>2</td>
</tr>
<tr>
<td>On</td>
<td>Button</td>
<td>On</td>
<td>On</td>
<td>1</td>
</tr>
<tr>
<td>On</td>
<td>Button</td>
<td></td>
<td>Button</td>
<td>1</td>
</tr>
<tr>
<td>Button</td>
<td>Button</td>
<td>On</td>
<td>Button</td>
<td>1</td>
</tr>
<tr>
<td>On</td>
<td>Button</td>
<td>Button</td>
<td>Button</td>
<td>1</td>
</tr>
</tbody>
</table>

Table 4-3: Combinations most tried during scenario 2

<table>
<thead>
<tr>
<th>Follow</th>
<th>Rotation</th>
<th>Attraction</th>
<th>Auto-Centering</th>
<th>Count</th>
</tr>
</thead>
<tbody>
<tr>
<td>On</td>
<td>Button</td>
<td></td>
<td></td>
<td>5</td>
</tr>
<tr>
<td>On</td>
<td></td>
<td></td>
<td>On</td>
<td>4</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>On</td>
<td>2</td>
</tr>
<tr>
<td>On</td>
<td>Button</td>
<td></td>
<td>On</td>
<td>2</td>
</tr>
<tr>
<td>Button</td>
<td>Button</td>
<td></td>
<td>On</td>
<td>1</td>
</tr>
</tbody>
</table>

Behavior Preferences

We also looked into the individual ranks given by the participants. After the final trial, we asked participants to order each behavior in terms of decreasing utility. We then computed the average rank, assigning weights 3, 2, and 1 to the first, second, and third behaviors (follow was not considered since it was necessary to complete the task). Rotation was considered the most useful, followed by Auto-Centering and Attraction (Figure 4-15).
Behavior Characteristics and Perceptions

To investigate the reasons for the preferences, we sub-coded all passages describing those behaviors. This allowed us to capture specific meanings, to keep track of the bigger picture, and to establish dimensions for those codes. Table 4-4, Table 4-5, and Table 4-6 list the codes which appeared at least twice during the analysis. The Auto-Centering behavior was deemed convenient and easy, but had the perceived disadvantage of taking some time to respond and not allowing precise control of the windows. On the other hand, the combination [Follow-On, Rotation-Button] was judged as fast and precise, at the expense of requiring manual control. The attraction behavior was considered safe and appropriate for stationary conditions. However, it was hard to predict where the windows would attach and it was not useful on lateral walls.
Table 4-4: Most frequent codes for the Auto-Center behavior

<table>
<thead>
<tr>
<th>Code</th>
<th>Representative Quotations</th>
</tr>
</thead>
<tbody>
<tr>
<td>Takes too long to align</td>
<td>“I would turn they would not show up. It might take a second for them to show”</td>
</tr>
<tr>
<td>Does not allow control of</td>
<td>“I found that if you are using auto-centering and you stop, you don’t get the</td>
</tr>
<tr>
<td>position</td>
<td>windows where you want”</td>
</tr>
<tr>
<td>Is convenient</td>
<td>“So, auto-centering I like because it always kept the information in front of me ...”</td>
</tr>
<tr>
<td>Works as long there is no</td>
<td>“It worked because we did not do anything in the real world, only read information from</td>
</tr>
<tr>
<td>interaction</td>
<td>the virtual”</td>
</tr>
<tr>
<td>Allow freedom to look around</td>
<td>“The rotation has a problem that it follows your gaze. I found that auto-centering was</td>
</tr>
<tr>
<td></td>
<td>the best option”</td>
</tr>
<tr>
<td>Is easy</td>
<td>“The auto-centering was pretty straightforward”</td>
</tr>
<tr>
<td>Hard if walking fast</td>
<td>“Some windows were moving a lot...I needed to slow down a little bit to read”</td>
</tr>
</tbody>
</table>

Table 4-5: Most frequent codes for the Rotation on B behavior

<table>
<thead>
<tr>
<th>Code</th>
<th>Representative Quotations</th>
</tr>
</thead>
<tbody>
<tr>
<td>Allows precise control (with</td>
<td>“when you press the b-button [assigned to rotation] you get the windows immediately</td>
</tr>
<tr>
<td>button)</td>
<td>where you want”</td>
</tr>
<tr>
<td>Restrictive if always on</td>
<td>“If I leave it on, I cannot see all the windows”</td>
</tr>
<tr>
<td>Fast</td>
<td>“So, I like to basically tap B to reset it quickly in front of me”</td>
</tr>
<tr>
<td>Requires automatism</td>
<td>“Sometimes I forgot about pressing the b-button”</td>
</tr>
</tbody>
</table>

Table 4-6: Most frequent codes for the Attraction behavior

<table>
<thead>
<tr>
<th>Code</th>
<th>Representative Quotations</th>
</tr>
</thead>
<tbody>
<tr>
<td>Is hard to predict/ control</td>
<td>“The attraction was hard to know which wall it was going to pick ...”</td>
</tr>
<tr>
<td>Is good on frontal but not</td>
<td>“Attraction is not very useful if you are not moving towards the wall”</td>
</tr>
<tr>
<td>lateral walls</td>
<td></td>
</tr>
<tr>
<td>Is appropriate for stationary</td>
<td>“I can see that as pretty useful function if you are on a table.”</td>
</tr>
<tr>
<td>conditions</td>
<td></td>
</tr>
<tr>
<td>Makes walking safer</td>
<td>“You are not going to walk through a wall anyway, so it is safer”</td>
</tr>
<tr>
<td>Required stopping</td>
<td>“I could not really walk and read the windows. I need to stop”</td>
</tr>
</tbody>
</table>
In the second cycle, we engaged in axial coding, searching for relationships between the categories and identifying thematic axes. We found three major dimensions which characterize the appropriateness of the behaviors for the scenarios we studied: 1) compatibility with other real-world activities, 2) compatibility with vision-body coordination involved in walking, and 3) the tradeoff between usability and expressiveness of the interface (Figure 4-16).

**Figure 4-16: Core themes resulting from the axial coding**

- **Compatibility with other real-world activities** - Participants noted, during the study, that having many windows too close blocked the view of the surroundings. It also made interaction with the world harder. For example, when talking to someone or manipulating object on a desk: “if I need to be doing some delicate operation it would be more user-friendly keeping them on the closest wall and when I require I could just look at that.” Participants also mentioned during the last phase of Scenario 1 that it was hard to get the windows in a good place when using auto-centering: “auto-centering is not very useful mostly because it can get between you and other people very easily and prevent you from seeing what is happening at close quarters”.


• **Compatibility with vision-body coordination involved in walking** - One of the challenges of the task was to read the content while navigating around desks and other obstacles in the room. Even for those with some familiarity with the space, walking was hard if they could not look where they were walking. “if the windows takes [sic] the whole view of the corridor, you would not be able to see if someone was coming”. Even though most participants claimed they could see clearly through the windows, a few participants mentioned that they would like the windows to “occlude less of the actual visual space.” “The good side [of attraction] is that it was not always popping in front of me. I could see the space more clear in front of me…where I am going.” Participants also mentioned that it was hard to read a window on the side (which was positioned there, in the process of moving to another location, or attracted to a lateral wall). “Attraction is not very useful if you are not moving towards the wall, because you need to keep your head at 90 degrees’ angle, and you cannot see what is front of you. You are more likely to bump into objects.”

• **Tradeoffs between usability and expressiveness** - The main advantage recognized by the users for the auto-centering and attraction modes was the fact that they did not require any input form the user. Some participants felt that using the button was an additional overhead: “I liked auto-centering better, because with the rotation I had to keep pressing the b-button.” Occasionally, they forgot to press the button and one expressed the concern that it could become tiring over time. The ability to perform multiple functions was recognized for the [Follow-on, Rotation-button] configuration as a benefit.

**Interaction Patterns – How did participants use the behaviors?**

To find out how participants were using the behaviors, we looked into the L1 codes for those related with participants’ actions. We applied the C-Family coding (context, causes, contingencies, consequences and conditions) to reconstruct the temporal course of actions with the goal of achieving a specific purpose (Glaser 1978; Böhm 2004; Strauss & Corbin 1990).
We considered “actions” the participant’s response to a phenomenon. Following the recommendation of Strauss and Corbin we did a functional analysis and considered both conscious and unconscious actions (Böhm 2004). The events or conditions that led to the phenomenon and participant action were considered “causes”. If a condition acted as a modifier to the phenomenon, or delimited the set of possible user action, they were classified as a part of the “context” (Corbin & Strauss 1998).

**Attraction to Layout:** “When we started walking I kinda put them on the wall, because they zoom away and you have a better field of view.”

**Attraction to See:** “if you put them on the walls, you are not going to walk through a wall anyway, so it is safer”. “If I need to be doing some delicate operation it would be more user-friendly keeping them on the closest wall”

**Rotation to Move:** “When I was doing the floor map, I got a little time to align then together [tried to place the virtual floorplan on top of the drawing].”
**Tap Follow/Rotation:** “I like to basically tap B to reset it quickly in front of me” “If I want them immediately because I want to gather some information I could press follow and have the in front of me.”

**Auto-Centering:** “If I want them immediately because I want to gather some information I could press follow and have the in front of me.”

**Adjust Window Layout:** “I felt that I needed to move the windows, because there was another window in front of the one I needed.”

**Head Turn:** “Attraction was used and understood as a good behavior to use in a frontal wall … if not in the front, it caused strain”

**Visual Search:** “I felt that I need to stop because I put it large, so I need to stop walking to find them. The notes and the price list.” “But the price list has twelve, something, and I need to rotate my head because it was out of the field of view.”
Interaction Framework – To what end were behaviors used?

We found four core categories related to the interaction patterns participants adopted during the study.

**Core Category 1: Obtaining Information**

This category groups the strategies used to obtain the information necessary to answer the questions asked during the study. Participants stated and were observed using the combination [Follow-On, Rotation-Button], [Follow-Button, Rotation-Button], and Auto-centering to achieve this goal. All these combinations brought the windows to the locations the participant previously established. The tradeoff was the occlusion of the surrounding by the windows. In addition, when they were close enough to read, it was not possible to fit all the windows in the field of view, which required the participants to turn their heads.

**Core Category 2: Visualizing Information**

This category groups the actions taken to organize the windows so that they could be clearly visible and help participants find where to obtain the information they needed. If the target window was close, but in an undesired position, some participants would drag to rearrange them again. This often happened because the initial configuration set the windows was too far away or on one of the sides. Hand adjustment of the layout also happened when windows would accidentally overlap each other. Another strategy used to organize information was to open the Menu and enable Attraction so that the windows would move to the nearest wall.

**Reduced Awareness:** “I still can see the environment but I walk slower because of my attention.”
Since switching between [Follow-Button, Rotation-Button] or Auto-centering and Attraction was not possible without opening the Menu, this was not a common solution. However, participants did express the desire to switch between those (see suggestions analysis in this section).

Core Category 3: Compensating for Reduced Awareness

This category groups the actions taken to compensate for the lack of visibility of the physical world or the toll of attention caused when focusing on the windows to obtain information. When faced with this issue, participants either slowed down the pace or stopped. The lack of sensory data about the world happened for two reasons: 1) when users tightly packed the windows so that they could not see very well through it 2) when users were not looking in the direction they were walking (because the information was on the side or on a lateral wall). In addition, participants also indicated they slowed down or stopped because of the attention spent on reading / thinking about the content (even if it was right in front). In all cases, the act of slowing down and eventually stopping seem to have happened both consciously and unconsciously, as some participants reported to be unaware of doing so. Enabling attraction was also used as a way to cope with the lack of awareness of the environment. By moving the windows to the wall, a participant reported he was more comfortable walking.

Core Category 4: Harmonizing Augmented Reality with the World

This category groups the strategies used with the goal of making other interactions with the world more sensible or efficient. For example, when writing on the physical board one participant wanted to align the virtual floorplan with the physical floorplan to compare. Participants also expressed the desire to move windows out of the way to talk to someone or otherwise see the environment to perform other tasks. Two strategies were used or mentioned: 1) use the combination [Follow-On, Rotation-Button] to manipulate the windows and adjust their position in space, and 2) move the windows to the walls so that they naturally integrate with the physical surfaces. Figure 4-17 illustrates the detailed interactions that were observed between the core categories.
Figure 4-17: Interaction between the core categories

Other Related Topics

Suggestions for Interface Improvement

We asked participants to imagine which features they would like the interface to possess or how they would like behaviors to operate in an ideal case. The codes identifying suggestions were reviewed to find common themes (Table 4-7). Regarding the behaviors, two recurring suggestions were 1) to make the adaptation faster: “I would like for auto-centering to perform faster. Because now it takes a while” and 2) modify the attraction so that it would always attract to the frontal wall: “It should be rather than the closest wall, it should display the things on the wall which I am looking at.” Participants were also distracted by the number of windows and suggested that there should be a way to close the ones not being used: “The biggest problem is that I cannot have many windows in front of me. At most three windows in order to read information from that.”
### Table 4-7: Frequency of suggestion codes

<table>
<thead>
<tr>
<th>Code</th>
<th>Applicability</th>
<th>Description</th>
<th>Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>Frontal Wall</td>
<td>Attraction</td>
<td>Attract to the frontal wall instead of the closest one or control</td>
<td>6</td>
</tr>
<tr>
<td>Less Windows</td>
<td>General</td>
<td>Reducing the number of windows, providing means of closing them, or reducing the “space” they occupy</td>
<td>4</td>
</tr>
<tr>
<td>Switch Behaviors</td>
<td>General</td>
<td>Way to quickly switch between different behavior combinations</td>
<td>4</td>
</tr>
<tr>
<td>Faster</td>
<td>Auto-Centering / Attraction</td>
<td>Making the adaptation faster when moving to the walls or when the user rotates</td>
<td>3</td>
</tr>
<tr>
<td>Natural Interaction</td>
<td>General</td>
<td>Using gestures or voice instead of the controller to active behaviors</td>
<td>2</td>
</tr>
<tr>
<td>Freeze</td>
<td>General</td>
<td>Way to freeze the windows in a position</td>
<td>2</td>
</tr>
<tr>
<td>Transparency</td>
<td>General</td>
<td>Control the transparency of the windows instead of moving them away</td>
<td>1</td>
</tr>
</tbody>
</table>

*Perception of the Visual Aspects*

The windows used in the study had a wide variety of content, text size, typography, colors, and layout. To make sure the content itself was not imposing an undue burden, we asked participants to comment on the visual aspects and on content legibility when stationary. All the participants reported that reading the content on the windows was “easy,” “pretty easy,” or “very easy.” The windows containing the news, webpage, and floorplan were considered the easiest ones. One participant mentioned that the weather window had too much information, and another that the text was small. The size of the text on the price list was also mentioned by one participant to be small. Another participant mentioned that the handwritten typography on the notes did not help. We observed that once the desired window was located, obtaining the information was a two-step process. First the participant need to find the information inside the window. For example, the notes required the user to find the corresponding room name in the list, then read the associated information. Although all the content could be read easily, some windows were more amenable to the searching step: “Because [the price list] have several items, I needed to scan them to find the one that had the same price.” For this, the structure of the information was essential: “The calendar was very fine because it was...
patterned.” “I guess the floorplan and the notes were pretty easy to read. Because of the content. The notes the font was easy to read, could scan top/down.”

**Task Effect on Walking Performance**

We also observed participants evaluating the impact of the tasks on their performance navigating the environment. Most participants adapted their walking patterns by slowing down or stopping altogether. Some of them mentioned that: “I feel that stopping is safer. I don’t want to hit an object that I did not pay attention to,” or “the auto-centering seemed to obstruct my vision while walking so I was more worried about bumping into something.” Although the majority of the participants were able to keep away from obstacles, three participants did collide with objects (with no serious consequences). Two participants hit a table and another one hit the leg of a tripod.

**SUMMARY**

In summary, this study allowed us to identify four fundamental interaction patterns in AR walking scenarios. Those comprehend the process of visualizing the information layout, obtaining the desired information, compensating for the lack of awareness of the environment, and adjusting virtual information so that it is presented harmonically with the world.

Most participants appreciated the ability to have information with them while walking, in particular to quickly look up information. The majority also opted for automatic following, instead of moving the windows on demand. Auto-centering was considered convenient and straightforward to use, and the combination [Follow-On, Rotation-Button] was deemed more precise. Among the limitations, participants mentioned the fact that the attraction to the nearest wall was not intuitive. Many also expressed the desire to switch between different combinations while walking, which our system did not support.

In addition, we were able to isolate the ideal set of properties interfaces should have to be effective. We found that interfaces should 1) be compatible with the vision-body coordination involved in walking, 2) be compatible with secondary real-world activities, and 3) display a good tradeoff between usability and expressiveness.
4.8 AR Adaptive Workspace – Version II

With Study I, we found that while trying to accomplish the scenario tasks, participants engaged in a more fundamental set of interaction patterns. However, since our system was not specifically designed to support those patterns, they reached the system’s limits while doing so. After a reflection, participants suggested different changes such as the ability to quickly switch between sets of behaviors, changes in specific behaviors, and other improvements. In this section, we describe how we used our general findings and participant feedback to guide the design and development of an improved version of the system.

Selecting the Changes

To evaluate the impact of implementing each suggestion in our scenario, we crossed the effect of the most frequent suggestions against each of the goals we isolated in Study I (We decided not to pursue the other two suggestions. Even though faster adaptation was mentioned by many participants, this modification would not allow qualitative changes in the interaction (allow them to do new things) and would only reduce the time to acquire information (since participants would wait less time for the windows to move). The other suggestion we decided not to implement was reducing the number of windows. That would likely make the task much easier and introduce new interaction aspects which were not the focus of this study. In addition, in the second study we planned to run a single scenario involving all the six windows.

Table 4-8). From that we chose to implement two changes:

1. The ability to quickly switch between behaviors: Implementing this suggestion would allow participants to choose on demand the appropriate behavior to support the current goal. For this reason, we expected it to have the highest impact among all changes.

2. Modify the Attraction behavior to use the frontal wall: Attraction was used (or mentioned) in the context of three of the four main goals. Improving it would reflect directly on the ability to execute them.
We decided not to pursue the other two suggestions. Even though faster adaptation was mentioned by many participants, this modification would not allow qualitative changes in the interaction (allow them to do new things) and would only reduce the time to acquire information (since participants would wait less time for the windows to move). The other suggestion we decided not to implement was reducing the number of windows. That would likely make the task much easier and introduce new interaction aspects which were not the focus of this study. In addition, in the second study we planned to run a single scenario involving all the six windows.

Table 4-8: Correspondence between suggested changes and goals

<table>
<thead>
<tr>
<th>Change</th>
<th>Organize</th>
<th>Obtain Information</th>
<th>Compensate for reduced awareness</th>
<th>Harmonize AR/RW</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ability to switch behaviors*</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Attraction on frontal wall*</td>
<td></td>
<td>X</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>Reduced number of windows</td>
<td>X</td>
<td></td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>Faster adaptation</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

We also reviewed the analysis of the individual behaviors. The main advantage cited for the combination [Follow-On, Rotation-Button] over Auto-Centering and Attraction was the ability to precisely control the position of the windows by temporarily pressing the button. So, we decided to provide a similar functionality for the Attraction behavior, allowing the users to specify the position of the windows on the wall.

**IMPLEMENTING THE NEW BEHAVIORS**

Our changes extended the automatic Auto-Centering and Attraction behaviors to make them adjustable by using two controller buttons. Pressing the corresponding button also allowed users to toggle between them. In Mode A, windows would auto-center if the A button was not pressed. Pressing it temporarily switched to the Rotation mode, allowing
the user to instantly position the windows along the new view direction (Figure 4-18 A). In Mode B, windows would be attracted to the current wall if the B button was not pressed. By pressing the B button, a combination of Attraction and Rotation was engaged, allowing the user to position the windows in a specific place on the wall being viewed (Figure 4-18 B). In both modes, Follow was always enabled.

**Adjustable Automatic Behaviors**

![Diagram of adjustable behaviors](image)

**Figure 4-18: Overview of adjustable behaviors. During Mode A (left), the button press enables control of the windows using rotation. During Mode B (right), the button press enables control using a combination of attraction and rotation.**

To give more freedom to the user and attend to one of the suggestions, we also included a Freeze button, which disabled all behaviors (including Follow), making the windows stay still in space. The final state machine is shown on Figure 4-19.

---

7 A video demonstrating these behaviors can be found in the support materials.
4.9 Study II

We recruited nine university students with corrected or normal vision. Six participants were males and three were females. Four of them had used AR just once or twice. In this study, the participants followed a single path within the room and were asked all the questions in the first study, thus requiring the use of all windows. In addition to the whiteboard task at the end (where participants were asked to mark the location of the furniture), this combined scenario also asked participants to sit in a chair while talking to the experimenter about possible days to go to the beach.

Procedure

Upon arrival, participants were asked to read and sign the consent form. Next, they were asked to fill a background questionnaire containing demographic questions, a question about experience with VR and another about experience with computer games. Then the experimenter explained the goals of the study, described the scenario and interaction method. Next, they were introduced to the HoloLens, guided on how to adjust it and on how to perform the HoloLens calibration procedure. The experimenter then walked the
participants through the two modes, activating one at a time and explaining how each one worked. The experimenter also trained the participants to grab and move the windows around to configure the layout. Participants then performed a series of trials. Each trial consisted of adjusting the windows, walking and performing the tasks, and a semi-structured interview (Figure 4-20). The participants were instructed to follow the experimenter as he walked along the path at a comfortable pace. If the participant stopped to answer, the experimenter asked the participant to start walking again (if they did not do so after answering). The experimental material for this experiment can be found in Appendix D.

![Diagram of the experimental procedure of Study II.](image)

**CODING METHOD**

The coding was performed as in Study I. In the first cycle, the textual data was coded per sentence level using open coding and also theoretical coding (Thornberg & Charmaz 2014). We applied theory codes using the following definition:

**Visualize Information** – When participants expressed doing an action with the purpose of finding specific windows or visualizing all the windows.

**Obtaining Information** – When participants expressed doing an action so that they would be able read some specific content.
Compensating Reduced Awareness – When participants expressed doing an action so that they could observe the environment to walk.

Harmonizing with the world – When participants expressed doing an action so that they could do something besides walking, or so that windows felt natural in the world.

**FINDINGS**

Here we report the main findings of Study II.

Use of Modes

To understand how participants used the different modes, we developed a code matrix crossing each theoretical code to the four interaction options. Table 4-9 shows the use reported by each participant.

**Table 4-9: Participants mention to the use of each mode for a specific goal**

<table>
<thead>
<tr>
<th></th>
<th>Mode A</th>
<th>Mode B</th>
<th>Freeze</th>
<th>Adj.Layout</th>
</tr>
</thead>
<tbody>
<tr>
<td>Organize Information</td>
<td></td>
<td>P2, P3, P6, P7</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Obtain Information</td>
<td>P3, P5, P6</td>
<td>P3</td>
<td>P2</td>
<td>P4, P5</td>
</tr>
<tr>
<td>Comp. for Awareness</td>
<td>P3</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Harmonizing AR/RW</td>
<td>P3, P5, P8, P9</td>
<td>P3, P5, P7</td>
<td></td>
<td>P5</td>
</tr>
</tbody>
</table>

All actions associated with Mode A were regarding obtaining information. Mode B, on the other hand, was used for all actions previously identified. However, it was mostly used to organize information and match windows with the real world. The option to freeze the windows was most used as a resource for harmonizing with the environment. Both freezing and adjusting their position by hand, though, were also used to obtain information. Table 4-10 lists representative quotes from this study.
Table 4-10: Quotations of the use of each goal code

<table>
<thead>
<tr>
<th>Code</th>
<th>Representative Quotations</th>
</tr>
</thead>
<tbody>
<tr>
<td>Organize Information</td>
<td>“I would say that B was cooler, just because it added the zoom away aspect so I can actually see everything all together”</td>
</tr>
<tr>
<td>Obtain Information</td>
<td>“The A button is good to get [the windows] immediately in front of you.”</td>
</tr>
<tr>
<td>Comp. for Awareness</td>
<td>“If I was required to do more physical tasks, perhaps like running … I don’t want the windows to be anywhere in my field of view”</td>
</tr>
<tr>
<td>Harmonizing AR/RW</td>
<td>“Fixing the window on a wall is similar to real life. We usually post something on the wall”</td>
</tr>
<tr>
<td>Freezing</td>
<td>“Freezing windows made me feel slightly like the object [window] is on the real world”</td>
</tr>
</tbody>
</table>

The results suggest that the modes were used as predicted by the interaction framework identified in Study I, with Freezing replacing the use of [Follow-On, Rotation-Button] in the “Rotation to Move” category. Figure 4-21 illustrates how the two modes and freeze were used to achieve the goals.

![Diagram](image)

**Figure 4-21:** Use of the new behaviors. Modes A, B, and Freeze were used in the expected way to accomplish the interaction goals.
Mode A was observed as being the main mode for obtaining information. On the other hand, its tradeoffs were explicitly mentioned by the participants: 1) not being comfortable to get an overview of the windows: “When I press A they are too close to me and I need to turn my head to see everything,” and 2) impairing the ability to walk/run: “If I was required to do more physical tasks, perhaps like running, where I had to be very aware of my surroundings, perhaps I don’t want the windows to be anywhere in my field of view.”

The fact that participants did not use Mode B for obtaining information is also in accord with our previous study, basically due to the lack of walking compatibility. The same can be said of its extensive use in helping to organize the information and harmonize it with the world. We noted, however, that only one participant reported an action which would indicate the use of Mode B to compensate for reduced awareness while walking (Participant 03). We believe that this is a reflection of the study design: since there were more tasks in the same path, participants had little time to only walk. They were engaged in acquiring information most of the time. This one participant became very proficient in switching between the two modes, and adopted a reverse interaction style, stopping to read on the wall when possible: “When I need to focus on specific information I fix it on the wall, so that I don’t need to pay attention to objects in the world.”

Freeze was considered a good feature when used stationary. This acted as a replacement of the [Follow-On, Rotation-Button] present in the initial design. It was used either while in the Mode A or Mode B to place the windows in specific position and decouple the adaptation from the user. In Mode A, an observed strategy was to activate the rotation behavior (by pressing the button) and then freezing them out of the way. In Mode B, it was used to immobilize the windows on a wall while the participant walked closer to them or walked away to draw on the board.

Finally, we noticed two strategies for the hand interaction. The first was to finely adjust the windows to specific positions (e.g., aligning it on the board to facilitate the comparison). The second reason was to bring windows closer which were too far during the task. Some participants tried to place the windows slightly further away so that they could see more of them inside the field of view.
Attention and Walking Performance

This time, only one participant tripped on a video cable which was on the floor. Initially he reported being afraid to look down and move the windows out of place:

“Only once, when we were walking there. It was pretty narrow. That was the only time when I felt - Am I going to trip? Am I missing something? - Otherwise in the open space it was OK. This is the first time I used this. I was reluctant to look down to see that there was nothing, because I did not want the windows to move down. So, I would stay looking forward so that they stay in front of me. I think I did not have enough confidence” (emphasis ours).

In this particular occasion, he was aware that he did not have enough information about his surroundings and consciously avoided looking down. After being explained that the windows would not move along the y-axis, he became more confident to look around while walking:

“This time I moved my head more, in terms of looking at the environment, looking down, I was more comfortable doing that”

However, as he became more confident, he started to walk faster and dedicate more attention to the task. Eventually, he tripped:

“I actually did trip this time, so I was not entirely aware of my environment. Perhaps if I did not have the glasses, I could have realized there was a cable on the floor. Maybe ... I guess the other times I was stopping to read because of attention and focus. This time I was more comfortable” (emphasis ours).

This time he knew that he could look down, but was distracted by the attention given to the scenario tasks. This process of gaining confidence and neglecting the environment was similar to what was reported by one participant of the first study, who mentioned:

“I was able to read the information quickly because of the auto-center. Very nice. I did not feel that I need to stop and that is why I bumped into the objects, because it was very easy to read while walking, but I learned to stop and pay attention” (emphasis ours).

Attention and Walking Performance

Again, the content of the windows was considered easy to read: “The notes was easy. The weather was fine, but it had two stages so it took some time. The price has less text so it
was easier to read.” The webpage with the news, however, gave more trouble to the participants:

“If things are spaced out they are easier to read. The news were hard to read. The prices were definitely easier” – Participant 05

“The one that gave me some trouble was the news. When you asked me about science, I needed to look around to find the science.” – Participant 07

“When I was walking, I could not read science because there was a lot of information on that page” – Participant 04

“The news is harder because I am not used to read news in English” – Participant 06

The reason seems be due to the way information was laid out in the News window, and search time was what actually made it hard. In addition, two participants noticed issues with rendering quality while walking (color separation): “As I walked, because the windows were following me, I saw that they were a little blurry. I stopped just because they were blurry.”

**SUMMARY**

This second study allowed us to evaluate the effectiveness of the new adjustable behaviors and gather additional evidence for the overarching interaction patterns discovered in our first study. We observed that users had no difficulty using the two behaviors to achieve their goals. As expected, users predominantly used Mode A for accessing information while walking, and Mode B for managing the information complexity and position in space. However, we found little supporting evidence of actions to compensate for the lack of awareness while walking, which we believe was a consequence of the changes in the study design.

4.10 **DISCUSSION**

Overall, our research showed that position adaptation is a promising way to tackle the challenges of walking. Participants had no issues learning our final set of behaviors or using them to accomplish the tasks. However, we learned that some properties are
important in the performance of these behaviors, such as stability, predictability, speed, and control. Our implementation could still be optimized in those aspects.

Due to the use of potential fields, our windows moved smoothly and with low acceleration. Thus, they took some time to rest at final positions. The auto-centering behavior also had a small latency due to the way we decoupled the body movement from the head movement (simple average). We believe that the smaller accelerations helped users to track the windows before they left the field of view. However, quickly moving the windows (instead of using a potential field) should allow techniques to perform better.

In the final iteration, participants had individual control of each mode through different buttons on the controller. This design choice gave them freedom to use the modes individually and also to adjust the position of the windows without switching modes. Since both modes showed themselves useful in the end, an application could further simplify the input by using just one button. This button (or gesture) could be used to toggle between modes and still allow position adjustment during the transition.

Looking at the findings of the first and second studies, we can also begin to formulate a model to link the effects we observed (Figure 4-22). At the center we have the dual-task required by the scenario, which consists of obtaining information while walking (Figure 4-22 a). When confronted with the need of completing both tasks, participants often slowed down or stopped. A few became distracted and bumped or tripped on obstacles. These observable task impacts (Figure 4-22 b, c) were caused by reduced capacity to dual-task. However, participants could also perform intentional actions to compensate (Figure 4-22 d).
Figure 4-22: Hypothesized relationship between user actions, dual-task costs, and task impacts. Participants engaged in a dual task (a) which created cognitive and motor impacts (b, c). On the other hand, they actively compensated for those challenges by using the interface to achieve specific goals (a, d).

The primary impact of the dual-task was a performance penalty on walking. This result is consistent with previous reports of reduced walking speed in dual-task conditions, which was observed for diverse cognitive tasks and in different populations (Al-Yahya et al. 2011). Another result was a higher attentional demand, which resulted in the lack of environmental awareness (c). This observation is consistent with the participants’ notes about “lack of attention” which eventually led to a few trips and bumps. It has been shown that users talking on the phone or texting have reduced awareness (Lin & Huang 2017; Lim et al. 2015). However, it is also possible that they may have been caused by the direct impact in the gait parameters by the dual-task (Yoge-Seligmann et al. 2008a).

Some participants reported difficulty seeing the environment through the windows due to occlusion. In some cases, participants compensated the lack of awareness by moving the windows away. In others, they slowed down. For the participants who did try to observe the environment, but did not engage in any active strategy to manage the interface, the layout of the windows might have helped. As users were free to adjust the windows, they could opt to leave more of the visual field empty. However, even a good layout required users to switch the gaze between the windows and the environment. Although locomotion
does not require continuous input from vision (Laurent & Thomson 1988), the study design placed high emphasis on the cognitive tasks.

Finally, several participants intentionally slowed down or stopped to answer questions. Some could be afraid to trip and some mentioned that visual artifacts made reading uncomfortable when they were moving too fast (e.g., color separation). Additional studies are needed to separate these different causes and also separate more clearly intentional and unconscious impacts.

4.11 Designing For Walking

Our study revealed fundamental interaction patterns and qualitative features of the techniques for the usability of techniques for AR workspace management. These can be combined into the following design guidelines:

Incorporate support to compensate for reduced awareness and matching with the real world by:

- **Allowing content positioning** – Users may need to move AR content to a specific position in space to reflect new priorities, increase the comfort of the current task, or increase the semantic association with the world.

- **Allowing physical interactions** – Unlike VR, AR users are still engaged in the world. They may want to interact with objects (e.g., write something), engage in a conversation, etc. Make sure that AR interface allows unobstructed view when necessary.

- **Not occluding the walking path** – As part of world activities, users will walk around. To give users the confidence to walk, the interface should not block the user path. For example, allow users to look around the interface if possible.

- **Supporting multiple uses** – Users will be performing multiple tasks, so the cognitive and motor load of the interface should be kept at minimum. One possibility is to use automatic behaviors or provide a single operation with multiple uses.

These guidelines can be used to steer the development of new techniques. We applied them to create the following technique ideas:
**MOSE – MULTIPLE OBJECTS SEPARATION**

This technique uses a fluid metaphor to automatically clear the view for physical interactions. All the windows stay in the current position until the user extends the hand to interact with the world (for example, to push a door knob). When the arm extension is detected by the headset, all the windows move apart to clear the view around and below the arm position. When the arm retracts, the windows move back to their original locations. This technique adapts to allow physical interaction, which is something that was only indirectly achieved with our first design.

**SLOW – STABLE LINEAR OFFSET WINDOW**

This technique moves all the windows together in the current FOV. It works by opening a window within the workspace, which can then be moved in any direction. The user grabs any point of the current workspace and moves it to either side. All the windows slide linearly in that direction and stay there until the reverse movement is made.

**WIND – WINDOWS INDIRECT DIVISION**

This technique uses a wind metaphor to automatically blow away the windows placed on the side towards the direction of movement, dividing it into two sections. It is similar to the MOSE technique, but divides the entire vertical field of view automatically based on speed. Once the user stops, the windows come back to the original layout.

### 4.12 LIMITATIONS

There are a few limitations that must be considered when interpreting the results of this research.

**SCENARIO DESIGN**

Our scenario in both contextual studies was mostly designed to assess performance while walking. As such, most participants adopted the strategy to prioritize fast access to information instead of visibility or comfort. For this reason, care should be taken to generalize the results to scenarios where the information access is sporadic. In addition, the scenarios required very little real world interaction. In the first study, participants were only asked to mark an X on the board. In the second they were also asked to sit down in the middle of the path. Those were not frequent enough to change the strategy
participants would use if they were required to interact more constantly with the environment. Some AR uses might depend on that (e.g., assembly tasks). We also forced participants to keep all six windows with them, even when not being used. Even though participants could move them away, it may have forced some of them into non-optimal positions. Allowing the “focus” window to be at the best spot may lead to different results.

**Device Limitations**

The device we used had a small FOV, contrast, resolution, and still had considerable weight. A larger FOV might change the way users arrange the windows before the task, so they would not block so much of the central field of view. It could also allow users to find the windows by using peripheral vision instead of rotating the head, which can be more comfortable and efficient. The display quality, although very good, was still limited in contrast and resolution. A better display may enable users to have the windows farther away than what we observed. A few users also complained about color separation while moving, which is specific to the device we used. Finally, the headset still represents a considerable weight, which may have discouraged participants from naturally looking around. Although we tried to ensure the headset was comfortably and securely attached, some participants were not confident wearing it.

**Content and Interaction**

Another limitation of our study concerns the content. We tried to cover a wide variety of content, including images, diagrams, different typography, and different backgrounds. However, an application requiring only one type of content might perform significantly differently than the average case we evaluated. In addition, more complex applications will likely involve some degree of interaction, either through gestures, voice, or controllers. This could potentially create new cognitive and motor demands that would likely change the way users behave. Depending on the interaction modality selected, users may need to stop for motor reasons as well.
4.13 CONCLUSION

Our research showed that adaptation can be a valuable approach in the design of mobile AR workspaces. Based on a careful study of the design space, we designed and implemented several behaviors that demonstrate different applications of the concept. We also selected a minimum set of behaviors capable of supporting user activities as they move around using mobile AR. Through contextual studies we evaluated those behaviors in life-like scenarios and discovered that they are easy to understand and use. In addition, we also learned particular strategies that users apply while walking and performing tasks in AR. Finally, the set of interaction patterns and guidelines we derived can be used to direct the development of future interfaces.
5 CONCLUSION AND FUTURE WORK

In this dissertation, we have explored different aspects of walking in mixed reality. Through controlled experiments and design based research, we investigated how the ability to walk affects different tasks that users would like to perform in VR and AR. We began by discussing the results of a study comparing walking vs. manipulation (Chapter 2), then we presented a technique for marking in augmented reality (Chapter 3), followed by a study of augmented reality workspaces for walking (Chapter 4).

5.1 SUMMARY OF THE CONTRIBUTIONS

Here we summarize the primary contributions of our three studies:

TRADEOFFS BETWEEN WALKING AND MANIPULATION

We compared the performance of participants during a cognitively demanding visual analysis task while walking vs manipulating a 3D dataset. We found that the performance of the participants was driven by individual differences in spatial ability and game experience. Participants with high spatial ability had similar performance in both conditions. However, participants with low spatial ability and low game experience performed better when walking, while participants with low spatial ability and high game experience performed better rotating the dataset with a controller. We argued that optimal design decisions for this tradeoff should consider the role of training and individual differences.

MODEL-FREE MARKING IN AUGMENTED REALITY

We evaluated different augmented reality techniques for marking points without any model of the environment. We found that geometric triangulation can be more accurate than perceptual methods. However, its precision may be affected by head pointing tremors. To increase the precision, we designed a new technique that takes multiple samples to obtain a better estimate of the target position. The effectiveness of this
approach was demonstrated with a controlled AR simulation and with actual outdoor AR tests.

**MANAGING WORKSPACES IN AUGMENTED REALITY**

We developed and evaluated a system that demonstrates how adaptation can be used to create mobile AR workspaces. We created a taxonomy to describe the design space of adaptive interfaces, and based on this taxonomy, we designed and implemented an adaptive workspace manager. Our evaluation of this system, by means of contextual studies, indicated that the design represents a minimum viable set of operations for this scenario. We also identified a set of key interaction patterns and desirable properties of adaptation-based techniques, which were used to create a set of guidelines to help the design of future walking-centered workspaces.

5.2 **FUTURE WORK**

This dissertation can be extended in many directions.

**PERCEPTION AND COGNITIVE ISSUES**

In our study on dual-task performance (Chapter 2) we used a carefully designed VR environment. It was completely dark and there were few visual cues that participants could use to calibrate their position in space. Most users used only the dataset itself. The virtual environment was also free of distractors. Although these conditions were appropriate for a first study, real environments in VR and AR will probably be different. Future studies could investigate more realistic environmental conditions (for example, a virtual room) or try to replicate our results in AR with wide field of view.

Other cognitive aspects could also be further investigated. Since cognitive abilities are interconnected, it is possible that other factors may have a higher influence in the task performance. For example, we suspect that visual short memory capacity will affect the performance in one or more conditions. If that is so, we may be able to obtain a more accurate model for user performance by accounting for this factor. In addition, we did not block verbal memory in our study. Although we do not have any evidence that it played a role (losing count was not common, but miscounting the features was), a similar experiment blocking verbal memory would help to isolate the results to visual factors.
We could also increase the walking challenge by modifying the path around the dataset to make it narrower or adding no-stepping zones. This would keep the experiment controlled, but require more attention to be directed to walking. It would make the study more similar to walking in real environments. In addition, the attention mechanism itself could also be investigated, as prior work has shown its relevance to walking tasks.

Another interesting aspect to study would be whether some benefits of walking and spatial interaction hold with the mix of real and virtual objects. For example, investigating spatial memory with virtual objects in real environments or using virtual constructions to support memorization in real environments. It is possible that some of the mechanisms responsible for the observed walking benefits are affected by AR fidelity components such as FOV, occlusion, or tracking stability. Finally, we did not look into many interface parameters that could affect the perception of the environment and content in AR (color, contrast, composition, etc.).

**AR Content Creation Techniques**

Our research on model-free point marking (Chapter 3) only scraped the surface of the design space for content creation in wide-area AR. We looked into the most fundamental primitive, a point in 3D space and the VectorCloud technique proved robust to reduce point errors and viable to use while walking. However, it could still be improved in several ways. First, our method for sampling and selecting the rays for intersection is very simple. When walking, the viewing direction does not need to be sampled at uniform intervals. Current head orientation could be used to only sample the views which are pointing in a different direction. Sampling could also be filtered by several factors, such as angular distance from the previous rays (to reject outliers) or the position in the gait cycle (to pick stable head points). Once stored, the rays can undergo a more refined selection. Currently, the VectorCloud splits the ray buffer into two parts and intersect all rays. While simple, it is possible that progressively intersecting the rays from the extremes until the variance stabilize would be better. Or simply use a rejection window for the middle rays.

Another limitation of our current technique is the lack of feedback of marked position. In the experiment, the sampling interval was fixed and in the qualitative studies the
participants were not able to judge how many samples to collect. For this reason, one major feature of the technique (the ability to collect an arbitrary number of samples) cannot be used well. The VectorCloud could be modified to provide a real-time boundary of the point being marked, allowing the user to determine if more samples or a change in position are needed.

We chose the task of marking a point because it is enough for simple annotations and can also be used to build more complex primitives. However, this does not mean that aggregating points into higher level primitives would lead to efficient and intuitive techniques. Future research could look directly into the design of more complex markings such as curves, areas, or volumes. For example, instead of specifying three points to build a plane, a new technique could specify just one point and a vector.

**AR Workspaces**

After the two design iterations, we reached an interesting set of techniques which were simple to use and efficient (Chapter 4). However, it still required the use two buttons to alternate between the two modes. One clear next step would be to investigate automatic switching between those modes. Switching could be driven by any of the elements in our adaptation taxonomy. It could also employ more sophisticated user models that consider spatial history: if the user always switches to Mode A in the corridor and Mode B in the office, it could be done automatically. However, this strategy opens up the possibility of frustrating the user by stealing control and performing the wrong adaptation. A better approach would be to simplify the interaction by eliminating options that cannot work (e.g. attraction without a wall). Designers need to be sure, though, that it will not impede multiple uses.

In our study, we only considered position and orientation adaptation. Perhaps the more promising line for future work is to explore content adaptation. The needs for information will change between walking and stationary conditions. In our experiment, for example, the webpage content could be simplified to show only the headlines while walking, since most other content would require more uninterrupted attention. In addition to improving search and reading time, content adaptation can also reduce the screen area occupied by the interface, improving the visibility of the environment. Content adaptation can consist
of many changes, such as position, color, transparency, presentation form, etc. This sort of adaptation can also follow the position adaptation we studied. For example, when windows move into the wall, they expand and show more detail. When they are surrounding the user, they can assume a more compact form.

Finally, we did not spend time on other features that could be useful in a fully featured workspace manager. For example, the ability to assign different behaviors to individual windows. This could create interesting possibilities, for instance, having some high priority information following you in the personal workspace (e.g., a teleconference call window) while low priority items update their position with less frequency on the walls (e.g., a shared meeting board).
REFERENCES


Lorente de Nó, R., 1933. Vestibulo-ocular reflex arc. *Archives of Neurology & Psychiatry*.


factors in computing systems. pp. 2683–2686.


# Appendix A: Tables

<table>
<thead>
<tr>
<th>Code</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Application</td>
<td>The applicability of the idea to a real scenario</td>
</tr>
<tr>
<td>Automatic vs Manual</td>
<td>Tradeoffs of automatic vs manual behaviors</td>
</tr>
<tr>
<td>Button Familiarity</td>
<td>Use or familiarity with the button input</td>
</tr>
<tr>
<td>Collision</td>
<td>Collision of the windows with the walls</td>
</tr>
<tr>
<td>Context</td>
<td>The content/layout inside the windows</td>
</tr>
<tr>
<td>Content</td>
<td>Different use applicability depending on the context/goals</td>
</tr>
<tr>
<td>Distraction</td>
<td>Distraction caused by the system or world</td>
</tr>
<tr>
<td>FOV</td>
<td>Field of View</td>
</tr>
<tr>
<td>Hand Interaction</td>
<td>Use of hand with the system</td>
</tr>
<tr>
<td>Head Rotation</td>
<td>Head rotation</td>
</tr>
<tr>
<td>Layout</td>
<td>Arrangement of the windows regarding each other</td>
</tr>
<tr>
<td>Learning Curve</td>
<td>Perception of performance change with use</td>
</tr>
<tr>
<td>Occlusion</td>
<td>Windows preventing from seeing the world</td>
</tr>
<tr>
<td>Overlap</td>
<td>Windows occluding each other</td>
</tr>
<tr>
<td>Preference</td>
<td>Something/configuration that was preferred</td>
</tr>
<tr>
<td>Reading</td>
<td>Reading text or acquiring information from the windows</td>
</tr>
<tr>
<td>Strategy</td>
<td>Specific way to approach the task / use the windows</td>
</tr>
<tr>
<td>Suggestion</td>
<td>Suggestion to improve / alter the system</td>
</tr>
<tr>
<td>Technical Issues</td>
<td>Limitations of the prototype or technical problems</td>
</tr>
<tr>
<td>UX</td>
<td>Overall experience with the task / system</td>
</tr>
<tr>
<td>Visual Aspects</td>
<td>Quality of the rendering / display</td>
</tr>
<tr>
<td>Walking Impact</td>
<td>Something caused by walking</td>
</tr>
<tr>
<td>Window Movement</td>
<td>How windows moved or stayed fixed in space</td>
</tr>
<tr>
<td>Interaction with world</td>
<td>Tasks not directly related to AR</td>
</tr>
</tbody>
</table>

Table 0-1 - Descriptive codes created during the initial analysis
MEMORANDUM

DATE: June 19, 2017

TO: Doug A Bowman, Bireswar Laha, Wallace Santos Lages

FROM: Virginia Tech Institutional Review Board (FWA00000572, expires January 29, 2021)

PROTOCOL TITLE: Validation of MR Simulation for the Effects of Display Fidelity in Analysis of Tracheal Systems of beetles from the Platynus and Pterostichus genera

IRB NUMBER: 13-137

Effective June 19, 2017, the Virginia Tech Institution Review Board (IRB) Chair, David M Moore, 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: April 17, 2017
Protocol Expiration Date: April 16, 2018
Continuing Review Due Date*: April 2, 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.
I. Purpose of this Research Project

This research project is intended to investigate the effects of different properties of virtual reality (VR) display systems and compare strategies for analyzing data in virtual reality (VR) systems. This research will help us to gather quantitative and qualitative insights into, and generalize possible benefits for the analysis of 3D volume datasets at higher levels of display fidelity (visual immersion). 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 Moss Arts Center, in the Perform Studio, Experience Studio or Sandbox. It 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 a virtual reality (VR) headset such as the HTC Vive, which will be adjusted to fit you. Next, you will do some training with tasks similar to the ones in the main part of the experiment. Tasks will involve some physical movement, including looking around virtual objects and manipulating them. The main part of the study consists of two blocks, during which the experimenter will record your answers. There will
be a break between each block. After all tasks are completed, you will be interviewed about your experience.

III. Risks

Using VR technology can produce symptoms of illness 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 the properties of VR systems, so that effective user interfaces can be designed for real-world 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.

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
When the experimental part is concluded, we will randomly select five participants to receive a $40 gift card each. The odds of winning are 1 in 6. The winners will be contacted by e-mail when we complete the study with all participants.

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 subject, or need to report a research-related injury or event, you may contact the VT IRB Chair, Dr. David M. Moore at moored@vt.edu or (540) 231-4991.

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
Block 4

Virtual Reality Background Questionnaire and Spatial Visualization Test

This survey consists of three parts that helps us to understand better the results of the study that you are going to participate:

1. A background questionnaire with a few questions about yourself
2. A cube comparison test, to evaluate spatial ability
3. A surface unfolding test, to evaluate visualization ability

The questionnaire should take 1 minute, the first test has 2 parts of 3 minutes each, and the second test has one part of 3 minutes.

The test contains images that will not display correctly in a smartphone. Please use a PC, laptop or large tablet to do the tests.

Default Question Block

Background Questionnaire

Gender

☐ Male
Female

Age

Do you wear glasses or contact lenses?
- Neither
- Glasses
- Contact Lenses

Are you:
- Right handed
- Left handed
- Ambidextrous

Occupation (if student, indicate graduate or undergraduate)

Major / Area of specialization

Rate your expertise with computers

How often do you use computers:
How often do you play video games (on computers or consoles like Xbox, Nintendo, etc)

Never

How pleasant is your experience watching 3D stereo movies, in theaters or on TV:

Bad (headaches, dizziness, etc)

How many times have you tried virtual or augmented reality?

Never used

Once or twice

3 to 10 times

More than 10 times

Would you like to be invited for other studies involving augmented or virtual reality?

Yes

No

This is the end of the Background Questionnaire.

Please proceed to the Paper Folding Test.

Block 7
Paper Folding Test

In this test you are to imagine the folding and unfolding of pieces of paper. In each problem there are some figures drawn at the top and others drawn at bottom. The figures at the top represent a square piece of paper being folded, and the figures on the bottom have one or two small circles drawn on it to show where the paper has been punched.

Each hole is punched through all the thicknesses of paper at that point. One of the five figures on the bottom shows where the holes will be when the paper is completely unfolded. You are to decide which one of these figures is correct and mark the checkbox below it.

Now try the sample problem below. (In this problem only one hole was punched in the folded paper).

![Diagram of folded paper with options A to E]

The correct answer to the sample problem above is C. The figures below show how the paper was folded and why C is the correct answer.
In these problems all of the folds that are made are shown in the figures at the top, and the paper is not turned or moved in any way except to make the folds shown in the figures. Remember, the answer is the figure that shows the positions of the holes when the paper is completely unfolded.

Some of the problems on this sheet are more difficult than others. If you are unable to do one of the problems, simply skip over it and go on to the next one. Your score on this test will be the number marked correctly minus the number marked incorrectly. Therefore, it will not be to your advantage to guess unless you have some idea which choice is correct. Work as quickly as you can without sacrificing accuracy.

You will have three minutes for each of the two parts of this test. Each part has one page. There will be a timer on the top and bottom of each part. The form will auto-advance when the time is over.

Please proceed to the first part when ready.

**Block 5**

**Part 1 (3 minutes)**

Time Remaining:

*These page timer metrics will not be displayed to the recipient.*

First Click: 0 seconds
08
A B C D E

09
A B C D E

10
A B C D E

Block 9
You may rest a while and proceed to the second part when ready.
As before, you will have 3 minutes to work on the next page.

Block 6
Part Two (3 minutes)

Time Remaining:
These page timer metrics will not be displayed to the recipient.
First Click: 0 seconds
Last Click: 0 seconds
Page Submit: 0 seconds
Click Count: 0 clicks

03 00

01
A B C D E

02
A B C D E

03
A B C D E
These page timer metrics will not be displayed to the recipient.
First Click: 0 seconds
Last Click: 0 seconds
Page Submit: 0 seconds
Click Count: 0 clicks

Block 8

This is the end of the Folding Test.

Rest for a minute then proceed to the Cube Comparison Test.

Block 6

Cube Comparison Test

Instructions
Wooden blocks such as children play with are often cubical with a different letter, number, or symbol on each of the six faces (top, bottom, four sides). Each problem in this test consists of drawings of pairs of cubes or blocks of this kind. Remember, there is a different design, number, or letter on each face of a given cube or block.

Compare the two cubes in each pair below.

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td><img src="image1" alt="Cube 1" /></td>
<td><img src="image2" alt="Cube 2" /></td>
</tr>
<tr>
<td>S</td>
<td>D</td>
</tr>
<tr>
<td><img src="image3" alt="Cube 3" /></td>
<td><img src="image4" alt="Cube 4" /></td>
</tr>
<tr>
<td>S</td>
<td>D</td>
</tr>
</tbody>
</table>

The first pair is marked D because they must be drawings of different cubes. If the left cube is turned so that the A is upright and facing you, the N would be to the left of the A and hidden, not to the right of the A as is shown on the right hand member of the pair. Thus, the drawings must be of different cubes.

The second pair is marked S because they could be drawings of the same cube. That is, if the A is turned on its side the X becomes hidden, the B is now on top, and the C (which was hidden) now appears. Thus the two drawings could be of the same cube.

**Note:** No letters, numbers, or symbols appear on more than one face of a given cube. Except for that, any letter, number, or symbol can be on the hidden faces of a cube.

Work the three examples below.

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td><img src="image5" alt="Cube 5" /></td>
<td><img src="image6" alt="Cube 6" /></td>
</tr>
<tr>
<td>S</td>
<td>D</td>
</tr>
<tr>
<td><img src="image7" alt="Cube 7" /></td>
<td><img src="image8" alt="Cube 8" /></td>
</tr>
<tr>
<td>S</td>
<td>D</td>
</tr>
<tr>
<td><img src="image9" alt="Cube 9" /></td>
<td><img src="image10" alt="Cube 10" /></td>
</tr>
<tr>
<td>S</td>
<td>D</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td>---</td>
<td>---</td>
</tr>
<tr>
<td>X</td>
<td>A</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td>S</td>
<td>D</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td>G</td>
<td>I</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td>S</td>
<td>D</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td>J</td>
<td>K</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td>S</td>
<td>D</td>
</tr>
</tbody>
</table>

The first pair immediately above should be marked D because the X cannot be at the peak of the A on the left hand drawing and at the base of the A on the right hand drawing. The second pair is "different" because P has its side next to G on the left hand cube but its top next to G on the right hand cube. The blocks on the third pair are the same, the J and K are just turned on their side, moving the Circle to the top.

In summary: Mark S if they could be the same. Mark D if they must be different.

Your score on this test will be the number marked correctly minus the number marked incorrectly. Therefore, it will not be to your advantage to guess unless you have some idea which choice is correct. Work as quickly as you can without sacrificing accuracy.

You will have 3 minutes to work on the next page. There will be a timer on the top and bottom of each part.

Continue when ready.

**Block 5**

Remaining Time:

*These page timer metrics will not be displayed to the recipient.*

- First Click: 0 seconds
- Last Click: 0 seconds
- Page Submit: 0 seconds
- Click Count: 0 clicks
These page timer metrics will not be displayed to the recipient.
First Click 5 seconds
Last Click 0 seconds
Page Submit 0 seconds
Click Count 0 clicks

Block 4

Proceed to submit your answers and finish this survey.

If you have not done yet, take a time to review the Informed Consent Document before your appointment in the lab. We hope to see you soon!
Validation of MR Simulation for the Effects of Display Fidelity in Analysis of Tracheal Systems of beetles from the Platynus and Pterostichus genera

Post-Experiment Questionnaire

Please rate your feeling with the display by the end of the experiment:

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
</tr>
</thead>
<tbody>
<tr>
<td>Highly comfortable</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
</tr>
</tbody>
</table>

Highly uncomfortable

What in your opinion, makes counting branches challenging if you would do it in real life?

Were there any issues with the technology, device, or technique that made it even more challenging? Which things do you wish you were able to do?
Regarding the techniques:

With which technique do you think you were more ACCURATE:

- [ ] Walking around
- [ ] Using the wand

Why?

Which technique was EASIER to use?

- [ ] Walking around
- [ ] Using the wand

Why?
Which technique was FASTER to use:

- Walking around
- Using the wand

Which strategy or strategies did you use to avoid miscounting while WALKING?

Which strategy or strategies did you use to avoid miscounting while using the WAND?
How would you compare both techniques?

Regarding COMFORT?

- The wand was clearly more comfortable
- The wand was slightly more comfortable
- Both wand and walking were equally comfortable
- Walking was slightly more comfortable
- Walking was clearly more comfortable

Regarding your CONFIDENCE IN THE ANSWERS?

- I was definitely more confident in my answers with the wand
- I was slightly more confident using the wand
- I was equally confident either using the wand or when walking
- I was slightly more confident when walking
- I was definitely more confident in my answers when walking
Regarding the difficulty to obtain the desired VIEWPOINTS?

- It was definitely easier with the wand
- It was slightly easier using the wand
- It was equally easy either using the wand or when walking
- It was slightly easier when walking
- It was definitely easier when walking

Regarding the difficulty to REMEMBER which parts of the dataset you had looked?

- It was definitely easier with the wand
- It was slightly easier using the wand
- It was equally easy either using the wand or when walking
- It was slightly easier when walking
- It was definitely easier when walking

End of Block

This is the end of the questionnaire
MEMORANDUM

DATE: April 30, 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 April 13, 2018, the Virginia Tech Institution Review Board (IRB) approved the Continuing Review 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, 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.

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.
View management and user interface optimization for wide-area mobile augmented reality

You are invited to participate in a research study with virtual reality (VR) and augmented reality (AR) technologies. Our team is investigating the best ways to interact with AR and VR, and this study compares various methods of interaction with virtual data and information. The study will make use of state-of-the-art VR and AR technologies such as the HTC Vive and Microsoft Hololens.

Participants in the study will come to the Perform Studio, room 159 in the Moss Arts Center. Participants will be asked to complete interaction tasks such as navigation, object manipulation, and object selection in either AR or VR. Participants will be asked for their comments about these innovative interaction techniques. The entire experimental session will take about one hour.

Participants need to be at least 18 years old and have normal vision (glasses and contact lenses are fine).

If you are interested in participating, please contact Lee Lisle at llisle@vt.edu to schedule a session. The project is supervised by Dr. Doug Bowman in Computer Science.

This experiment has been approved, as required, by the Virginia Tech Institutional Review Board.
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
MEMORANDUM

DATE: March 27, 2018
TO: Douglas Andrew Bowman, Wallace Santos Lages
FROM: Virginia Tech Institutional Review Board (FWA00000572, expires January 29, 2021)

PROTOCOL TITLE: Interface management for wide-area mobile augmented reality
IRB NUMBER: 18-176

Effective March 8, 2018, the Virginia Tech Institution Review Board (IRB) approved the New Application 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) 6,7
- Protocol Approval Date: March 8, 2018
- Protocol Expiration Date: March 7, 2019
- Continuing Review Due Date*: February 21, 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.

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.
You are invited to participate in a research study with augmented reality (AR) technologies. Our team is investigating the best ways to work in AR, and with this study we will try to understand how future AR workspaces should be. The study will make use of state-of-the-art AR technologies such as the Microsoft HoloLens.

Participants in the study will come to the Sandbox, room 160 in the Moss Arts Center. Participants will be asked to perform physical actions, such as walking and drawing, and tasks in AR such as reading, and moving windows. Participants will be asked for their comments and preferences about different interaction alternatives using brief, audio recorded interviews. The entire experimental session will take about one hour.

- Participants need to be at least 18 years old and have normal vision (glasses and contact lenses are fine).
- The study also takes place in the Moss Arts Center lobby, when there may be non-participants present. There is a risk of nausea, motion sickness, etc. with augmented reality technologies.
- Your participation is voluntary and confidential. The results of this experiment may be published and used in a dissertation.

If you are interested in participating, please contact Wallace Lages at wlages@vt.edu to schedule a session. The project is supervised by Dr. Doug Bowman in Computer Science.

This experiment has been approved, as required, by the Virginia Tech Institutional Review Board.
Title of Project: **Interface management for wide-area mobile augmented reality**

Investigator(s):  
Wallace Lages  
wilages@vt.edu  
(540-750-9380)  
Name  
E-mail / Phone  

Douglas Bowman  
dbowman@vt.edu  
Name  
E-mail / Phone  

I. Purpose of this Research Project

This research project is intended to evaluate a variety of methods for managing workspaces in augmented reality (AR) systems. This research will help us understand the best ways to interact with windows while the user move in space. The findings we may help to establish design guidelines for future AR user interfaces. 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 Sandbox (Moss Arts Center room 160) and other common areas of the building. It 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 an augmented reality (AR) headset such as the Microsoft HoloLens. Using these device, you will then complete a series of tasks to evaluate different AR interaction alternatives. Tasks will involve physical movements including looking around the environment and walking. It will also include interacting with virtual objects in AR. After each alternative, you will be interviewed about your experience; these interviews will be audio recorded.
III. 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 will warn you, if we think you are unaware of a nearby physical obstacle.

IV. Benefits

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

Your voice will also be recorded during the interview. The audio will be transcribed and then deleted. The transcription will not have any identifying information assigned to it.

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
Interface management for wide-area mobile augmented reality

BACKGROUND QUESTIONNAIRE (questions on the online questionnaire)

Q1 Gender
   ○ Male
   ○ Female

Q2 Age
   ________________________________

Q3 Do you wear glasses or contact lenses?
   ○ Neither
   ○ Glasses
   ○ Contact Lenses

Q4 Are you:
   ○ Right handed
   ○ Left handed
   ○ Ambidextrous
Q5 Occupation (if student, indicate graduate or undergraduate)

________________________________________________________________

Q6 Major / Area of specialization

________________________________________________________________

Q8 Rate your expertise with computers

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
</tr>
</thead>
<tbody>
<tr>
<td>Beginner</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>o</td>
</tr>
<tr>
<td>Advanced</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Q9 How often do you use computers:

<table>
<thead>
<tr>
<th></th>
<th>Not at all</th>
<th>Once a month</th>
<th>Once a week</th>
<th>Several times per week</th>
<th>Daily</th>
</tr>
</thead>
<tbody>
<tr>
<td>for work</td>
<td>o</td>
<td></td>
<td></td>
<td></td>
<td>o</td>
</tr>
<tr>
<td>for fun</td>
<td>o</td>
<td></td>
<td></td>
<td></td>
<td>o</td>
</tr>
</tbody>
</table>

Q10 How often do you play video games (on computers or consoles like Xbox, Nintendo, etc)

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
</tr>
</thead>
<tbody>
<tr>
<td>Never</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>o</td>
</tr>
<tr>
<td>Everyday</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Q11 How pleasant is your experience watching 3D stereo movies, in theaters or on TV:

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>Excellent</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Bad (headaches, dizziness, etc)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Q12 How many times have you tried virtual or augmented reality?

- Never used
- Once or twice
- 3 to 10 times
- More than 10 times
Interface management for wide-area mobile augmented reality

I. Interview Prompts

How did you arrange your windows before we go?
Which Behaviors did you use now?
Why did you use/not use the controller to activate behaviors?
Did you notice anything that made the task difficult?
Did you feel the windows prevented you from perceiving the environment?
Did you think that the windows went to positions that made them hard to read?
What about the freeze button?
What do you think went well?
Did you feel you need to stop walking to read or adjust the windows?
What you would like to different in the next time?
Which button would be most useful, A or B? Why?
What is the proportion of each mode you used?
Ideally how would like the windows to behave in this scenario?
How easy was it to read each window?
Do you have any suggestion for how to control them in this scenario?
What do you think was the best thing about the experience? and Worst?
Do you think the scenario we used could represent a real situation?