Reconciling Pixels and Percept

Improving Spatial Visual Fidelity with a Fishbowl Virtual Reality Display

by

Qian Zhou

B.Sc., Tianjin University, 2011
M.Sc., Georgia Institute of Technology, 2014
M.Sc., Shanghai Jiao Tong University, 2014

A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF
THE REQUIREMENTS FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

in

The Faculty of Graduate and Postdoctoral Studies

(Electrical and Computer Engineering)

THE UNIVERSITY OF BRITISH COLUMBIA

(Vancouver)

June 2020

The following individuals certify that they have read, and recommend to the Faculty of Graduate and Postdoctoral Studies for acceptance, the dissertation entitled:

Reconciling Pixels and Percept: Improving Spatial Visual Fidelity with a Fishbowl Virtual Reality Display

submitted by **Qian Zhou** in partial fulfillment of the requirements for the degree of **Doctor of Philosophy** in **Electrical and Computer Engineering**.

Examining Committee:

Sidney Fels, Electrical and Computer Engineering
Supervisor

Robert Rohling, Electrical and Computer Engineering
Supervisory Committee Member

Alan Kingstone, Psychology
University Examiner

Boris Stoeber, Mechanical Engineering
University Examiner

Additional Supervisory Committee Members:

Septimiu E. Salcudean, Electrical and Computer Engineering
Supervisory Committee Member

Panos Nasiopoulos, Electrical and Computer Engineering
Supervisory Committee Member

Abstract

Virtual Reality (VR) has fundamentally changed how we can perceive three-dimensional (3D) objects in a virtual world by providing pictorial representations as 3D digital percepts rather than traditional 2D digital percepts. However, the way we perceive virtual objects is fundamentally different from the way we perceive real objects that surround us every day. Therefore, there exists a perceptual gap between the virtual and real world. The research described in this dissertation is driven by a desire to provide consistent perception between the two worlds.

Bridging the perceptual gap between virtual and physical world is challenging because it requires both understanding technical problems such as modeling, rendering, calibration and sensing, but also understanding how human perceive 3D space. We focus on a Fishbowl VR display to investigate the perceptual gap by introducing new techniques and conducting empirical studies to improve the visual fidelity of digital 3D displays.

To create a seamless high-resolution spherical display, we create an automatic calibration approach to eliminate artifacts and blend multiple projections with sub-millimeter accuracy for a multiple-projector spherical display. We also perform an end-to-end error analysis of the 3D visualization, which provides guidelines and requirements for system components.

To understand human perception with the Fishbowl VR display, we conduct a user experiment ($N = 16$) to compare spatial perception on the Fishbowl VR display with a traditional flat VR display. Results show the spherical screen provides better depth and size perception in a way closer to the real world. As the virtual objects are depicted by pixels on 2D screens, a perceptual duality exists between the on-screen imagery and the 3D percept which potentially impairs perceptual consistency. We conduct two studies ($N = 29$) and show the influence of the on-screen imagery causing perceptual bias in size perception. We show that adding stereopsis and using weak perspective projection can alleviate perceptual bias. The explorations from this dissertation lay the groundwork for reconciling pixels with percept and pave the way for future studies, interactions and applications.

Lay Summary

Virtual Reality (VR) fundamentally changes our ways of viewing and interacting in a computer-generated 3D world. The underlying motivation of VR is to present digital content as if it exists in the real world, thus leveraging humans' built-in capabilities practiced every day. While users perceive a 3D object, they are actually looking at digital pixels on a 2D display. Unfortunately, there are still significant challenges in using pixels to generate realistic visualization, which is preventing this technology from being used in various applications such as training and product design. In this work, we addressed challenges in providing consistent perception from both technical and perceptual aspects, including: creating an accurate display calibration method, characterizing the visual error, demonstrating the influence of the screen shape and the effect of on-screen pixels. By overcoming the major obstacles in providing consistent visual experience, we will make VR more accessible and practical for future research and applications.

Preface

Much of this dissertation have been published elsewhere. A complete list of all publications can be found in Appendix A. The co-authorship of contributions is discussed here. All of the research work in this dissertation was conducted in the Human Communication Technologies Laboratory at the University of British Columbia, Point Grey campus. All user studies and associated analysis were approved by the University of British Columbia Behavioural Research Ethics Board with the original certificate (H08-03005) and post approval activities (H08-03005-A017 and H08-03005-A019).

Chapter 3 and Appendix B have been published in [154] as listed below. I wrote the source code for the calibration, performed the evaluation, and wrote the text. Dr. Miller assisted with analysis and writing. Dr. Fels provided editorial feedback on the manuscript. Mr. Wu and Ms. Correa assisted with the implementation of the demo shown in [154] as well as the supplementary video.

> [154] **Qian Zhou**, Gregor Miller, Kai Wu, Daniela Correa, and Sidney Fels. Automatic calibration of a multiple-projector spherical fish tank VR display. In *2017 IEEE Winter Conference on Applications of Computer Vision (WACV)*, pages 1072–1081. IEEE, 2017. →video

Versions of Chapter 4 have been published in [155] and [37]. The error analysis has been published in [155]. I formulated the visual error and performed the simulation and analysis in consultation with Dr. Fels. I wrote the manuscript and the source code of the error model. Mr. Wu assisted with the implementation of the demo shown in [155] as well as the supplementary video. The design and implementation of the spherical prototypes described in Section 4.1 has been published in [37]. Mr. Fafard, Mr. Chamberlain, Mr. Hagemann and I jointly developed the software and hardware of the system. Mr. Fafard wrote the manuscripts. I assisted with the data collection, analysis and writing. Dr. Miller, Dr. Stavness and Dr. Fels provided editorial feedback on the manuscript.

> [155] **Qian Zhou**, Gregor Miller, Kai Wu, Ian Stavness, and Sidney Fels. Analysis and practical minimization of registration error in a

spherical fish tank virtual reality system. In *Asian Conference on Computer Vision (ACCV)*, pages 519–534. Springer, 2016. →video

[37] Dylan Fafard, **Qian Zhou**, Chris Chamberlain, Georg Hagemann, Sidney Fels, and Ian Stavness. Design and implementation of a multi-person fish tank virtual reality display. In *Proceedings of the 24th ACM Symposium on Virtual Reality Software and Technology (VRST)*, page 1-9. ACM, 2018. →video

A version of Chapter 5 has been published in [152]. I was responsible for the experimental design, data collection and analysis, as well as manuscript composition. Mr. Hagemann assisted with the implementation of the planar display prototype used in the experiment. Dr. Fels and Mr. Fafard were involved in the discussion of the experimental design. Dr. Stavness and Dr. Fels provided editorial feedback on the manuscript.

[152] **Qian Zhou**, Georg Hagemann, Dylan Fafard, Ian Stavness, and Sidney Fels. An evaluation of depth and size perception on a spherical fish tank virtual reality display. *IEEE Transactions on Visualization and Computer Graphics (TVCG)*, 25(5): 2040–2049, 2019. →video

An early version of Chapter 6 has been published in [156]. I was responsible for the experimental design, data collection and analysis, as well as manuscript composition. Dr. Fels and Ms. Wu were involved in the discussion of the experimental design. Dr. Stavness and Dr. Fels provided editorial feedback on the manuscript.

[156] **Qian Zhou**, Fan Wu, Sidney Fels, and Ian Stavness. Closer object looks smaller: Investigating the duality of size perception in a spherical fish tank vr display. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, pages 1–9. ACM, 2020.

Early versions of Section 3.2 and Section 4.1 have been published as interactive demonstrations in [157] and [153]. I developed and tested the applications jointly with Mr. Fafard, Mr. Wagemakers, Mr. Chamberlain, Mr. Wu and Mr. Hagemann. Mr. Fafard, Mr. Wagemakers and Mr. Chamberlain developed the rendering infrastructure required to run the application. I wrote the manuscripts with the assistance of Dr. Miller, Dr. Stavness and Dr. Fels. Dr. Stavness and Dr. Fels were involved in the discussion of the design of applications and provided editorial feedback on the manuscript.

[153] **Qian Zhou**, Georg Hagemann, Sidney Fels, Dylan Fafard, Andrew Wagemakers, Chris Chamberlain, and Ian Stavness. Coglobe: a co-located multi-person ftvr experience. In *ACM SIGGRAPH 2018 Emerging Technologies*, page 5. ACM, 2018.

[157] **Qian Zhou**, Kai Wu, Gregor Miller, Ian Stavness, and Sidney Fels. 3dps: An auto-calibrated three-dimensional perspective-corrected spherical display. In *Virtual Reality (VR)*, 2017 IEEE, pages 455–456. IEEE, 2017.

Table of Contents

Appendices

List of Tables

List of Figures

Glossary

1D One Dimensional

2D Two Dimensional

3D Three Dimensional

ANCOVA Analysis of Covariance

ANOVA Analysis of Variance

AR Augmented Reality

CAD Computer-Aided Design

CI Confidence Interval

DOF Degree of Freedom

FTVR Fish Tank Virtual Reality

HMD Head-mounted Display

MR Mixed Reality

RE Real Environment

RMS Root Mean Square

SAR Spatial Augmented Reality

VE Virtual Environment

VR Virtual Reality

Acknowledgements

Much of this work has been supported by the University of British Columbia, Natural Sciences and Engineering Research Council of Canada (NSERC) and B-Con Engineering.

This work would not have been possible without the invaluable guidance and supervision from my research supervisor Professor Sidney Fels, who has been my role model in my development as a researcher. Thanks for providing an excellent environment for research and supporting me with perceptive insight and enthusiasm of research.

I am grateful to my committee members and examiners: Robert Rohling, Tim Salcudean, Panos Nasiopoulos, Alan Kingstone, Boris Stoeber and Kevin Ponto (my external examiner). Their expertise in various aspects has been very helpful to refine this work. Special thanks go to Kellogg Booth for taking the time to review the dissertation and providing constructive feedback through in-depth conversations.

I would like to thank Gregor Miller and Ian Stavness for their guidance and continuous support for the work in Chapter 3 and 4, as well as Chapter 5 and 6, respectively. I would also like to thank my fellow colleagues at the HCT lab for their support in user studies and special thanks go to Fan Wu and Georg Hagemann for the friendship and support through the tough times.

Lastly, I would like to thank my parents and my husband Weipu for the endless supplies of encouragement and generous support. Thanks for putting up with me through both the good and bad times. Because of their unconditional love and patience, I have got the opportunity to complete this dissertation.

Chapter 1

Introduction

Virtual Reality (VR) has fundamentally changed the way we perceive 3D objects in a virtual world by providing pictorial representations as 3D digital percepts rather than traditional 2D digital percepts. The underlying motivation of VR is to present virtual 3D content as if it exists in the real world, thus leveraging humans' built-in capabilities practiced and well-developed in the 3D world that surrounds us every day. One of the most important factors to support such experience is the *visual fidelity*, defined as *"the degree to which scenes presented on a display system are perceived in the same manner as an equivalent real-world scene"* [128]. High level of *visual fidelity* provides users consistent perception between the real and digital world, which is crucial in VR.

For 3D displays, providing high *visual fidelity* is non-trivial, and often involves technical challenges in rendering, calibration and tracking, as well as perceptual challenges which require a deep understanding of how humans perceive the 3D space. In this dissertation, we investigate *visual fidelity* by examining a spherical Fish Tank Virtual Reality display. Fish Tank Virtual Reality (FTVR) displays [30] enable head-coupled perspective on high-resolution desktop screens. In comparison to other 3D displays, FTVR display is promising to provide consistent perception because it allows users to see virtual contents situated in the real world. Using a spherical FTVR display, we propose approaches to improve *visual fidelity* from both technical and perceptual standpoints towards bridging the perceptual gap between real and virtual world.

1.1 Motivation

The research described in this dissertation is motivated by a desire to visually provide perceptual consistency between the virtual and real world by identifying technical challenges and understanding the way humans perceive spatial structures in the real and virtual environment. Providing consistent perception enables a user to utilize real-world skills in a virtual environment.

It also helps to transfer learning from the virtual environment back into the real world [129].

It is important to provide high *visual fidelity* because our cognition and actions depend on what we perceive. Every canonical manipulation task in 3D involves different aspects of spatial perception. A selection task, for example, is the task of selecting a target object from the entire set of objects available. The real world counterpart of the selection task is picking and grabbing an object with a hand. It requires accurate depth perception so that the user knows the position of the target object. As another example, a scaling task is the task of resizing an object until the size of the object meets certain criterion. It requires precise size perception knowing the scale of the object. Providing high fidelity of depth and size perception is important since users can leverage their interaction skills in the real world when displays effectively provide perception in a consistent way as in the real world.

Despite of their abstract forms, these canonical tasks are the building blocks of many 3D applications. For example, Computer Aided Design (CAD) systems have an assembly task which involves selecting objects, placing them in the desired locations, and resizing them along various dimensions [145]. Misperception of the distance and size of virtual models would decrease the performance of these interactions. Providing accurate spatial perception is more than supporting 3D interactions in CAD. Most importantly, it helps preserving consistent shape perception between the design and fabricated object, such that when a virtual object is fabricated in the real world, the user is confident that the fabricated object matches the look of the design. Medical visualization is another example that requires perceptual consistency for various tasks ranging from diagnose to surgical training and planning. Accurate depth and size perception plays a critical role in identifying anatomic landmarks, performing resection of tissue, and closing the incision [26]. Effective surgical training allows trainees to progress partway along the learning curve before performing on a real patient. To ensure the learning from the virtual environment can be transferred back into the real world, it is important to provide spatial perception consistently between the training and real environment, which require high *visual fidelity* supported by the display.

Driven by these use cases, our work seeks to recognize technical and perceptual challenges to provide perceptual consistency in the virtual environment. A virtual entity goes through a sequence of processes before it is ultimately perceived by the viewer. Problems may arise from image creation to percept formation, making the virtual entity deviated from the way it was intended to look like. There are steps that need to be taken to ensure the

perceptual consistency. Our work is to identify and understand these steps to improve *visual fidelity* in the virtual environment.

1.2 Fishbowl VR Display

Recent advances in display technologies make it possible to present virtual content with high resolutions, fast refresh rates, and stereoscopic capabilities. Depicting the virtual scene with high definition pixels helps to build up the perceptual vividness of 3D space. However, it is still challenging to provide the same visualization as in the real world. While we perceive a virtual object, we are actually looking at pixels on a 2D screen. This is fundamentally different from the way we perceive a real object, because the light is not emitted from true 3D space, but rather from 2D pixels. Solutions are available to date, in the form of volumetric displays, which produce volume-filling 3D imagery so that each voxel emits visible light from the region where it appears [40]. However, these displays often suffer from limited resolution and low brightness with a small viewing angle, which spoils the 3D experience and fails to meet the promise of high *visual fidelity* . Therefore, pixel-based 3D displays are still the mainstream of VR in providing high quality viewing experience.

To bridge the perceptual gap between the virtual and real world, there are technical issues and perception issues. From the technical standpoint, display systems have registration error, tracker jittering, system latency and display resolution that can cause different artifacts in the virtual environment. These artifacts do not exist in the real world and hence impairs the virtual experience. On the other hand, we perceive various forms of feedback in the real world. The visual information alone is formed by lots of depth cues, such as perspective, occlusion, shadows, motion parallax and stereopsis. In contrast, the virtual environment can only provide a limited collection of them. Furthermore, within this collection, information can conflict with each other. The stereoscopic vergence-focus problem, for example, is a conflict between the disparity and vergence information caused by the 2D screen [142].

Confronted with so many obstacles, how can we improve the *visual fidelity* in the virtual environment? Fortunately, for the visual information, research has found that the way we use spatial information depends greatly on the purpose of perceiving [27, 142]. It stands for the reason that if we can understand and use the spatial information which is significant for critical tasks in an application, users may still have the correct spatial perception

in the absence of some depth cues. Of all the depth cues, stereopsis has been long favored for providing acuity in relative depth judgements [142], positioning objects in a 3D space [19], and improving size-constancy in the virtual environment [91]. In addition to stereopsis, motion parallax has been found to be an independent depth cue for both shape and relative depth perception in the absence of others [119]. It is believed that motion parallax is the one that will most likely enables users to see more information, particularly for certain cognitive tasks [142].

While most 3D displays today provide both stereopsis and motion parallax cues, few work have met the promises for providing perceptual consistency between the real and virtual environment. Immersive 3D displays, such as head-mounted displays (HMDs) and CAVE displays, create deep engagement and strong immersion in the virtual environment. They surround and isolate users from the real world, making it challenging to provide consistent perception between the real and virtual world. Unlike immersive 3D displays, FTVR displays create a 3D illusion that is situated within the real world, allowing easy transition between real-world and virtual information on the FTVR display. See-through HMDs can create similar effect, but presently support a limited field of view with a reduced brightness caused by the additive blending [11].

Previous studies have compared task performances of FTVR displays with other VR forms like CAVE and HMD [31, 41, 106]. The results of these studies show mixed preferences, suggesting that user performance is task related. Users tend to perform better in immersive 3D displays with exploration tasks [41], while in FTVR displays they perform better with tasks that require understanding of spatial structures [31, 106]. Nevertheless, most of these user evaluations have relied on fixed, single screen FTVR, where the user's head motion generates limited motion parallax cues and is generally within a viewing angle of 45 degrees off-axis. Limited head-motion will favor non-head-coupled rendering because the user's perspective does not significantly deviate from a default fixed perspective. Multi-screen FTVR display can create a larger viewing space than a single screen, increasing the head-coupled motion parallax. A previous study with a cubic FTVR has shown potentially better performance with a 3D path-tracing task [126] and mental rotation task [86]. However, the presence of seams between multiple displays discourages users to better take advantage of the multi-screen aspect [86, 126].

Spherical FTVR displays have the advantage to generate seamless motion parallax with the 360 degrees of visibility, creating a virtual "fishbowl" experience using head-tracked rendering. It maintains a metaphor of virtual

Figure 1.1: Fishbowl VR display with multiple rear projectors. The stereoscopic imagery is perspective-corrected based on the viewer's left and right eye positions. The viewer perceived a 3D scene by looking at the projection on the spherical surface.

objects contained within the bounds of the spherical display. This metaphor helps the virtual content to appear more naturally situated within the real world, as if it were real objects within a glass globe or display case. As a display situated in the real world, we believe the spherical FTVR display is promising to provide perceptual consistency between the real and virtual environment. By convention, a spherical fish tank is equivalent to the word "fishbowl". We will use *Fishbowl VR display* in this dissertation to represent spherical Fish Tank VR display.

Our investigation starts with a spherical display using multiple mini-projectors shown in Figure 1.1. Tiling multiple projectors on the spherical screen can increase the display resolution and make the system scalable. The challenge lies in stitching and blending multiple projection to create a seamless display. We present an automatic calibration approach that achieved sub-millimeter accuracy as discussed in Chapter 3. With head-tracking, the calibrated display renders correct perspective based on the viewpoint. How-

ever, it is not clear how accurate the calibration and tracking needs to be in order to create effective visualization. Even small inaccuracies induced by the display or tracker may break the 3D illusion and cause visual artifacts. Therefore, we conducted an end-to-end error analysis of a Fishbowl VR display, which provides detailed descriptions of the criteria for different components in the system in Chapter 4. Together with the calibration approach, we established the technical foundation of a Fishbowl VR display to provide 3D experience with high *visual fidelity* .

As the display system fidelity has been established, we conducted studies to understand the way humans perceive spatial structure using the Fishbowl VR display in Chapter 5. The spherical screen has several unique properties that can potentially improve spatial perception and enhance the 3D experience, such as the enclosing shape, consistent curved surface and borderless views from all angles. It is not clear whether these natural affordances can improve spatial perception in comparison to traditional flat FTVR displays. Therefore, we conducted an experiment to see whether users can perceive the depth and size of virtual objects better on a Fishbowl VR display compared to a flat FTVR display in Chapter 5.

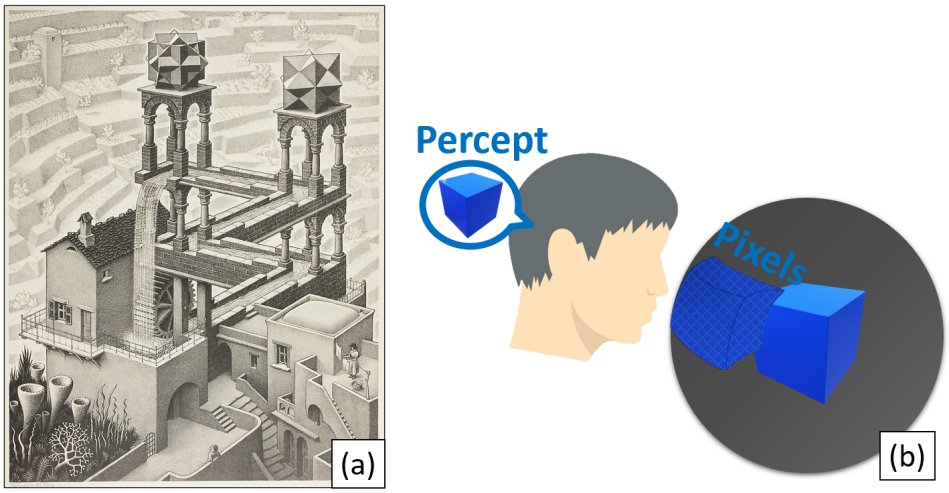


Figure 1.2: Illustration of perceptual duality. (a) M. C. Escher's *Waterfall*. Copyright © The M.C. Escher Company B.V. Baarn. The Netherlands. (b) A user perceives a 3D cube by looking at a 2D image rendered in pixels on the surface of the Fishbowl VR display. The user knows it is a 2D picture but also knows it is a 3D structure. These two concurrent understandings represent the perceptual duality when perceiving 3D out of 2D.

Virtual objects are depicted by pixels on the screen. When we perceive a 3D object, we are actually looking at digital pixels (Figure 1.2(b)), which is fundamentally different from the way we see objects in the real world, when light directly comes from the object, not from pixels with an incorrect depth. As human beings, we have the ability to perceive 3D out of a 2D picture before 3D displays are invented. Escher's *Waterfall* (Figure 1.2(a)), is an example of our ability to perceive 3D structure that can even be physically impossible by looking at a 2D image. We know this is a 2D picture, we also know this is an impossible 3D structure. The duality between these two concurrent understandings may cause perceptual inconsistency when it mismatches with viewer's expectation. We conducted studies to demonstrate this perceptual duality and proposed methods to alleviate the perceptual inconsistency in Chapter 6.

Providing high *visual fidelity* in the virtual environment is difficult and inherently has considerable technical and perceptual challenges. To fully overcome these challenges, years of research are required with the concerted effort from numerous researchers. We expect that our effort of exploiting some unique characteristics of such unusual display can help understand way people perceive objects, and build the perceptual bridge between the virtual and real world.

1.3 Contribution

This dissertation investigates *visual fidelity* with a Fishbowl VR display from the technical aspect and the human factor aspect. We developed technical approaches to create a multi-projector Fishbowl VR display (Chapter 3 and 4). The contribution of Chapter 3 is based on computer vision techniques and Chapter 4 based on computer graphic techniques. We conducted studies to understand the way humans perceive spatial structures using the Fishbowl VR display (Chapter 5 and 6). Publications resulting from this work are listed in Appendix A. The four main contributions are summarized here.

Automatic calibration of a multiple-projector spherical display

i. **Created a novel automatic calibration approach of a multiple-projector spherical display.** We developed an automatic calibration method to blend multiple projections in creating a seamless display for a multiple-projector spherical display. Using the correspondence between the projected pattern and the observed pattern from a

single camera, we reconstruct the 3D position of each projected pixel on the display.

ii. **Applied and evaluated the calibration approach to a Fishbowl VR prototype.** We applied and evaluated the calibration approach with a Fishbowl VR prototype. The results achieved sub-millimeter calibration accuracy.

Error analysis of a Fishbowl VR display

i. **Formulated visual error of a Fishbowl VR display.** We formulated the visual error of a Fishbowl VR display in terms of display calibration error and head-tracking error. Using this model, we analyzed the sensitivity of the user's visual error to each error source in the system.

ii. **Established design guidelines for FTVR displays.** Using this model, we provided guidelines and requirements for the tracking and display system to minimize visualization error for FTVR displays.

Evaluation of spatial perception in a Fishbowl VR display

i. **Compared spatial perception in a Fishbowl VR display with a traditional flat FTVR display.** We conducted an experiment and demonstrated that users perceived the depth and size of virtual objects better on a Fishbowl VR display compared to a traditional flat FTVR display.

ii. **Demonstrated superiority of the spherical display in providing better size-constancy.** We showed that the perception of size-constancy is stronger on the Fishbowl VR display than the flat FTVR display.

Evaluation of perceptual duality in FTVR displays

i. **Demonstrated the influence of on-screen imagery on size perception.** We conducted two studies and found that the size of on-screen imagery significantly influenced object size perception, indicating there is a perceptual duality between the on-screen pixels and the 3D percept.

ii. **Demonstrated the influence of stereopsis on size perception.** We conducted a study and found adding stereopsis can mitigated the perceptual bias of size perception caused by the on-screen imagery.

iii. **Compared size perception under different projection matrices.** We conducted a study and found weak perspective projection significantly reduced the perceptual bias of size perception and was strongly preferred by users compared to perspective projection.

1.4 Outline

This dissertation is structured around the four main research contributions. Chapter 2 is an overview of literature related to the topics of 3D displays and perception studies in the virtual environment. Chapter 3 describes the automatic calibration approach to blend multiple projections on the spherical display. Chapter 4 details the design and implementation of the Fishbowl VR prototypes with visual error model, simulation results, and guidelines for FTVR displays. Chapter 5 describes the methodology and experiment in evaluating depth and size perception on the Fishbowl VR display. Chapter 6 describes the methodology and experiments in evaluating the influence of the on-screen imagery on perceived object size. Chapter 7 summarizes the dissertation contributions, describes directions for future work, and provides concluding remarks. The appendices provide additional background material. Appendix A lists the publications, research talks and demonstrations associated with the dissertation. Appendix B describes the optimization formulation of the calibration approach discussed in Chapter 3. Appendix C provides the subjective questionnaires of user experiments described in Chapter 5 and 6. Appendix D provides the tables of results from the statistic analysis in Chapter 5 and 6.

Chapter 2

Related Work

Numerous effort has been made in researching and developing 3D display technologies. By providing additional depth information, 3D displays introduce the idea of reproducing the visual experience in the digital world as we have in the real world. Recently, the rapid advances in computer graphics and display technologies boost the development of 3D displays, making various 3D displays such as Oculus Rift [54] and Hololens [61] accessible and affordable. These emerging high quality implementations have made it possible to deliver high level of visual experience in the virtual environment.

In this chapter, we review the visual display technologies and perception studies that are crucial to provide high *visual fidelity* in the virtual environment. The first part of the chapter surveys existing 3D displays to provide an overview of the strengths and weakness of each display technology in supporting perceptual consistency between the virtual and real world. As providing high fidelity visual experience requires understanding of human visual perception, the second part reviews existing effort in evaluating spatial perception in the virtual environment focusing on depth and size perception. The last part reviews technical approaches such as calibration and tracking techniques to support high fidelity visual experience on 3D displays.

2.1 3D Visual Displays

Display devices present information to one or more modalities of human perceptual system with most of the displays focusing on stimulating the visual, auditory or haptics [17]. In this dissertation, we focus on 3D visual displays which provide additional depth cues such as stereopsis and motion parallax compared to traditional 2D visual displays. Therefore, the term of *displays* represents *visual displays* in this dissertation. Among different types of 3D visual displays, the head-mounted display is the most well-known one. Head-mounted displays (HMDs) can provide a strong sense of immersion and engagement with an "inside-out" 3D experience (looking outwards from inside the 3D volume). As an alternative design to early immersive HMDs, Fish Tank Virtual Reality display was developed to provide

an "outside-in" 3D experience (looking inwards from outside the 3D volume) [106] with affordable and effective technologies for exploring 3D content and tasks. While being non-immersive, it renders virtual imagery situated in a real-world context, which is important for providing consistent perception between the virtual and real environment.

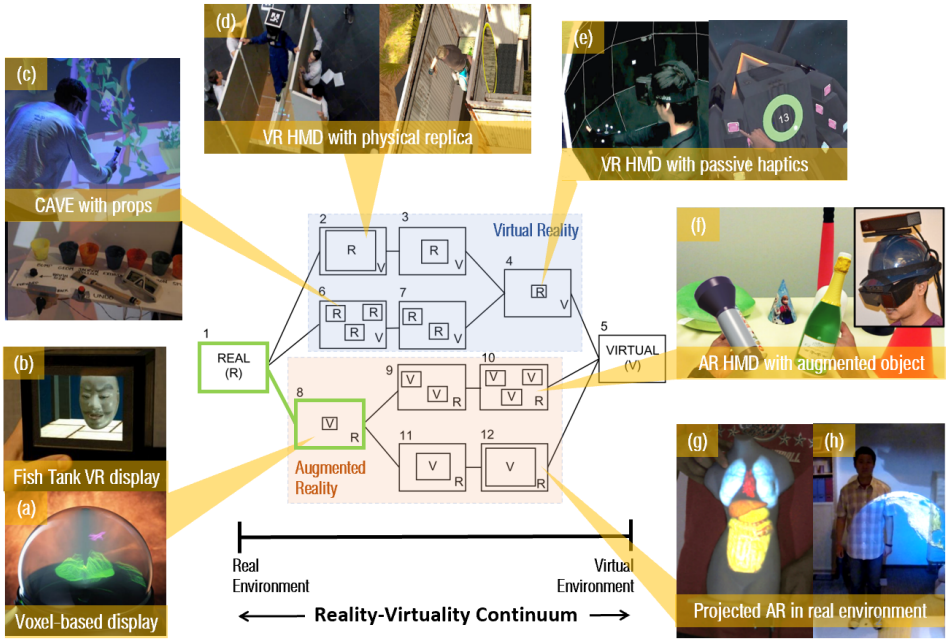


Figure 2.1: Milgram's reality-virtuality schematic continuum [94] with representative 3D displays discussed in Section 2.1. Real environments are shown at the left end of the continuum with virtual environments at the opposite extremum. The schematic continuum shows the variety of ways in which the Real (R) and Virtual (V) components could be mixed. Examples of 3D displays include: (a) FTVR display *pCubee* [126] © ACM (2010). (b) *Perspecta* volumetric display [40] © IEEE (2005). (c) *CAVE Painting* with physical props [72] © ACM (2001). (d) *Turkeck* with physical replica [25] © ACM (2015). (e) *Sparse Haptic Proxy* [24] © ACM (2017). (f) *Annexing Reality* [50] © ACM (2016). (g) Spatial AR display *pMomo* [158] © ACM (2016). (h) *Dyadic AR* display [11] © ACM (2015). All figures adapted with permission.

Despite of the implication from its name, perhaps Fish Tank Virtual Re-

ality (FTVR) display is better classified as an Augmented Reality display. In Virtual Reality (VR), computer-generated graphics fully replace the real world, while in Augmented Reality (AR), the virtual graphics is blended and situated in the real world. We included Milgram's taxonomy of Mixed Reality (MR) displays [94] to better understand the relation between different 3D displays and the situated environments.

As shown in Figure 2.1, the "reality-virtuality continuum" is an abstraction which can be visualized as a line connecting the virtual and real environment, where the real environment is shown at the left end of the continuum, and virtual environment at the opposite extremum. In this continuum, both AR and VR displays are subsets of MR displays, with AR displays considered to be closer to the real environment. Milgram further expands the one-dimension continuum to a schematic collection of composites represented as blocks between the virtual and real environment. As shown in Figure 2.1, the same left-to-right continuum demonstrates the variety of ways in which the Real (R) and Virtual (V) components could be mixed, ranging from fully real to fully virtual environments with intermediate points on the continuum being mixed environments. The upper half of the scheme represents approaches creating real components in the virtual environment with the bottom half as creating virtual components in the real environment.

To better understand the role of FTVR displays, existing representative 3D display systems are enumerated and placed in the blocks of this scheme. Detailed discussions of these approaches are provided later in this section. VR HMDs such as Oculus Rift [54] and surround-screen projector-based VR system (CAVE) [29] provide immersive 3D experience fully replacing the real world. They are examples of Block 5. Some HMDs and CAVE systems [72, 124] use real-world objects to provide real haptic sensation in the virtual environment. They are examples of Block 4, 6 and 7. Other immersive HMDs with room-scale physical replica such as physical walls [24, 25, 90] are examples of Block 2 and 3. AR HMDs such as Hololens [61] superimpose virtual images on top of the real world. Depending on the amount of superimposed entities, they are examples of Block 8, 9 and 10. Spatial Augmented Reality (SAR) displays such as iLamp [114] and Dyadic Projected SAR [11, 158] use projectors to place the augmenting graphics over the physical objects. Depending on the scale of projection coverage, they are examples of Block 11 and 12. Finally, volumetric displays [40] and FTVR displays [143] are desktop 3D displays situated in the real world. Therefore, they are examples of Block 8.

According to the reality-virtuality continuum, we believe blocks closer

to the real environment shown in Figure 2.1 are promising to provide consistent perception as in the real world. The proximity of FTVR displays with respect to the real environment shows promise of providing consistent perception. In this section we provide a background on the types of 3D displays. Each 3D display approach has its unique properties to support various depth cues. By examining and comparing the properties of FTVR displays to other 3D displays, we will gain an understanding of how to provide consistent perception and improve the 3D experience in the virtual environment.

2.1.1 Fish Tank Virtual Reality Displays

Fish Tank Virtual Reality (FTVR) display was originally proposed as a single desktop display [143]. It is designed to be a small, bounded and fish-tank-sized artificial environment which corrects perspective to a tracked viewpoint. The important finding of the original FTVR display was a comparison study of different depth cues, such as motion parallax and stereopsis. While motion parallax and stereopsis together worked the best, motion parallax alone resulted in better performance compared to stereopsis alone [6, 143, 144]. Other research also found that motion parallax enhanced presence [81], improved comfort [88], and decreased rotation error [87].

A number of studies have been conducted to compare users' performance among FTVR, CAVE and HMD. Qi et al. compared the performance of HMD and FTVR using simulated volumetric data [106]. They found that FTVR provided better accuracy at judging the shape, density and connectivity of objects. Demiralp et al. compared CAVE and FTVR in a scientific visualization application and found users performed an abstract visual search task significantly more quickly and accurately on FTVR over CAVE [31]. Using similar experimental setup with a different task, Prabhat et al. found the opposite result. Participants performed significantly better on CAVE and uniformly preferred CAVE [104]. The mixed results suggest that user performance is task related. However, most of these user evaluations have relied on fixed, single screen FTVR displays, where the user's head motion generates limited motion parallax cue and is generally within a viewing angle of 45 degrees off-axis. Limited head-motion will favor non-head-coupled rendering because the user's perspective does not significantly deviate from a default fixed perspective.

An important extension of the FTVR concept used multiple screens to construct multi-sided FTVR displays. The first multi-screen FTVR displays used three flat Liquid Crystal Display panels arranged into convex

corner [71] or three projectors projecting into a concave corner [32]. The advantage of the multi-screen display is that it allows for a larger range of head movement around the screen and therefore enhanced the use of motion parallax cue. The concept was further refined with screens on five sides of a box. Cubee and pcubee provided the illusion of an enclosed volumetric display allowing a viewer to view all sides of the 3D objects "inside" the display [126, 127]. However, it has been reported [85] that the occlusion caused by seams between screens discouraged users from changing their view from one screen to another. This makes the shape of the display an important form factor. Among different shapes of displays, the spherical FTVR display has a promising shape as it has no seams between screens to provide an unobstructed view from all angles. Spherical FTVR displays have advanced recently with improved calibration and rendering techniques [134, 139]. These advances have enhanced FTVR experience in 3D.

2.1.2 Projector-Based 3D Displays

Video projectors have been utilized to build scalable and immersive 3D displays. As a pioneering work, CAVE provided room-scale immersive visualization by projecting content on the six surrounding walls of a cubic empty room [29]. Following CAVE, various work have extended this concept with different display shapes, such as dome [1, 121] and cylinder [122]. As CAVE requires an empty room setup, other work also explored the possibility of projecting view-dependent content onto dynamic surfaces in a room with furniture and objects [68].

In contrast to CAVE which completely surrounds and isolates the user from the real environment, Raskar first introduced the idea of augmenting the appearance of everyday objects with video projections [114, 116, 117]. This approach is called Spatial Augmented Reality (SAR) [13]. It augments and enhances physical objects and spaces in the real world by projecting images onto their visible surfaces. The idea of overlaying virtual content onto real objects is followed by various work. Mine et al. [95] used projection-based augmentation to create spatially augmented 3D objects and spaces that enhance the theme park experience. Zhou et al. [158] extend projection-based augmentation on moveable objects. Benko et al. [9] demonstrated a working implementation of a projected SAR tabletop by accounting for the deformations caused by physical objects on the table. Later they extended this approach to the entire room and provide simultaneous perspective views to two users [11]. Similar to Benko's work, Jones et al. [69] proposed IllumiRoom to create peripheral projected illusions in the living room by

augmenting the area surrounding a television with projected visualizations to enhance the viewing experience.

In summary, projector-based approaches can support both VR and AR experience in a scalable way. The primary challenge for projector-based 3D displays is the multiple-projector calibration to produce a single continuous image across different projections.

2.1.3 Head-Mounted Displays

Head-mounted display (HMD) maximizes field of view by placing a stereo pair of displays directly in front of each eye. Although it is not until recent five years commercial HMD products become widely accessible, one of the most well-known HMD can be traced back to as early as 1968 by Sutherland [131]. Based on the reality-virtuality continuum [94], HMDs can be classified as VR and AR HMDs. VR HMD, such as Oculus [54], provided compelling 3D experience by placing the user in a fully immersive virtual environment. The result is the absence of the user's physical surrounding context and potential co-located collaborators [6], making it challenging to provide consistent perception across the real and virtual world. AR HMDs allow users to see the real world, by mounting cameras on the display, known as video-based see-though [58], or reflecting projected images into user's eyes, known as optical-based see-through [61]. Because AR HMD augments user's view by superimposing virtual objects on the real world, it requires high accuracy on the registration to align the virtual object in the real environment [7], and presently only supports a limited field of view with a reduced brightness caused by the additive blending [11].

A known perceptual issue for HMDs is the accommodation-convergence conflict. The user focuses on a screen close to their eyes, while their eyes converge toward 3D objects located far from the screens. The failure to present focus information correctly, coupled with the convergence, may cause motion sickness and eyestrain [142]. While this problem also exists in other 3D displays such as SAR and FTVR displays, it is believed to be less severe than HMDs as the content and displays both fit certain constraints with a smaller depth interval of the 3D space rendered by the displays [83].

2.1.4 Static and Swept Volumetric Displays

Unlike other 3D displays, static and swept volumetric displays produce volume-filling 3D imagery so that each voxel emits visible light from the region where it appears. Due to this property, static and swept volumetric

displays do not require supplementary hardware of head-tracking to present depth cues such as stereopsis and motion parallax. As 3D objects are rendered in-place via voxels, there is no accommodation-convergence mismatch. These displays became commercially available with relatively high technical requirements since the early 00s [40]. Comprehensive reviews and classification of volumetric displays can be found in [14, 40]. Despite of the capability of illuminating 3D points in their physical spatial locations, volumetric displays are in general limited in viewing angle, resolution and brightnenss with display artifacts compared to other 3D displays [45].

2.1.5 Spherical Displays

A number of interactive spherical display prototypes have been presented in various forms. Perspecta Spatial 3D System from Actuality Systems [40] utilized an embedded stationary projector to project imagery onto a rotating screen. Although it can generate volume-filling imagery, this type of system is expensive to build and has limitations in resolution and scaling. As an alternative method, some systems use rear-projection directly onto a spherical screen. Sphere [10], Snowball [15], and commercial products like Pufferfish [65], use one projector for rear-projection onto the spherical screen. While simple and fairly effective, these systems offer low resolution and lack of scalability. Spheree [133] extends this approach by mounting multiple pico-projectors under the spherical screen to increase the resolution and make the display scalable.

Besides the design of the display system, various work have investigated unique properties of spherical displays with comparisons to traditional flat displays. As one of the early work, Benko et al. [10] identified several unique properties of spherical displays in comparison to flat displays. The spherical shape provides easy 360-degree access for multiple users without occlusions. The continuous nature of the curved shape provides smooth transition both horizontally and vertically. The enclosing nature of the sphere inherently defines the virtual space within the screen. In contrast to flat displays, spherical displays do not have a fixed "sweet spot" of viewing. Following this study, Bolton et al.[16] investigated how spherical displays can support differences in sharing of information during competitive and cooperative tasks. They did not find superiority of spherical displays in competitive and cooperative tasks. Pan et al. [99] presented a tele-presence system using a spherical display. They found the spherical display allows more accurate judgment of the gaze direction in comparison to a planar display. Teubl et al. [133] designed a perspective-corrected spherical display. They integrated the spherical dis-

play with 3D design software. They found the free-viewpoint provides a highly immersive sense for users. Berard et al. [12] presented a hand-held spherical display and investigated its benefits for object examination. They found the object examination task did not benefit from the accurate and precise rotations offered by hand-held spherical displays compared to planar displays. The divergent results from these studies suggest that user performance with spherical displays can be task-dependent. Though considerable work has presented various interaction techniques to leverage the nature of spherical displays, previous literature reveals little conclusive findings on the fundamental spatial perception when using spherical displays.

2.2 Visual Perception Evaluation in the Virtual Environment

Living in a three-dimensional world, we are provided with abundant information everyday which helps us to perceive 3D objects and interact with them. The information we receive is inherently multi-modal in the forms of visual, auditory and haptic feedback. In some cases we also get olfactory feedback. To improve the sensory experience in the virtual environment, all the modalities need to be considered. Among them, the visual information, which stimulates users' visual system, is by far the most common modality studied by researchers to create high fidelity experience in the virtual environment due to its importance in the human sensory system [17]. In this dissertation, we focus on the visual modality to create a consistent perception between the real and digital world.

This section reviews existing literature which evaluate the spatial perception in the virtual environment. We first review definitions of fidelity in the virtual environment from previous literature to help understand and distinguish the *visual fidelity* discussed in this dissertation. Then we review existing work in terms of perceptual duality, depth cues, and size perception.

2.2.1 Definition of Visual Fidelity

Several research examined the fidelity of the virtual environment. Milgram discussed "reproduction fidelity" [94] in the taxonomy of mixed reality displays. The term refers to "the relative quality with which the synthesising display is able to reproduce the actual or intended images of the objects being displayed". As indicated in the reproduction fidelity dimension illustrated in Figure 2.2, this term emphasized on the visual modality by examin-

ing the implementation quality of both hardware and software to reproduce an image. It is a gross simplification of a complex concept, associated with several factors, such as display hardware and graphic rendering techniques, by comparing an intended image to an actual image being displayed. While related, it is different from our intention of providing consistent perception between the real and virtual world.

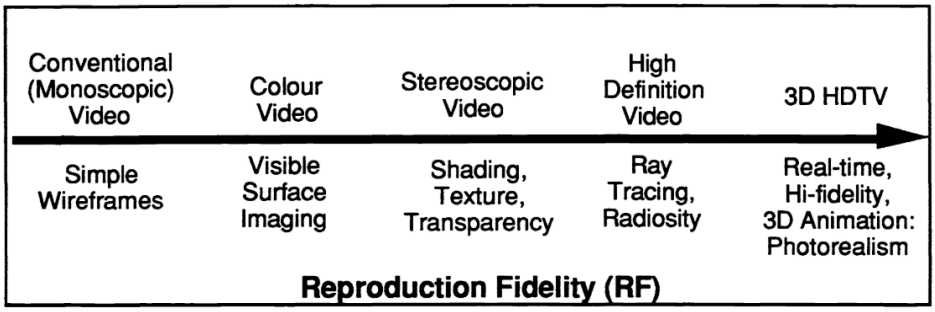


Figure 2.2: Milgram discussed "reproduction fidelity" [94] in the taxonomy of mixed reality displays © SPIE (1995). The term refers to "the relative quality with which the synthesising display is able to reproduce the actual or intended images of the objects being displayed".

Another body of literature defined "VR system fidelity", which consists of "display fidelity", "interaction fidelity", and "scenario fidelity" [84, 93, 108, 109]. McMahan et al. first proposed to use the term of "display fidelity" [93]. The term refers to "the objective degree of exactness with which real-world sensory stimuli are reproduced by a display system". It is distinguished from "interaction fidelity", which refers to "the objective degree of exactness with which real-world interactions are reproduced in an interactive system". The definition of "display fidelity" is broader than Milgram's reproduction fidelity because the former goes beyond the visual modality and deals with other modalities, such as haptic feedback, reproduced by display systems based on the definition, although McMahan's studies primarily examined the visual aspect. In their studies [93], they found that high level of display fidelity significantly improved users' performance as well as subjective judgments of presence, engagement, and usability.

Laha [84] and Ragan [108] further expanded McMahan's definitions to "scenario fidelity". It refers to "the objective degree of exactness with which behaviors, rules, and object properties are reproduced in a simulation as compared to the real or intended experience". It focuses on the realism of

the simulated scenario and the associated model data. As indicated by the name, it is a context-related concept associated with the realism provided by multiple modalities, with a desire to allow users to perceive and interact in a simulated scenario similar to the real or intended environment. Among the three definitions of fidelity in the work of McMahan [93], Laha [84] and Ragan [108], the term "display fidelity" is closest to our *visual fidelity* mentioned in this dissertation. However, there are two differences between them. First, the display fidelity includes multiple sensory modalities while *visual fidelity* only deals with the visual modality. The second difference is more subtle. The fidelity in the virtual environment can be either *reproductional* or *operational*. Being reproductional means exactly replicating the full sensory experience by reproducing equivalent stimuli in the virtual environment. Being operational means revoking the same behavior of users even if the sensory stimuli are different. McMahan's definition of display fidelity emphasized on reproducing the real-world sensory stimuli, while our *visual fidelity* is operational to revoke the same behaviors of users. Given the limitations of current display technologies of using pixels to depict virtual entities, one could hardly reproduce the real-world sensory stimuli as users directly see the pixels rather than the depicted object.

Among the previous work, perhaps our *visual fidelity* is closest to Stefanucci's definition. Stefanucci et al. [128] defined "perceptual fidelity" as "the degree to which scenes presented on a display system are perceived in the same manner as an equivalent real world scene". As indicated by the name, perceptual fidelity emphasized on the providing equivalent perception rather than generating equivalent stimuli. Therefore, high level of perceptual fidelity means a virtual object can be perceived and interpreted in the same manner as an equivalent real object. In this dissertation, we use Stefanucci's definition of perceptual fidelity. Since perceptual fidelity can be divided into visual, auditory and haptic fidelity, we focus on the visual modality, which is a subset of Stefanucci's perceptual fidelity.

2.2.2 Perceptual Duality

When users perceive 3D objects in screen-based 3D displays with various depth cues, they are actually looking at pixels on a 2D screen. A perceptual duality exists between the object's pixels and 3D percept such that users can either perceive the object in a 3D space, or as a 2D representation on the screen. In human vision science, similar perceptual duality has been studied in the real environment. They found that the inherent dual reality in paintings or photographs enables viewers to perceive a scene as 3D at the

same time see the flat surface of the picture [44, 46, 136]. In the virtual environment, few work has investigated this duality. Ware [142] discussed the duality of size perception in the virtual environment as "a choice between accurately judging the size of a depicted object as though it exists in a 3D space and accurately judging its size on the picture plane". Benko et al. [11] mentioned this ambiguity as object presence. Using projector-based 3D displays, they investigated when visualizing without the stereo cue, whether uses could perceive the presence of virtual objects as spatial rather than as 2D projections on the screen surface. Perhaps the most related work will be Elner's study on the phenomenal regression to the real object in the virtual environment [35]. When matching the perspective size of a virtual object to a real object at different distances, they found participants had a tendency to report the size towards the real size, indicating the spatial perception might be related to both the standard and perspective stimulus. As there has been limited work investigating the perceptual duality, it remains unanswered as how the on-screen imagery could impact spatial perception.

2.2.3 Evaluation of Depth Cues

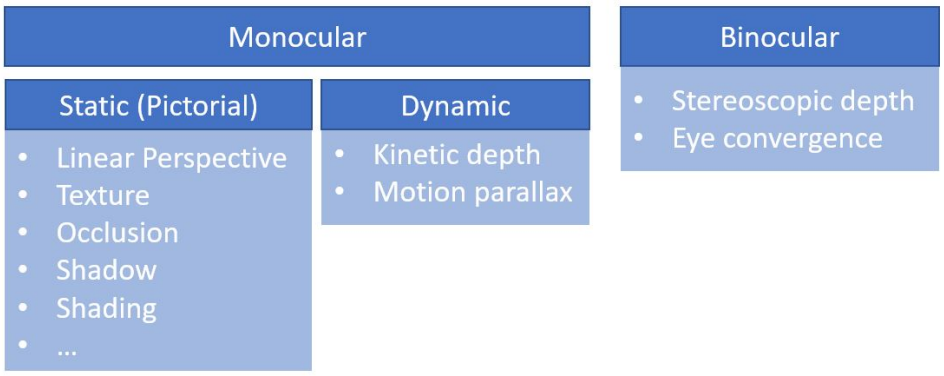


Figure 2.3: List of depth cues discussed in Section 2.2.3. A comprehensive review of the depth cue theory can be found in [142].

Depth cues are the visual cues that help us to perceive objects and scenes in a 3D space. There has been a large body of research which investigates the way the visual system processes depth cue information to provide an accurate perception of space. A list of important depth cues is shown in Figure 2.3, including monocular cues and binocular cues. Monocular cues require only

one eye to be perceived, which can be further divided into static and dynamic monocular cues. Static monocular cues, sometimes also known as pictorial cues, are classic depth cues widely presented in paintings, photographs and computer graphics, including but not limited to: shading, shadow, occlusion, texture and linear perspective. Dynamic monocular cues are the visual cues from the optical flow when objects or viewers are moving, such as kinetic depth effect and motion parallax. A classic example of kinetic depth effect is the spinning dancer illusion resembling a pirouetting female dancer [146]. It is ambiguous to interpret the dancer's movement as one direction or the other due to the kinetic depth effect. In comparison to monocular cues, binocular cues require two eyes to perceive, including stereoscopic depth and eye convergence. A comprehensive review of the depth cue theory can be found in Ware's *Information Visualization* [142].

These depth cues have been adequately simulated in the virtual environment. Computer generated images provide pictorial cues in various applications from 3D games to movies. Advanced 3D displays utilize head-tracking and shutter glasses to provide motion parallax and stereopsis. The relative importance of these depth cues in the virtual environment has been investigated by a number of work. As an early work, Wanger et al. [141] evaluated the influence of pictorial cues on perceived spatial relations in computer generated images by examining the accuracy of matching the position, orientation and scale of two virtual objects. They found the linear perspective and shadow cues are important in the positioning and orientation tasks, and the interaction between different pictorial cues is important in the scaling task. For non-pictorial cues, Arthur et al. [6] assessed the relative importance of stereopsis and motion parallax cues in a path-tracing tasks using a FTVR display. They reported that motion parallax is more important than stereopsis to understand 3D structures. Ware [144] reported similar findings in a graph-tracing task and also noted dynamic monocular cues such as motion parallax or kinetic depth cue can help in understanding 3D graphs. However, Arsenault and Ware found the stereopsis cue is considerably more important than motion cues to guide hand movements in a Fitts' Law tapping task [5]. The diverging results of previous studies indicate that the effectiveness of depth cues is task-dependent.

These early work identified important cues in different tasks and provided guidelines on how to effectively apply these cues in 3D designs [142]. Nowadays, due to the advances of display technologies, most depth cues such as stereopsis and motion parallax can be easily provided in 3D displays. CAVE and FTVR displays provide stereopsis cue via shutter glasses and high-frequency screens. HMDs provide motion parallax cue via head-

tracking. However, presenting necessary depth cues may still not be sufficient to provide consistent perception between the virtual and real environment. In particular, depth cues such as sterepsis, accommodation and motion parallax are rendered from the screen, not directly from the virtual entity in its physical spatial location. The spatial perception of the screen itself and the graphical image projected on the screen are inherently conflated. Hence it is important to examine the role of the screens in 3D displays to provide perceptual consistency between the virtual and real world.

2.2.4 Depth and Size Perception Evaluation

3D displays provide extra depth information, which can potentially facilitate 3D understanding. A number of work have been conducted to investigate the depth perception with different 3D displays such as HMDs and FTVR displays. While HMD has been classified as an "inside-out" display, which allow viewers to look outwards from inside the display volume, FTVR displays are "outside-in" displays, which support looking inwards from outside the display volume [106]. Due to the different visualization nature between HMD and FTVR displays, perception studies of these two categories of 3D displays have focused on different aspects of depth perception. Depth perception studies using an HMD mostly investigated the *egocentric distance*, defined as the interval between the viewer and the virtual object [28]. Much research used a verbal estimate task [4, 75, 96] as well as action-based tasks such as the blind walking or pointing task [20, 70, 74, 96] to evaluate absolute depth perception with HMDs. Egocentric depth has been measured mostly in the action space and frequently reported to be under-estimated in the VE [118]. In contrast to the extensive research on egocentric distance, there is limited work on *exocentric distance*, defined as the distance between two virtual objects. As one of the early work, Arthur et al. [6] evaluated exocentric depth using a path-tracing task with a FTVR display. They reported that depth cues, such as the stereopsis and motion parallax, helped users to perceive relative depth to understand the 3D graph structure. Ware et al. [144] reported similar findings in a graph-tracing study and also noted any structured motion cues such as head-coupled perspective can help in perceiving depth. Grossman et al. [45] evaluated exocentric depth on a volumetric display using a depth ranking task, a collision judgment task and a path-tracing task. They found volumetric displays enable significantly better user performance than other 3D display techniques. Geuss et al. [43] found both the egocentric and the exocentric depth are under-estimated in a turn-and-walk task using HMDs. Other researchers have used a depth-ranking and

path-tracing task to evaluate the exocentric depth with other 3D displays [9, 12, 126]. Being simple and straightforward, the depth-ranking task has been found to be effective to evaluate exocentric depth for closer distances.

While there has been considerable work studying depth perception with different 3D displays, the number of size perception studies in the VE is quite limited. Most of the research investigated size perception with an HMD using a size-matching and size-judgment task. Eggleston et al. [34] found size-constancy is weak in a VE compared to the real world using an HMD. Kenyon et al. [91] further found depth cues like stereopsis and familiar environmental objects can help establish size-constancy in the VE. They used a size-matching task to measure size perception in different viewing conditions. Ponto et al. [103] investigated the size perception using a shape-matching task and found that accurate perceptual calibration will significantly improve the size perception. Stefanucci et al. [128] used size-judgment tasks to assess the perceived size of virtual objects. They found the size in the VE is underestimated compared to the real world. Kelly [74] investigated the re-scaling effect caused by walking through the VE also using a size-matching task. They found walking through a VE causes rescaling of perceived space with an HMD. Benko et al. [11] evaluated size perception with Dyadic Projected Augmented Reality and found participants are able to deduce the size and distance of a virtual projected object with a size-judgment task. Elner and Wright [35] reported a direct measure of VE visual quality in a distance and size estimation task with an HMD. To summarize, size perception in the virtual environment is measured via size matching or judgment tasks, mostly using CAVE and HMD systems, with the results showing a trend of size underestimation [28].

Despite of different 3D displays utilized in previous studies, it is promising to use similar methodologies and tasks to evaluate spatial perception in the virtual environment due to the effectiveness demonstrated from the results. Yet most previous work focus on HMD and CAVE; there is a clear gap in the perception studies for FTVR displays.

2.3 Hardware and Software Display Techniques

To provide accurate perception in the virtual environment, it is important to ensure the display renders the correct perspective in real-time based on the viewpoint, which requires hardware and software support to create the 3D visual display. In this section, we review the technology to implement the 3D visual displays, with a particular focus on the tracking and calibration

techniques to provide the fidelity necessary for FTVR displays.

2.3.1 Multiple-Projector Display Calibration

Multiple-projector systems tile multiple projectors to create scalable displays, typically for large-scale screens. The challenge for multiple-projector systems lies in the stitching and blending of images from different projectors to create a seamless imagery. This requires geometric and photometric calibration. The geometric calibration of multiple-projector systems uses cameras to record correspondences from the known pattern to the observed projected pattern. We summarize a few approaches that are most closely related to our effort in calibrating the spherical FTVR display described in Chapter 3.

Despite substantial work on calibration techniques for planar screens [21, 23, 110, 112], the non-linearity of the curved screen is a challenge for multiple-projector system calibration. An early work [113] presents a calibration approach of the non-planar surface using a stereo camera pair to recover intrinsic and extrinsic parameters and reconstruct the non-planar surface. This approach has been further improved by focusing a subset of curved screens called a quadric screen to recover a quadric transformation [115]. Another approach [49] uses physical checkerboard patterns attached on the curved display to provide the camera a composition of 2D-mesh-based mappings. Their approach aims at a class of curved surface that can be bent or folded from a plane. However, the use of a physical marker on the display causes limitation in the application space.

Majumder et al. proposed a series of automatic calibration approaches for non-planar screens [120–122]. Using an uncalibrated camera, they computed a rational Bezier patch for a dome screen. This approach worked for various shapes such as extruded surfaces, swept surfaces, dome surfaces and CAVE-like surfaces. The camera is mounted on a pan-tilt unit to cover the entire display. Their approaches aim at large-scale immersive displays.

Teubl et al. developed Fastfusion [134] which automatically calibrated multiple-projector systems with different shapes of screens. Fixed warping is utilized to register imagery from projectors to avoid geometry reconstruction of the display. For a spherical shape, an implicit linear assumption has been made using a homography transformation between the camera and the projector. This causes observable misalignment and distortions in the overlapping areas.

2.3.2 Head Tracking Techniques

The importance of head tracking has been long appreciated in the practice of Human Computer Interaction. There has been a tremendous body of literature related to tracking technologies and sensor fusion. Comprehensive surveys on the head tracking and human sensing technologies can be found in [125, 132, 149]. Sensor-based head tracking requires use of sensors such as inertial measurement unit or magnetic sensors [102] mounted on the head to capture the movement of head, while vision based tracking such as Kinect [60] requires the acquisition of the head images using cameras. Both approaches have their advantages and disadvantages. Sensor based tracking can be uncomfortable to users as sensors are mounted on heads with physical contact, while vision based tracking is more user friendly but suffers from configuration complexity and occlusion problems [80]. As vision based tracking performs best with low frequency motion and sensor based tracking such as inertial based sensing performs better for measuring high-frequency rapid motion, a number of hybrid tracking approaches have been proposed to exploit the complementary nature [125]. The configuration complexity such as the synchronisation between multiple sensors has made it challenging to be widely applied. Recent advances of optical marker-based motion capture provides high-fidelity tracking with low latency. A number of commercial motion capture systems have become available and accessible [62, 66].

2.3.3 Error Analysis of 3D Displays

Several research has handled the error for different 3D displays [7, 51, 92, 150]. As a pioneering work, Holloway et al. [51] analyzed different error sources for a HMD using a set of parameters. They found the system delay as a combination of the tracking latency ($11ms$), graphic computation cost ($17ms$) and video sync delay ($16.7ms$) at 60 Hz caused significant registration error of $6cm$. MacIntyre et al. [92] presented a statistical method to estimate the error and further use the estimated error to improve the AR interface of HMDs. Bauer provided detailed analysis and approaches to estimate the tracking error in augmented reality systems [8]. For Non-HMD systems, Cruz et al. discussed tracking noise and delay as error sources in the CAVE [29], with comparisons to a HMD and traditional monitor. They found the monitor ($2.5°$) caused more eye angular error than HMD and CAVE ($0.5°$) for large viewing distance ($> 100cm$). However, for small viewing distance ($< 100cm$), a tracking error of $3cm$ may cause up to $5°$ of eye angular error on CAVE. Kindratenko investigated tracking error in the

CAVE and summarized techniques to correct the location error and the orientation error [79]. Vorozcovs examined sources of error and discussed techniques to reduce these errors in CAVE-like systems [137]. While most of the work focus on HMD and CAVE systems, limited consideration has been given to FTVR displays. It remains to be unanswered on the error analysis of FTVR displays and how different sources of error would contribute to the visual error. As technologies have advanced with better tracking and rendering support, we expect current 3D displays will have less visual error compared to the literature.

2.4 Summary

To summarize, we presented a review of 3D display techniques and discussed the strengths and weakness in support of high *visual fidelity* . Existing work discussed in Section 2.1.1 show the promise of FTVR displays to bridge the perceptual gap between the real and virtual world. Combining FTVR with a spherical display is promising to further enhance the 3D experience with unique properties of the spherical screen. Yet the spherical FTVR displays have challenges from both technical and perceptual standpoints.

We reviewed the technical challenges (Section 2.3.1 and 2.3.3) and the perceptual issues (Section 2.2.2 and 2.2.4) in the virtual environment. The primary technical challenge for multi-screen FTVR displays is to stitch and blend display content from multiple screens to create a seamless display. Even small inaccuracies and errors induced by the display or tracker may break the 3D illusion and cause visual artifacts. It remains to be unanswered on the error analysis of FTVR displays and how different sources of error would contribute to the visual error.

While considerable work have presented spherical display prototypes, previous literature reveals little conclusive findings on the fundamental spatial perception when using spherical displays. Like other screen-based 3D displays, FTVR displays create 3D illusions by rendering view dependent imagery on 2D screens. When users perceive 3D objects, they are actually looking at pixels on a 2D screen. A perceptual duality exists between the on-screen pixels and 3D percept such that users can either perceive the object in a 3D space, or as a 2D representation on the screen. Yet there has been very limited work investigating this perceptual duality and its potential influence on size perception.

To clear these obstacles in support of high *visual fidelity* , we have identified four areas to contribute to this thread of research with two as technical

contributions and the rest two as human factor contributions. They are: (1) an automatic calibration approach to blend multiple screens, (2) an error analysis of a spherical FTVR display to establish design guidelines, (3) an evaluation of spatial perception in a spherical FTVR display and (4) an evaluation of perceptual duality in FTVR displays.

Chapter 3

Multiple-projector Spherical Display Calibration

In the previous chapter, we outlined a number of interesting and unique properties of spherical displays identified by existing work, making them a promising platform to support high fidelity of visual experience in the virtual environment. In particular, the capability to provide an unobstructed view from all angles is ideal for creating FTVR visualizations. Providing high resolution, uniformly spaced pixel imagery on the spherical screen is important for constructing Fishbowl VR displays. One approach is to tile multiple projectors on the spherical screen to increase the resolution and make the system scalable. The challenge for this lies in the stitching and blending of images from different projectors to create seamless imagery. This requires geometric and photometric calibration of the multiple-projector system.

3.1 Introduction

Geometric calibration of a multiple-projector system typically uses a camera to record correspondences from the known pattern to the observed pattern. Systems with a planar screen take advantage of 2D homography transformations to linearly establish the projector-to-display correspondences. While there has been substantial previous work on calibration techniques for planar screens [21, 23, 110, 112], automatic calibration for a curved screen has not received as much attention. A few have investigated approximate correspondence through 2D parameterization either with a linear approximation [134] or physical markers on curved screens [49]. Others have attempted to recover the 3D geometry of the display to establish the mapping [2, 113, 115, 155], but this usually requires a substantial amount of manual interaction.

In addition, previous work has primarily targeted large scale immersive displays like domes to create a sense of immersion. These displays consist of multiple front-projecting projectors and cameras with pan-tilt units to cover the entire display. For relatively small scale desktop FTVR display, projec-

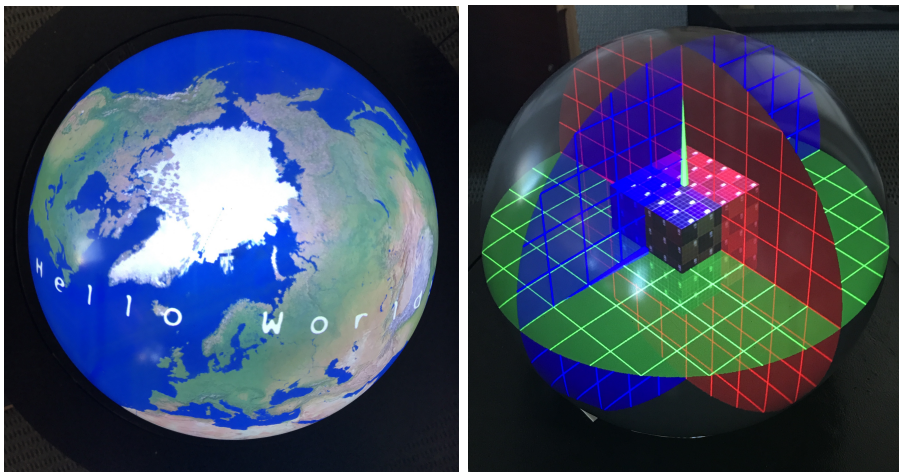


Figure 3.1: Our goal is to calibrate a multiple-projector spherical display with a single camera to allow for a seamlessly blended image. For spherical rendering (left), blending is the most important issue. We target the application of Fishbowl VR display (right) which uses single-person perspective-corrected rendering and requires a more accurate calibration method to provide a higher quality experience to the user.

tors are used in a rear-projection configuration through a small projection hole at the bottom of the spherical screen [10, 133, 155]. Existing calibration methods will not work because the camera's view is mostly blocked by the edge of the projection hole. Applications in FTVR display require accurate geometry registration to support perspective-corrected viewpoints and subtle interactions in real-time. Taken together, the configuration of the Fishbowl VR display makes the calibration challenging in the following aspects:

- **Scale** The Fishbowl VR display is a desktop system with a small-scale screen compared to large-scale immersive displays. Space is quite limited for the camera and projectors beneath the screen.
- **Visibility** The camera view will be occluded by the edge of the small projection hole at the bottom.
- **View-dependent applications** The calibration results will support view-dependent applications by generating perspective-corrected imagery in real-time.

This chapter provides an automatic calibration method that meets these requirements and supports applications in a desktop Fishbowl VR display. We start with a semi-automatic approach that solves these problems. This approach begins with a pre-calibration step that outputs the intrinsic parameters of the camera and the projectors. Then each projector is paired with the same camera to form a stereo pair. For each pair, a pattern projected onto the display is captured by the camera to recover extrinsic parameters via the essential matrix. Using intrinsic and extrinsic parameters, we triangulate projected features and compute the sphere's pose via Weighted Least Squares. The parameters of the sphere (pose) and the camera/projector pairs (intrinsics and extrinsics) are further refined via a nonlinear optimization. Finally we recover the 3D position of each pixel on the display surface via sphere-ray intersection for each pixel per projector.

Up to this point this approach is semi-automatic because it requires additional work to calibrate the intrinsics of the projectors. We provide further improvement of this approach by avoiding the separate calibration of projectors: by estimating the fundamental matrix using the projected pattern on the sphere, we recover the absolute dual quadric for each projector, which is then used to recover the intrinsic parameters of projectors.

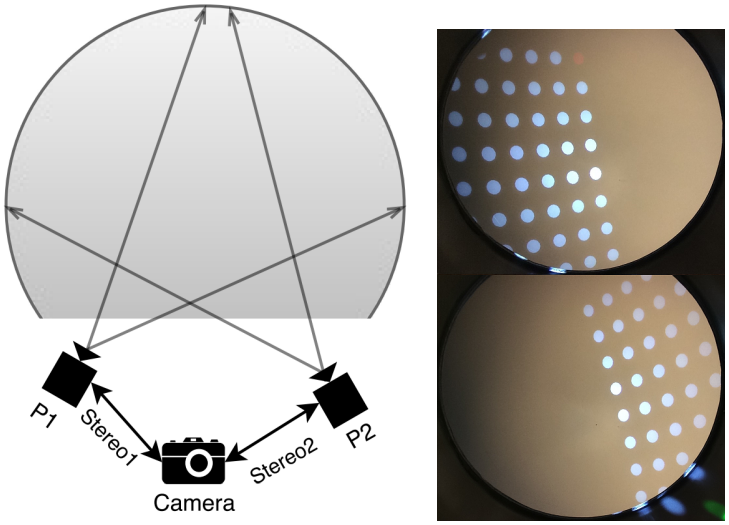


Figure 3.2: left: Multiple-projector spherical display layout with an example of two projectors (P1 and P2) showing overlapping back-projections with respect to the camera. right: Projected blob patterns on the spherical display surface observed by the camera, for each of two projectors.

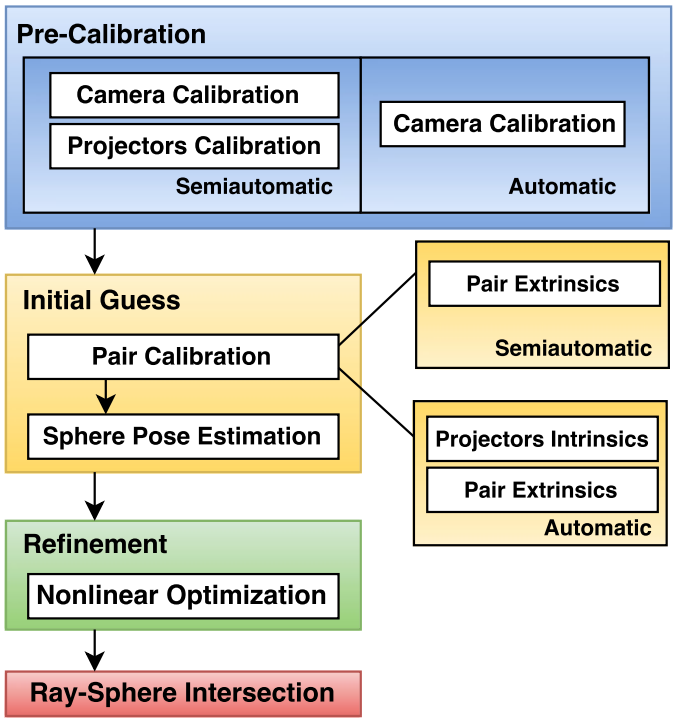


Figure 3.3: Calibration pipeline of a semi-automatic and automatic approach for a desktop Fishbowl VR display. For semi-automatic approach, projectors and camera are calibrated in the pre-calibration step. For automatic approach, only the camera is calibrated in the pre-calibration step. An additional step of blending (intensity normalization) which adjusts the intensity of overlapping area between adjacent projections is followed by ray-sphere intersection based on [113].

We also introduce a practical evaluation method using the camera to estimate the accuracy of our approach, using on-screen metrics instead of reprojection error. We can measure the misalignment by matching points and lines between the observed pattern and the expected pattern.

3.2 Calibration Approach

Our Fishbowl VR system consists of multiple projectors and a spherical screen. As shown in Figure 3.2(left), the projectors work in rear-projection mode through a small projection hole at the bottom of the spherical screen.

Perspective-corrected images are generated on the sphere based on the viewer's position.

The calibration approach we propose is presented in Figure 3.3. We start with a semi-automatic approach that supports applications in a desktop Fishbowl VR display. Then we enhance the method to make it automatic by avoiding the separate calibration of projectors. This approach is realized by recovering the absolute dual quadric for each projector using the fundamental matrix.

3.2.1 Semi-Automatic Calibration Approach

Pre-Calibration The camera and the projectors are pre-calibrated to determine the intrinsic parameters. The camera is calibrated using a standard checkerboard calibration approach [151] to estimate the nine intrinsic parameters including focal length (2), principal point (2) and lens distortion (5). The checkerboard-based calibration approach [151] typically yields the re-projection error of one pixel. Each projector is calibrated using a plane-based calibration approach [39] with the help of the calibrated camera to estimate the four intrinsic parameters including focal length (2) and principal point (2). The plane-based calibration approach [39] typically yields the re-projection error of 1-2 pixels. After this step, the intrinsic parameters of camera and projectors are recovered.

Pair Calibration Each projector P_i and the camera C are paired as a stereo pair S_i as shown in Figure 3.2(a). Although we have the pre-calibrated intrinsics for each pair, the extrinsics are still unknown. In this step, we project blob patterns onto the spherical screen, detect them as blob features in the camera, and record feature correspondences for each pair as shown in Figure 3.2(b).

Using these correspondences, the essential matrix E_i can be recovered since we know the intrinsics for the pair S_i. Then we extract the rotation R_i and translation T_i from the essential matrix E_i using Singular Value Decomposition. Although there are four possible solutions for the calibrated reconstruction from E_i, only one solution is the correct reconstruction that has 3D points in front of both C and P_i. Thus testing with a single point to determine if it is in front of both C and P_i is sufficient to select the correct solution [48] for S_i. The camera center is chosen to be the origin for all projectors and the camera with all three axes aligned with the camera.

However, the translation vector T_i is recovered up-to-scale such that the coordinates between pairs are still up-to-scale. To solve this problem,

we choose one pair S_0 as the "standard" pair that is assumed to have a translation with norm of 1, then we estimate scale factors for other pairs with respect to S_0. These scale factors are computed using the knowledge that points between pairs are on the same sphere. We first triangulate blob features for each pair using the up-to-scale extrinsics. Then we fit a sphere for each pair, and compute each scale factor using Linear Least Squares based on the recovered sphere poses (center position and diameter) from S_0 and S_i. After this step, each pair has extrinsics in the same camera-centered coordinates.

Sphere Pose Estimation With intrinsics, extrinsics and 3D points in the camera-centered coordinate system, the sphere pose can be recovered by fitting a sphere with these 3D points using a Weighted Linear Least Squares [130]. The weighting comes from the re-projection error in the triangulation step so that a large re-projection error results in a small weight in determining the sphere pose.

From these steps we have calculated a full set of parameters that affects this system: intrinsics, extrinsics and sphere pose. However, parameters like extrinsics and sphere pose are roughly estimated. Thus, using these parameters to compute a 3D position for each pixel on the display sphere may cause significant errors. So we use these results as an initial guess for a nonlinear optimization to refine them.

Nonlinear Optimization Parameters are refined using a non-linear optimization with the previous result serving as an initial guess. We now describe the error function we use for the non-linear optimization.

Assume we have 1 camera and N projectors. Camera parameters $\vec{p}_c$ have 9 degree of freedom (DOF): 4 for the focal length and the principal point; 5 for lens distortion [33]. Each projector has parameters $\vec{p}_{p_i}$ with 10 DOF: 4 for the focal length and the principal point; 3 for rotation and 3 for translation. Sphere parameters $\vec{p}_s$ have 4 DOF: 3 for the center position and 1 for radius.

For each pixel $\vec{x}_{p_{ij}}$ in projector P_i with the subscript j representing jth pixel, a ray is back-projected and intersects with the sphere at the point $\vec{X}_{ij}$. The back-projection and ray-sphere intersection can be expressed as a function f based on variables $\vec{p}_{p_i}$ and $\vec{p}_s$:

$$\vec{X}_{ij} = f(\vec{x}_{p_{ij}}; \vec{p}_{p_i}, \vec{p}_s) \tag{3.1}$$

Then the 3D point $\vec{X}_{ij}$ is observed by the camera at pixel $\vec{x}_{c_{ij}}$ on the

image plane. This can be expressed as a function g based on variables $\vec{p}_s$ and $\vec{p}_c$:

$$\vec{x}_{c_{ij}} = g(\vec{X}_{ij}; \vec{p}_c) \tag{3.2}$$

Substituting equation (3.1) into equation (3.2), we get a function F that models this whole process:

$$\begin{aligned} \vec{x}_{c_{ij}} &= g(f(\vec{x}_{p_{ij}}; \vec{p}_{p_i}, \vec{p}_s); \vec{p}_c, \vec{p}_s) \\ &= F(\vec{x}_{p_{ij}}; \vec{p}_c, \vec{p}_s, \vec{p}_{p_i}) \end{aligned} \tag{3.3}$$

Since we know exactly which pixel $\vec{x}_{p_{ij}}$ has been projected from projector P_i, the error function is formulated as the re-projection error in the camera:

$$E = \sum_i \sum_j d(x_{c_{ij}}, F(x_{p_j}; \hat{p}_c, \hat{p}_s, \hat{p}_{p_i}))^2, \tag{3.4}$$

where $x_{c_{ij}}$ is the detected point from camera and $F(x_{p_j}; \hat{p}_c, \hat{p}_s, \hat{p}_{p_i})$ is the estimated point based on parameters $\hat{p}_c, \hat{p}_s$ and $\hat{p}_{p_i}$. The analytic expressions and derivations of equation 3.1, 3.2, and 3.4 can be found in Appendix B.

For a system with N projectors, there are $13+10N$ variables to refine. We use a Levenberg-Marquardt algorithm to solve this non-linear least square problem. The solver is initialized using our previous results.

Ray-Sphere Intersection After refining the parameters, we compute the 3D position for each pixel on the display via ray-sphere intersection with rays coming from each projector by solving a quadratic equation B.4. The geometric result for each pixel is stored in a look-up table.

When there are two real solutions of the quadratic equation B.4, which indicates there are two points intersected between a ray and a sphere, we always choose the point further from the ray origin as we use rear-projection through a projection hole. When there is no real solution, which indicates the ray does not intersect with the sphere, we store a value of $(0, 0, 0)^T$, indicating the corresponding pixel is off sphere and should be discarded in rendering. In a rare case when there is one real solution, indicating the ray is tangent to the sphere, we store the result of that solution. It should be noted that the reconstruction does not determine the scale of the display. In practice, the geometric result is normalized and stored as RGB values in an image format for each projector as shown in Figure 3.4.

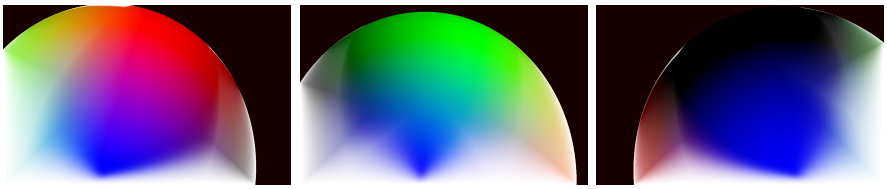


Figure 3.4: Example of calibration result stored as images for a three-projector spherical display. Black values represent off-sphere pixels and the circle pattern shows the projection hole. The 3D positions are stored in RGB channels of the image and the intensity weight for blending is stored in the alpha channel.

Alpha Blending To create a seamless imagery on the screen, the perceived intensity in the overlapping area between adjacent projections needs to be normalized. We implemented the alpha blending approach by Raskar [113] using the geometric result of pixels from the previous step. We create an alpha mask for each projector with a weight factor between 0 and 1 for each pixel. For the same point on the screen, the sum of weights from all projectors will add to unity which yields to normalized intensity. The weight A_{ij} for the jth pixel of projector i can be computed as:

$$A_{ij} = \frac{d_i(i,j)}{\sum_{k=1}^{N} d_k(i,j) * b_k(i,j)} \tag{3.5}$$

where $d_i(i,j)$ is the arc distance from the jth pixel of projector i to the nearest edge of projected image from projector i, and $d_k(i,j)$ is the arc distance from the jth pixel of projector i to the nearest edge of projector k. $b_k(i,j)$ is a binary value of 0 or 1. It equals to 1 when the jth pixel of projector i is within the frustum of projector k, and to 0 otherwise.

When the jth pixel of projector i does not overlap with other projections, $b_i(i,j) = 1$ and $b_k(i,j) = 0$ for all $k \neq i$. Therefore the weight A_{ij} will be 1. When the jth pixel of projector i overlaps with other projections, the weight A_{ij} will be a value between 0 and 1. A_{ij} is inclined to 0 when the jth pixel is close to the edge of projected image from its own projector i. Therefore more weights are given to the pixels from other projectors at the same point on the screen. The alpha blending result for each pixel is stored in the alpha channel of the calibration result.

To summarize, our semi-automatic method uses projector and camera intrinsics as prior knowledge to estimate the extrinsics and display pose parameters. These estimations are then used as an initial guess in a nonlinear optimization to refine the result. In this process, the camera and projectors are calibrated once and can be used afterwards. If there is disturbance on these devices, the system can be re-calibrated by projecting blob patterns and using the non-linear optimization. However, as described, this semi-automatic approach still requires manual work in the pre-calibration step to calibrate the intrinsics of the projectors. We now describe our automatic approach that recovers projector intrinsics together with the extrinsics directly from projected patterns on the spherical display.

3.2.2 Automatic Calibration Approach

In this section we revisit the calibration pipeline in Figure 3.2 and present techniques to make the workflow automatic.

Pre-Calibration As illustrated in Figure 3.3, only the camera's intrinsics are determined for use as prior information in the same way as described in Section 3.2.1.

Pair Calibration In this step, we first determine the internal projector parameters directly from the uncalibrated images. This is essentially an auto-calibration problem described by Hartley and Zisserman [48]. For each pair S_i, we project the same blob pattern as in the semi-automatic approach. The fundamental matrix F_i is recovered using these correspondences. We obtain a projective reconstruction for each pair by choosing the projection matrices as $P_0 = [I|\mathbf{0}]$ and $P_i = [[e'_i]_\times F_i|e'_i]$, where e'_i is the epipole in the projector view, P_0 for the camera and P_i for the projector. The reconstruction $\{P_i, X_i\}$ for each projector is up to a projective transformation. Our goal is to find the projective transformation H_i such that $\{P_i H_i, H_i^{-1} X_i\}$ is a metric reconstruction which is up to a similarity transformation. As described by Hartley and Zisserman [48], H_i can be expressed in the form:

$$H_i = \begin{pmatrix} K_c & \mathbf{0} \\ v^T & 1 \end{pmatrix} \tag{3.6}$$

where K_c is the intrinsic matrix of the camera. Since K_c is known, the only unknown is the vector v^T with 3 DOF. Because $v^T = -p^T K_c$ where p^T are the coordinates of the plane at infinity, this is essentially a problem to recover the plane at infinity, p^T, for each pair.

With further information such as the vanishing points, p^T can be recovered. However, vanishing points or parallel lines can hardly be observed in our case because the screen is curved. In most projectors today the principal point is vertically offset at the bottom-center of the image with zero-skew between axes so that the projection is not occluded by the table [97, 111]. We use the assumptions of zero skew and principal points [1] to provide additional constraints needed to solve for p^T.

Encoding the infinity plane p^T in a concise way, the absolute dual quadric Q^*_∞ for each projector under the transformation H_i can be expressed as:

$$Q^*_\infty = H_i \begin{pmatrix} I_{3\times3} & \mathbf{0} \\ \mathbf{0^T} & 0 \end{pmatrix} H_i^T = \begin{pmatrix} \omega_c^* & -\omega_c^* p \\ -p^T\omega_c^* & p^T\omega_c^* p \end{pmatrix} \tag{3.7}$$

where $\omega_c^* = K_c K_c^T$ is the dual image of absolute conic (DIAC) of the camera [48]. Each Q^*_∞ is related to the projector intrinsics in the form of:

$$\omega^*_{p_i} = P_i Q^*_\infty P_i^T \tag{3.8}$$

where $\omega^*_{p_i} = K_{p_i} K_{p_i}^T$ is the DIAC of the projector and P_i is the reconstructed projection matrix of the projector using the fundamental matrix.

In this case, constraints on projector intrinsics can be transferred to constraints on Q^*_∞. Our constraints are the known principal points and zero-skew. This results in three linear constraints for each projector on Q^*_∞:

$$\begin{aligned} \omega^*_{p_i}(1,3) &= pp_x * \omega^*_{p_i}(3,3) \\ \omega^*_{p_i}(2,3) &= pp_y * \omega^*_{p_i}(3,3) \\ \omega^*_{p_i}(1,2) &= pp_x * pp_y * \omega^*_{p_i}(3,3) \end{aligned} \tag{3.9}$$

where pp_x and pp_y are known principal points. This provides a direct solution for p^T. Once we recover p^T, we can get the DIAC and hence the intrinsics of each projector.

[1] For cases in which the principal point is not at the bottom-center, a general camera matrix is necessary. This will be dealt with by the semi-automatic method in which projector intrinsics are pre-calibrated.

The rest of the method is the same as our semi-automatic approach: estimate the extrinsics using these intrinsics, then triangulate and fit a sphere to find the sphere pose, followed by a nonlinear optimization to refine these parameters.

By doing these steps, we avoid the manual work to calibrate the projectors, and the whole calibration can be implemented by projecting blob patterns and detecting projected features automatically.

3.3 Evaluation

While various work have proposed calibration methods for curved screens, none of them have provided an evaluation for on-surface accuracy. Raskar et al. [115] evaluated their methods using root mean square (RMS) re-projection error, others use simulation to estimate percentage errors of the estimated camera and display parameters [120–122]. In this work, we propose an evaluation method that estimates on-surface accuracy empirically. We use this to evaluate our calibration approach and compare variations of our approach. We also include comparison with other work [115, 155] based on RMS re-projection error.

3.3.1 Metrics

We use three metrics to evaluate the result of calibration: *global point error*, *local point error* and *line error*. These three metrics are chosen to characterize the visual artifact and distortion caused by calibration inaccuracy with *point errors* describing the severity of the shadow effect and *line error* representing the non-linear distortion caused by the curved display. The fidelity of calibration is encapsulated in these metrics with the goal to create a seamless and undistorted imagery on a curved screen.

Global point error describes the overall misalignment of the display. We define the global point error to be the displacement between the expected and the actual position of a projected point. There are two units that can be applied to point error: one is the RMS re-projection error in pixels; the other is the arc length in, for example, millimeters, directly on the spherical screen. Because the arc length varies with the size of the spherical screen, we use radians of the sphere to describe on-surface misalignment.

Local point error describes the local misalignment between adjacent projectors. The effect of local point error is usually observed as a ghost effect in the overlapping projectors area. We define it as the displacement between a point from one projector and the same point from its adjacent projector.

Similar to global point error, we use both RMS reprojection error in pixel and on-surface error in radians to describe local point error.

Line error is used to describe the distortion of the overall display. The effect of line error is observed as distortions (i.e. straight lines appear to be curved). This is important for a FTVR system, because the distortion will cause perceptual discrepancies based on the viewpoint. We define the line error to be the angular difference between the expected lines and the observed lines. Ideally the projected line should be collinear with the expected line.

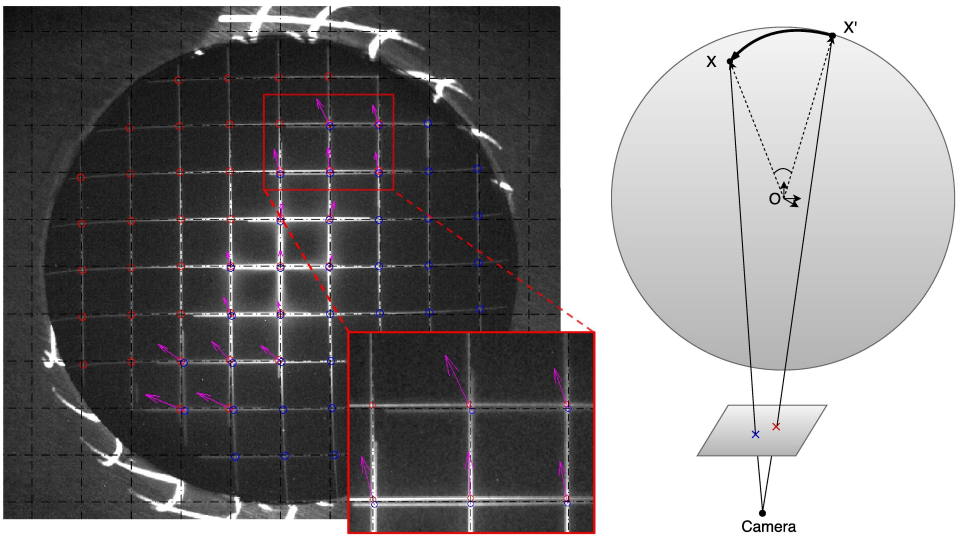


Figure 3.5: (left) Camera's view of projected grid patterns from two projectors (red and blue). Expected grid pattern is drawn with black dash lines and observed projected pattern is drawn with solid white lines. Purple arrows illustrate the local point error in overlapping area. (right) Back-projection of pixel error to estimate on-surface error

3.3.2 Error Measurement

To measure the error, a camera is introduced into the system that observes projected patterns on the spherical screen. The camera is regarded as a virtual viewer with known pose (relative to the display) that observes certain patterns. For example, if a grid pattern is expected from the viewpoint of the camera; then the color value for each pixel in the projector can be determined by projecting its associated 3D position onto the image plane of the camera.

Ideally, the camera will observe a grid pattern which is exactly the same as the expected pattern regardless of the curvature of spherical screen. Then we use the expected grid pattern as ground truth and compare it with the actual observed pattern. The point error is computed based on the locations of crossing features in the grid pattern, while line error is computed based on the angle between the projected line segment and the expected line segment at each crossing feature.

Ideally, a camera should be placed outside the display representing a user who is visualizing the display. In practice, we utilize the same camera used for calibrating the spherical display since the camera has already been calibrated with known intrinsic and pose parameters for the display. As the camera is placed underneath the display, it should be noted that this setup is different from the configuration of the idealized measurement.

Figure 3.5 shows the global point error with dashed black lines as ground truth and solid white lines as actual observations from two projectors with their corner features marked in red and blue circles. Crosses and line segments in the image are detected using template matching. The global point error is computed based on the displacement in pixels between black and white crosses. Shown as arrows, the local point error is computed based on the displacement in pixels between crosses from adjacent projectors in the overlapping area. Finally, the line error is computed based on the slope difference of observed white lines and the expected black lines.

Point errors are evaluated in the form of RMS reprojection error in pixels. Despite being simple and effective, the computed pixel error does not directly predict on-surface registration error [115]. To acquire an estimate of on-surface error, we back-project the displacement in the image onto the spherical screen and compute the arc length on the screen as shown in Figure 3.5. As the on-surface error varies with the size of display, the arc length is computed in radians.

3.3.3 Implementation

We implement the Fishbowl VR system with two mini-projectors and an acrylic spherical screen. The spherical screen has a diameter of 30 cm and a projection hole of 21 cm diameter. The projectors are ASUS P2B [53] with resolution of 1280 x 800. A host computer with a NVDIA Quadro K5200 graphic card sends rendering content to projectors. The rendering contents are created using OpenGL. We employ head-tracking to generate perspective-corrected views for applications in FTVR. The viewer is tracked using Polhemus Fastrak [64] . A two-pass rendering [113] approach based

on head position is chosen due to the non-linearity of the curved screen. We used a single gray-scale camera from FLIR Systems (Flea3 FL3-U3-13Y3M-C) [55] with a camera lens (Fujinon 3.8-13mm F1.4) from Fujifilm [56].

The calibration approaches are implemented using OpenCV [98] and the nonlinear optimization is implemented using Alglib [105]. The semi-automatic approach takes about 20 mins and automatic approach takes about 3 mins to calibrate a two-projector system. Our approach supports view-dependent and view-independent applications in Fishbowl VR displays. Figure 3.1(a) shows a view-independent application after calibrated and blended. The earth image is stitched from different projectors seamlessly. Blending is implemented using an alpha mask technique [113]. Figure 3.1(b) shows a view-dependent application. The viewer is tracked and presented a perspective-corrected images using our calibration result.

3.3.4 Result

We compare the results of the semi-automatic and automatic approaches in Table 3.1. For both, we see substantial improvement after nonlinear optimization. For the semi-automatic approach, the initial guess is computed using a pre-calibrated projector with more accurate intrinsic parameters; hence it has much less error than the initial guess of the automatic approach. This also explains a slightly smaller error after refinement compared with the automatic approach, although the difference is very small. For the automatic approach, although the error before optimization is quite large, the results are largely improved by nonlinear optimization. As a result, the final result of the automatic approach is close to the one from the semi-automatic approach but required no manual interaction.

Figure 3.6 shows the comparison between algorithms with respect to on-surface error. We convert the radian error in Table 3.1 to arc length in millimeters on a 30 cm diameter sphere. Both the semi-automatic and automatic approach can achieve accurate registration: the on-surface point error is less than 1 mm and line distortion is less than 1°. Our approaches are appropriate for Fishbowl VR display where viewers are usually 40 cm to 60 cm away from the screen, with an eye angular error to be no more than 0.14° [155].

Approach	Projectors	Global point error		Line error	Local point error	
		pixel	radian	degree	pixel	radian
Semi-auto	Projector 1	3.9093	0.0107	1.2610	5.6768	0.0167
	Projector 2	5.7653	0.0149	1.4643		
Semi-auto with NLO	Projector 1	1.2556	0.0036	0.8024	1.7165	0.0052
	Projector 2	1.4586	0.0053	0.8465		
Automatic	Projector 1	7.1373	0.0241	1.5683	15.7708	0.0453
	Projector 2	10.5067	0.0294	2.0423		
Automatic with NLO	Projector 1	1.6159	0.0051	0.8268	1.8222	0.0056
	Projector 2	1.7965	0.0064	0.9298		

Table 3.1: Comparing results of our semi-automatic and automatic approach before and after nonlinear optimization (NLO) in terms of global point error, local point error and line error. Measurements come from our implemented system with two mini-projectors. On-surface error is expressed in radians on the sphere.

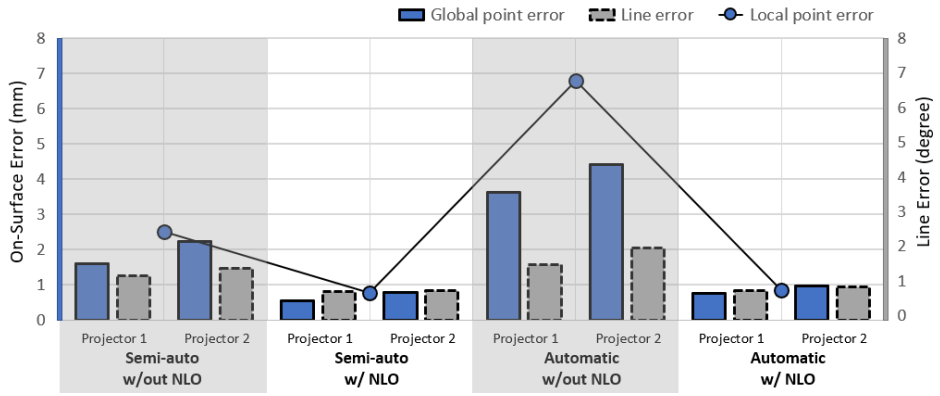


Figure 3.6: Comparing results of semi-automatic and automatic approach before and after nonlinear optimization (NLO). The global point error describes the distance between the expected point and observed point for each projector, while the local point error describes the inter-projector pixel distance in the overlapping area. On-surface error is estimated as the Euclidean distance in millimeters on a 30 cm diameter sphere.

The result is also compared with previous work based on RMS re-projection error in pixel. Existing work have used different metrics and setups to evaluate the accuracy of calibration, such as using the percentage error of pa-

rameters [121] for large-scale screens [122] as well as in simulation [120]. We chose Raskar's work [115] which used a hemispherical dome with a diameter of 1.5 meter to compare the RMS re-projection error in pixel. Our proposed automatic approach has the error of 1.706 pixels and the semi-automatic approach has the error of 1.357 pixels. Our error is higher than Raskar's work [115] with the re-projection error of 0.863 pixels, but lower than our previous work of 2.064 pixels [155], though all approaches are within the same scale. Comparisons are made in each group's own setup. Particularly, the setup of Raskar's work [115] allowed the camera to observe the entire curved screen, which is different from our setup with the challenge of the visibility from the camera's view.

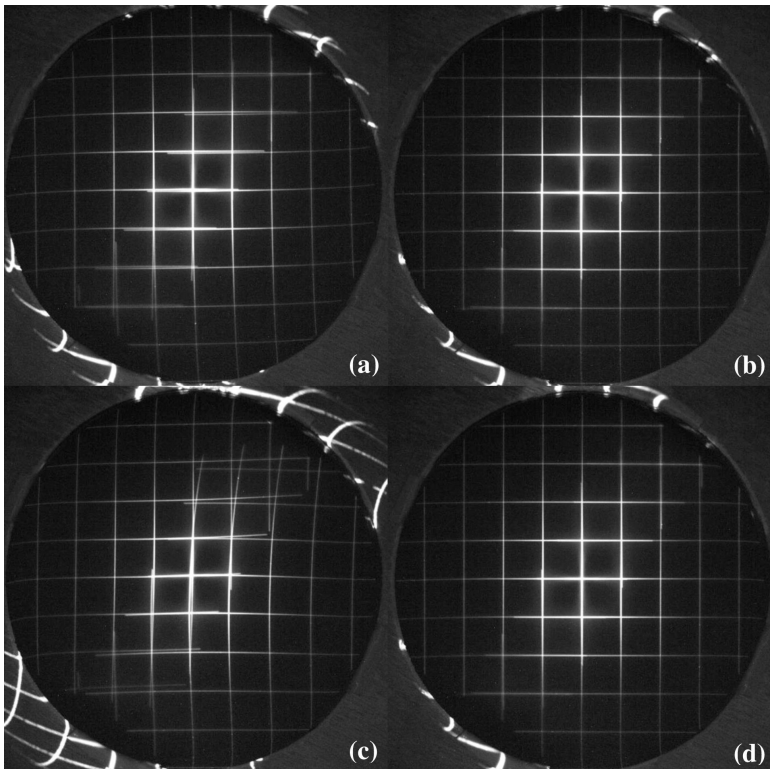


Figure 3.7: Observed grid pattern by camera using the semi-automatic approach (a) before nonlinear optimization and (b) after nonlinear optimization. Observed grid pattern using the automatic approach (c) before nonlinear optimization and (d) after nonlinear optimization

3.4 Discussion and Limitations

In this section, we discuss factors which may affect the results as well as the limitations of the proposed calibration approaches.

Scalability: Though we evaluate the approach with only two projectors, the result generalizes to more than two projectors. To add new projectors into the system requires adding another stereo pair directly registered to the world coordinate system for each projector. Because each projector is registered independently, this does not cause cascading error as the number of projectors increases. In practice, we applied this approach to a 12 inch three-projector system as well as a 24 inch four-projector system as discussed in Chapter 4. The result of rendering is provided in the video [1].

Visibility: One of the challenges to calibrate the Fishbowl VR system is the visibility of the camera. While it is possible to use multiple cameras outside the sphere which can potentially see the entire spherical surface, the setup and calibration of such multi-camera system could also introduce a lot of work. In practice we used a single grayscale camera with a camera lens as described in Section 3.3.3 with a 97 ° horizontal view. We observed more error on the portion invisible to the camera after calibration compared to the visible portion on the screen. To maximize the visibility from the camera and potentially increase the accuracy, we suggest using camera lens with strong distortions such as the fish-eye lens to calibrate the system.

Shape: While this work focuses on the spherical shape, the calibration approach can be applied to other shapes. For geometric primitives with prior knowledge on the shape, such as sphere, plane and cylinder, we can replace the sphere equation B.4 to other primitive equations, which requires modifications in the steps of sphere pose estimation, nonlinear optimization and ray sphere intersection in Figure 3.3. For arbitrary shapes, in addition to the above modification, the display surface needs to be reconstructed in the step of pair calibration after recovering the 3D coordinates of projected patterns via triangulation. However this can only recover visible portion of the display surface from the camera's view without using any geometry assumptions. It is also worth noting that in practice a perfect spherical screen can be hardly found. To some extent the physical screen will deviate from the geometric assumption. Depending on the quality of the screen,

[1] https://youtu.be/mtPR57DEMY8

when the deviation is large, the nonlinear optimization may converge to a local minimum with noticeable error. In that case, one may consider reconstructing the surface without using the geometric assumption. Future work could consider using simulation to evaluate the calibration approach by introducing distortions of the spherical screen similar to [121].

Uniform error: Among all the stereo pairs, there will always be a "standard" pair that has theoretically minimum error. As shown in Figure 3.6, the error in Projector 1 is always smaller than Projector 2, regardless of the metric used. This is due to the scale ambiguity of projector extrinsic parameters. When recovering extrinsic parameters for each pair, we choose one pair as the "standard" pair that has a translation with norm 1; then we estimate scale factors for other pairs with respect to the "standard" pair. The estimation of scale, however, is not accurate since we use the estimated sphere pose in each pair to get a linear solution of the scale as the initial guess, while those estimated sphere poses already contain errors from the previous step. For future work, we suggest an improved method will estimate the sphere pose together with all scale factors using a nonlinear least square solver.

Robustness: While the automatic approach can generate results very close to the semi-automatic one, the former is more sensitive to noise. In our system, if the fundamental matrix has not been estimated correctly due to incorrect feature correspondences or bad lighting, the automatic approach is more likely to fail than the semi-automatic approach. So a trade-off has been made between robustness and priors, which should influence the decision on which to use.

Uncalibrated camera: It is possible to calibrate the system without calibrating the camera. The initial guess of camera intrinsics like focal length can be acquired via EXIF tags of the capture image [121]. However, due to the small projection hole in our system, we use a camera with strong lens distortion to have a wider view on the spherical screen. The focal length is also adjusted to make sure that the unblocked portion is in focus. So in our case it is not appropriate to calibrate the system without having camera intrinsic parameters as priors even with the help of nonlinear optimization.

Lens distortion: For simplicity, we did not include the lens distortion of projectors in the calibration pipeline. In practice, we did not observe no-

ticeable error with the projectors we used (described in Section 4.1). But adding the lens distortion as additional parameters in the nonlinear optimization would ideally increase the accuracy of calibration and should be considered in the future work.

Evaluation metrics: Although direct comparison with existing work using proposed metrics is preferable, limitations inherent in the type of display we are targeting differ from the situations covered by other work, which has been identified as scale, visibility and perspective-correction in Section 3.1. These practical limitations of our multi-projector system are the primary motivations to create the proposed approach. Meanwhile, we include our additional proposed metrics and evaluation method for use by future researchers that have fewer limitations in their systems.

Color calibration: This work mainly focuses on the geometry calibration of a multi-projector spherical display with the assumption that the same model of projectors would be employed in a single setup with negligible inter-projector color difference. Future work should consider this color difference and propose additional color calibration steps to support a consistent view with considerations from both geometry and color aspects.

3.5 Summary

In this chapter, we presented an automatic calibration approach, as well as a practical evaluation method for a spherical multiple-projector display. We identified practical problems in calibrating the display to support the application of Fishbowl VR displays. Our proposed approach solves these problems and achieves less than 1 mm on-surface point error and less than 1° line error using a 30 cm diameter spherical screen. Although our calibration approach is for a spherical multiple-projector Fishbowl VR display, it can be applied to other multiple-projector systems. It should be noted that the proposed calibration approach does not determine the scale of the display. We see the estimation of the scale factor relative to the real world as a separate problem which will be addressed and discussed as a viewpoint calibration procedure [139] in Chapter 4.

Chapter 4

Visual Error Analysis for Fishbowl VR Displays

With a display calibration approach as described in Chapter 3, it is possible to reconstruct the display surface which can effectively support seamless rendering on the spherical screen. To render the perspective-corrected imagery in real-time, it further requires a fast and accurate tracking system to track the viewpoint. However, it is not clear what requirements and specifications of the display and tracking system in order to create an effective FTVR system. Even small inaccuracies induced by the display or tracker may break the 3D illusion and cause visual artifacts. These artifacts include, but are not limited to: distortion (straight lines appear to be curved), ghosting effect (doubled image) and floating effect (virtual objects that are supposed to locate at a fixed location appear to swim about as viewer moves head) [51]. This requires the understanding of how error sources can cause inaccuracies and visual artifacts.

In this chapter, we present the design and implementation of Fishbowl VR display with main results of an end-to-end error analysis. Previous chapter presents a display calibration approach to generate a seamless imagery without considering the viewpoint. In this chapter we introduce the viewpoint to the system and analyze the visual error from the viewpoint. While many FTVR systems have been proposed, few work has included formal analysis of the visual error. It is important to understanding the errors in FTVR displays for the following reasons. First, characterizing the nature and sensitivity of the errors helps to eliminate visual artifacts and provide correct perspectives. Second, when building the system and deciding calibration approaches, one may expect different criteria for system components based on the application. Error analysis provides guidelines when choosing these components.

4.1 Fishbowl VR Display

This section introduces in details the components of the Fishbowl VR display including the hardware such as projectors and screens, as well as the software, such as the graphics rendering approach and the display calibration approach. A diagram of the Fishbowl VR display with its hardware components is illustrated in Figure 4.1(left). A photo of a Fishbowl VR prototype with virtual fruit models is shown in Figure 4.1(right). Further implementation details can be found in [37].

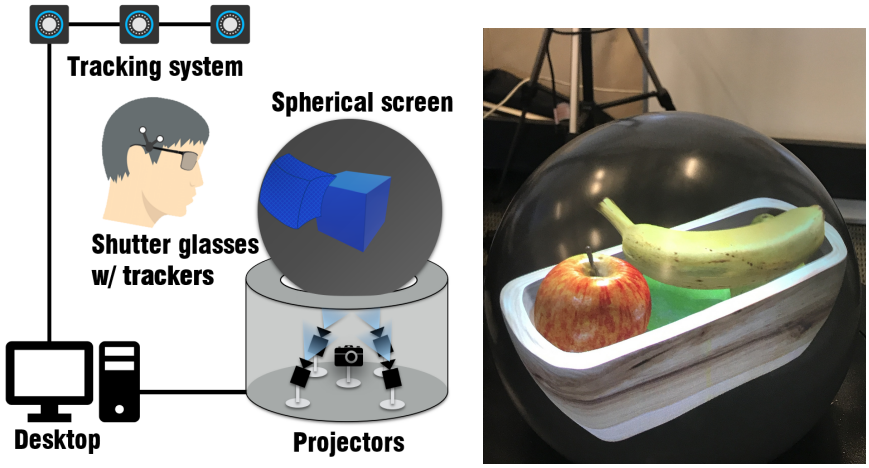


Figure 4.1: (left) Diagram of the Fishbowl VR display with its main hardware components. (right) Virtual 3D fruit models shown on a 12 inch spherical display prototype.

Spherical screen: Our display uses multiple projectors rear-projecting onto a spherical surface. The inner surface of a plexiglass sphere is coated with a translucent projection paint (created by B Con Engineering, Inc.). The imagery is projected on the inner surface of the screen with a thickness of 4 mm. A hole in the bottom of the sphere allows for rear-projection onto the inner surface of the sphere from projectors mounted below as shown in Figure 4.2. The hole size is a trade-off between projector coverage, projector placement and roundness of the final display. We found that a hole size of 75% of the diameter worked well for the projectors we use. We used two surface configurations: a 12 inch diameter sphere with a 9 inch diameter hole (Figure 4.2, left), and a 24 inch sphere with an 18 inch diameter hole (Figure 4.2, right).

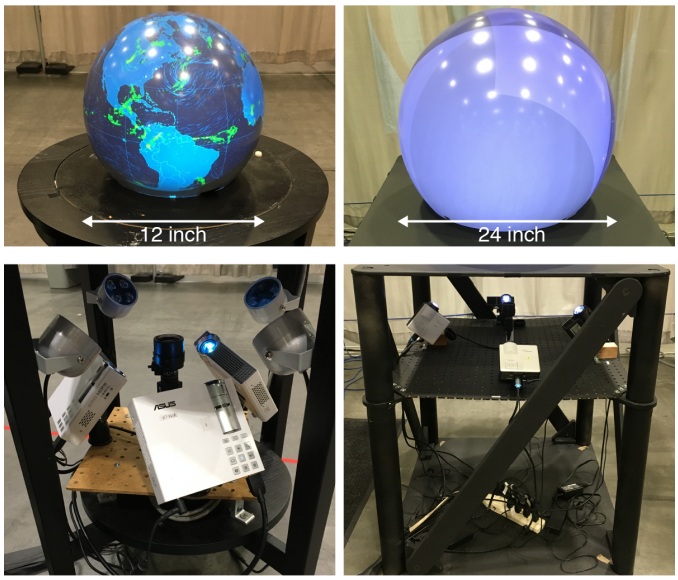


Figure 4.2: A 12 inch diameter spherical prototype built with three projectors (left) and 24 inch diameter spherical display built with four projectors setup (right).

Projector and chassis placement: Projecting onto a curved surface can raise a few concerns. First, consumer projectors have a flat focal plane and a shallow depth-of-field; therefore the corners of the projector when projected onto the sphere can be a bit out of focus. Second, the perceived light intensity of the projectors is somewhat view dependent, which can cause variation in projector brightness as one moves around the display. In practice, we have found that these optical issues are not noticeable, particularly for scenes that do not have a bright background. Chassis can be made from a table with a circular hole cut in the table top to allow rear-projection from the projectors mounted on a lower shelf below the tabletop. The placement of projectors should allow projectors to cover the spherical surface as much as possible and each projector overlaps its neighbor by ∼10%.

The optimal number of projectors depends on the size of the surface and the desired surface resolution. Covering the surface of the sphere with more projectors means that each projectors' visible patch on the surface will be smaller and therefore the effective resolution will be higher. We show two effective configurations: a 12 inch diameter sphere with three projectors (Asus P2B [53], 1024 x 768 resolution at 120Hz) at ∼63.49 pixels per inch

(PPI), and a 24 inch diameter sphere with four projectors (Optoma ml750s [63], 1024 x 768 resolution at 120Hz) at ∼34.58 PPI, with PPI computed as in [37]. The multiple projector approach allows the surface resolution to be scaled up by adding more projectors.

While in this work the placement of projectors relative to the screen is determined manually, future work can look for the optimal placement obtained automatically with a simulation of projection to avoid manual work, which can take parameters such as screen size and desired resolution as input, and output the optimal placement as well as number of projectors required.

Stereo glasses and projectors synchronization: We use short-throw mini-projectors for a compact design. To render stereo views, the projectors require a refresh rate of at least 120Hz. Fortunately, there are a few options for small stereo projectors, such as the Optoma ml750s [63]. We used active shutter glasses for our setup that are controlled wirelessly with a radio-frequency (RF) signal [67]. The NVIDIA Quadro graphics card generates a hardware synchronization signal controlling the stereo glasses. It is possible to use polarized 3D glasses to provide stereo visualization. However, we found the spherical screen interferes with the transition of the circularly polarized light, causing left eye and right imagery incorrectly received with the glasses. Special care on the lens of projectors and the coating of the screen is required to ensure the transition of the circularly polarized light.

Head tracking: FTVR requires tracking a user's viewpoint relative to the display screens. This could be accomplished by eye or face tracking from video, but for robust and low latency tracking it is often done using marker-based head tracking. We have explored two tracking systems in our prototypes. In the early stage of the prototype, we used Polhemus motion tracking systems [64] with electromagnetic trackers attached to the stereo glasses in the 12 inch prototype. While the tracking system has the accuracy of 0.7 mm and the latency of 4 msec, the tracker needs to stay within 30 inches of the electromagnetic source to ensure the tracking accuracy [102], which constrains the movement of users. In the later stage of the prototype, we used the OptiTrack [62] optical tracking system with passive markers attached to the stereo glasses for head tracking and active markers on handheld wands for manipulating virtual pointers and objects within the display. This provides 5 x 5 x 5 meters of tracking volume with sub-millimeter accuracy and the latency of 8 msec.

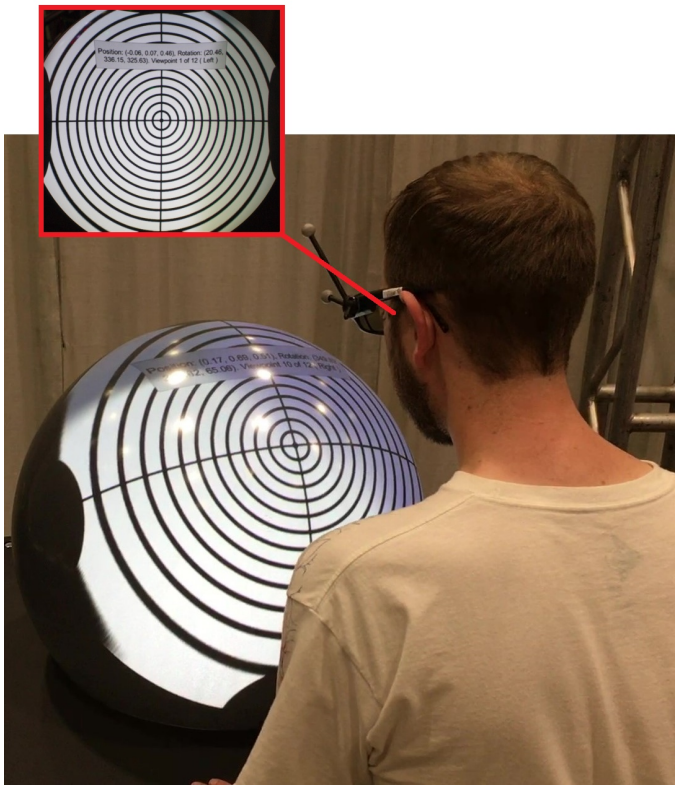


Figure 4.3: Viewpoint calibration procedure to estimate the parameters of the camera model for each eye with respect to the display.

Viewpoint calibration: To accurately register the tracked viewpoint with respect to the display, we employed an interactive perceptual viewpoint calibration approach [138]. During the calibration, a user visually aligns 2D patterns on the spherical surface using each eye at a time such that patterns will appear undistorted if viewed from a known calibration position as shown in Figure 4.3. This is repeated through a set of predetermined positions that define the parameters of the camera model for each eye. Note that the pattern and predetermined positions are generated in the display coordinate system, while the user is tracked in the tracking coordinate system. The viewpoint calibration approach also estimates the similarity transformation between the display and tracking system, which contains the scale factor of the reconstructed display surface. We then use the calibrated geometric model to render the view-dependent content for each eye.

4.2 Rendering

We have developed a general-purpose rendering engine for multi-screen FVTR displays. This section describes the rendering pipeline for Fishbowl VR. More details of the general-purpose rendering engine can be found in [38].

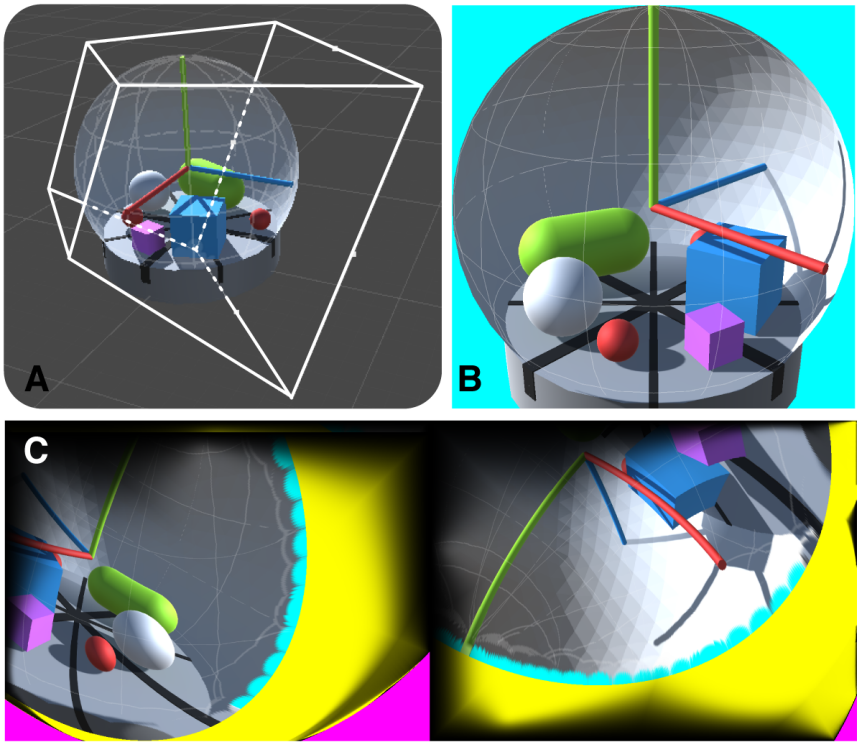


Figure 4.4: Overview of our rendering pipeline. We generate a view frustum for each tracked viewpoint of the scene (A) and render to an off-screen texture (B). The mapping from projector-space to sphere-space is used to non-uniformly sample the rendered texture to generate the pre-warped image for each projector (C, two projectors shown). For illustration, the pre-warped image (C) is colored to show different projector regions, including: magenta regions that are not visible on the spherical surface (because they do not pass through the bottom hole of the sphere), yellow regions that are visible on the spherical surface, but not from the current viewpoint, and black regions that are alpha blended for a seamless transition in the overlap between projectors.

The rendering system is built with the Unity game engine [135]. As illustrated in Figure 4.4, we use a render-to-texture pass and texture sampling to generate perspective-corrected imagery across multiple screens based on Raskar's work [113]. In the render-to-texture pass, a virtual camera is placed in the scene at the tracked position of the user and is rendered-to-texture with a normal rendering pass. For stereo rendering, there is an additional pass for the additional viewpoint. The texture is then sampled in the shader to generate per-pixel color for each projector.

In a second pass, pixel locations in projector-space are transformed to the sphere surface location using the projector-to-sphere map computed during multi-projector calibration (as described in Chapter 3). Sphere-space locations are transformed to the viewpoint camera's clip space with the standard `ModelViewProjection` matrix. These locations are then used to sample the rendered texture saved from the last rendering pass. Finally, we render the sampled color from the rendered texture after applying alpha blending based on the weights computed in multi-projector calibration (Chapter 3). The per-screen texture sampling is efficient as it does not involve 3D rendering, just 2D texture-space sampling.

The above process can be summarized as a two-pass rendering pipeline:

Pass 1

1. Place a virtual camera $\mathbf{C}$ at the viewpoint's position
2. Compute perspective projection from the camera $\mathbf{C}$
3. Render the projection to a texture $\mathbf{T}$

Pass 2

- ▷ For each projector $\mathbf{i}$:
  - ▷ For each pixel $\mathbf{p}$:
    1. Retrieve the 3D position $\mathbf{V_p}$ from a look-up table of projector $\mathbf{i}$ (computed as described in Chapter 3)
    2. Compute 2D coordinate $\mathbf{UV_p}$ via transforming the 3D position $\mathbf{V_p}$ to the clip space of camera $\mathbf{C}$
    3. Sample the texture $\mathbf{T}$ using $\mathbf{UV_p}$ to retrieve the color $\mathbf{Color_p}$ in RGB channels
    4. Apply alpha blending to $\mathbf{Color_p}$ in its alpha channel based on the intensity weight (also computed as described in Chapter 3)
    5. Render the pixel with $\mathbf{Color_p}$

We chose the two-pass rendering approach because of the spherical screen. The non-linearity of the spherical screen is not naturally supported by the perspective projection in a single pass. While variations of perspective projection exist, such as the off-axis perspective projection [82], they are still linear functions. However, screens with a polyhedron shape such as cube or pyramid can consider using the generalized perspective projection such as the off-axis perspective projection [82].

To simultaneously output to multiple projectors with the capability to support stereo visualization, we uses a workstation graphics card that can output frame-synchronized video for left and right eyes. An NVIDIA Quadro K5200 graphics card was used to output and synchronize projectors on each prototype. It is possible to scale up the system with multiple Quadro cards to frame-synchronized video for up to 16 outputs. We use NVIDIA Mosaic technology to unify the image of the projectors into one screen. We use an Nx1 typology (where N is the number of projectors) to represent the circular arrangement of the projectors under the spherical display. Importantly, NVIDIA Mosaic synchronizes all screens in resolution and framerate for stereo rendering. Each projector has a resolution of 1024x768, making a total Mosaic resolution of 4096x768.

4.3 Visual Error Analysis

In this section, we discuss the error model of a Fishbowl VR display. Let $\bar{E}$ represents a viewpoint looking at a virtual point P as shown in Figure 4.5(left). Then the intersection point $\bar{D}$ between the ray $\bar{E}P$ and the sphere is the corresponding pixel on the spherical display. With the existence of errors, D is the pixel actually displayed on the sphere. A number of error sources can cause this discrepancy, divided into two categories as the viewpoint error and display error:

- **Viewpoint Error** Error in the position of the viewpoint $\bar{E}$. There are multiple sources that can cause the viewpoint error, such as the tracker jittering, tracker latency and viewpoint offset between the viewpoint and tracker.
- **Display Error** Error made in displaying the perspective-corrected imagery, caused by the display calibration approach and projector/screen displacement.

In order to find out how these error sources can influence viewers' visual experience, we examine how the viewpoint and display error are applied to

the perceived image individually. Although it would be practical to use a binocular viewpoint model, the analysis of visual error would be complicated and is affected by other factors such as the accommodation-vergence conflict. For simplicity, we assume a monocular viewpoint. Despite of using a simplified model, the result provides practical guidelines for establishing criteria for system components and determining calibration fidelity.

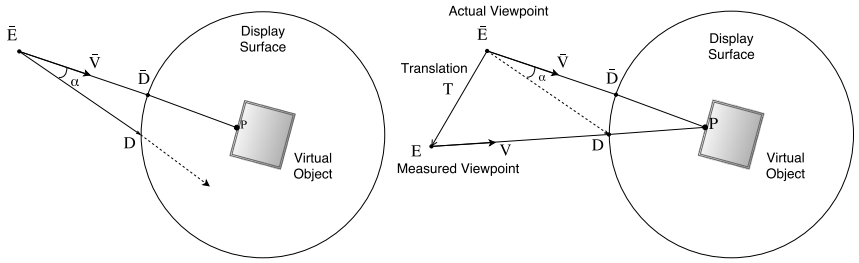


Figure 4.5: (left) Angular error α in the Fishbowl VR display. The discrepancy between the displayed pixel D and the desired pixel $\bar{D}$ is caused by the viewpoint error and display error. (right) Angular error α caused by viewpoint error T in the static visualization.

4.3.1 Metric

We use eye angular error α [29] as the metric of static visual error. The angular error α can be defined to be the displacement between the displayed pixel D and the desired pixel $\bar{D}$ based on the viewpoint $\bar{E}$:

$$\alpha = arccos(\frac{D\bar{E} \cdot \bar{D}\bar{E}}{\|D\bar{E}\|\|\bar{D}\bar{E}\|}), \tag{4.1}$$

As shown in Figure 4.5(left), $\bar{E}$ represents the viewer's position looking at a virtual point P. The discrepancy between the displayed pixel D and the desired pixel $\bar{D}$ is caused by the viewpoint error and display error.

4.3.2 Viewpoint Error

The viewpoint error can be modeled as a translation vector T between the measured position E and actual position of the viewpoint $\bar{E}$ as shown in Figure 4.5(right). Assuming we know the actual viewpoint $\bar{E}$, viewpoint error T and virtual point P, we can solve $\bar{D}$ via ray-sphere intersection

equations:

$$\begin{cases} \bar{D} = \bar{E} + \lambda(P - \bar{E}) \\ \bar{D}^T Q \bar{D} = 0 \end{cases} \tag{4.2}$$

where Q is the quadric surface matrix of the sphere, λ is the distance from the actual viewpoint $\bar{E}$ to the desired pixel $\bar{D}$. Similarly, we can solve for D using E, T, P and Q. With D and $\bar{D}$, the angular error α can be computed using equation 4.1 as a function of the virtual point P, measured viewpoint $\bar{E}$ and viewpoint error T.

4.3.3 Display Error

Display error can be represented by the displacement between the desired 3D position of pixel $\bar{D}$ and its actual location D on the display surface in Figure 4.5(left). This error depends on the accuracy of display calibration described in Chapter 3. While we could use empirical methods to measure the accuracy of calibration with additional cameras [22], we describe an analytical method in this section to estimate the accuracy using error propagation [48].

In equation 3.1, the 3D positions of projected pixels are computed using ray-sphere intersection based on projector projection parameters $\vec{p}_p$ and sphere pose parameters $\vec{p}_s$ in the form of column vector. To understand the influence of projector parameters $\vec{p}_p$ and sphere pose parameters $\vec{p}_s$ on the display error, we compute the covariance matrix of X as:

$$\Sigma_X = J_{X(p)} \Sigma_p {J_{X(p)}}^T, \tag{4.3}$$

where $J_{X(p)}$ is the Jacobian matrix of the vector function $X(\vec{p})$ with respect to the parameter vector $\vec{p}$ and Σ_p is the covariance matrix of $\vec{p}$. The parameter vector $\vec{p}$ can be projector parameters $\vec{p}_p$ or sphere pose parameters $\vec{p}_s$.

For the projector parameters $\vec{p}_p$, the covariance matrix Σ_{p_p} can be estimated using backward propagation of the re-projection error once the projector is calibrated. Assuming $\vec{p}_p$ is the vector that contains projector intrinsic K_p and extrinsic parameters $(R_p \ T_p)$, the camera matrix of the projector can be expressed as:

$$\vec{m} = K_p \begin{pmatrix} R_p & T_p \end{pmatrix} \vec{M}, \tag{4.4}$$

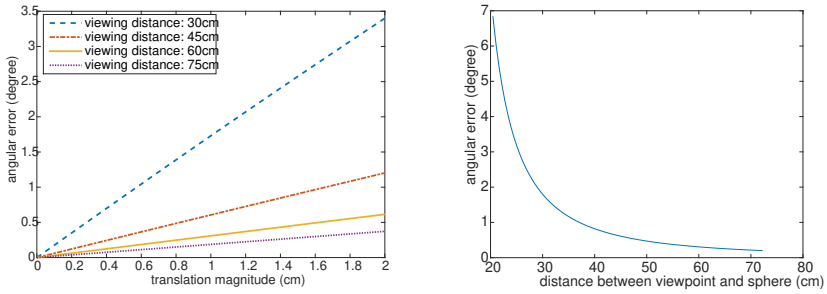


Figure 4.6: (left) Visual angle α increased by the magnitude of viewpoint error T when viewing a virtual point at the center of sphere. (right) Visual angle α increases as the viewpoint E moves closer to the sphere with 1 cm viewpoint error.

where $\vec{M}$ is the 3D coordinates of points and $\vec{m}$ is the resulting image coordinates in 2D. The covariance matrix Σ_{p_p} of $\vec{p_p}$ is computed using backward propagation from equation (4.4):

$$\Sigma_{p_p} = ({J_{m_{pp}}}^T \Sigma_{\mathbf{m}}^{-1} J_{m_{pp}})^{-1}, \tag{4.5}$$

where Σ_m is the re-projection error during projector calibration and $J_{m_{p_p}}$ is the Jacobian matrix of the equation (4.4) with respect to the projector parameter $\vec{p_p}$.

Hence the resulting covariance matrix Σ_X of equation 3.1 can be computed by plugging Σ_{p_p} back in equation 4.3.

The covariance matrix of sphere pose parameters can also be computed in a similar manner using a least square covariance computation. Assume $\vec{p_s}$ is the vector that contains sphere parameters x_s, y_s, z_s and r_s. The covariance matrix Σ_{p_s} can be estimated based on a weighted least square [130].

4.3.4 Result

Using the error model, we insert parameters based on the implementation of a 12 inch prototype described in Section 4.1. In this section, we present the simulation results of our Fishbowl VR design.

Effect of viewpoint error: Angular error α is maximized when the direction of T is perpendicular to the EP. In the following sections, we assume

EP is tangent to T so that we are evaluating based on the worst case. Figure 4.6(left) shows the simulation result of α as a function of $|T|$ while the viewer is looking at the virtual point P placed at the sphere center. The slope increases as the viewer gets closer to the sphere, meaning α becomes more sensitive to translation error if the viewer is closer to the display.

Effect of viewpoint position: Figure 4.6(right) illustrates the simulation result of how the viewpoint position influences α when the viewer is looking towards a virtual point P placed at the center of the sphere. As the viewpoint $\bar{E}$ moves away from the sphere along $\bar{E}P$, α decreases dramatically. To control α, a minimum viewing distance can be established depending on the applications. For example, for interactive applications, the viewing distance is likely shorter than ones only with visualizations. Hence, the FTVR display will require a more accurate tracking system with precise calibration approaches.

Effect of virtual point position: We assume that the virtual point P can be placed both inside and outside the sphere. Figure 4.7 illustrates the changes of angular error α when the virtual point P moves from P_i to the viewpoint $\bar{E}$, where $i = 1 \sim 5$ indicates five different locations of the virtual point. As a function of the virtual point position P_i, the angular error α first decreases when P_i travels towards $\bar{D}_i$ as shown in Figure 4.7 (left), so that the virtual point away from its corresponding display pixel will increase α. When P_i arrives at $\bar{D}_i$, there is no angular error since D_i overlaps with $\bar{D}_i$, suggesting that points on the display surface towards the viewer will not cause angular error even with the existence of viewpoint error. As P_i moves out of the sphere further toward $\bar{E}$, α starts to increase rapidly, meaning points out of the sphere are more sensitive to viewpoint error. Since α depends on the position of virtual point P_i, there exists an "optimal" rendering region with relatively small angular error. For instance, to set a maximum angular error of $0.4°$, virtual objects should be rendered within a region which is less than 18 cm inward away from its corresponding pixel, or 5 cm outward away, denoted as -5cm to +18cm.

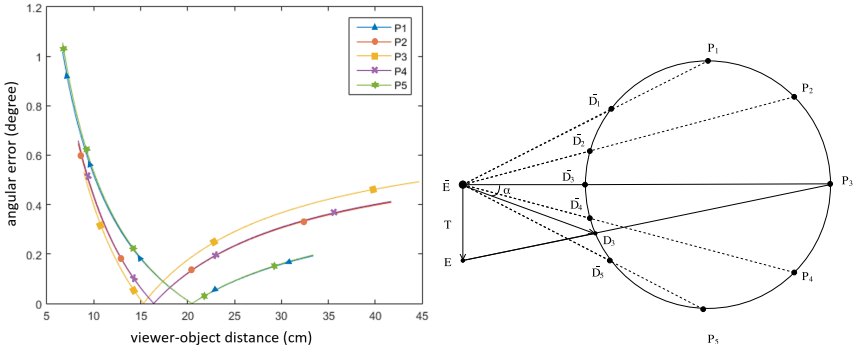


Figure 4.7: Effect of the virtual point position on the angular error when the viewing distance is 30 cm and viewpoint error is 0.2 cm. (right) The geometric layout of different virtual points from P_1 to P_5 in a top-view with the viewpoint E. (left) The visual error α caused by the viewpoint error T changes as a function of the viewer-object distance between the viewpoint E and virtual point P_i. D_i represents the corresponding pixel on the screen depicting the virtual point P_i. α is zero when the virtual point P_i is on the screen overlapping with D_i. α increases as P_i moves away inward from D_i, and increases more dramatically as P_i moves away outward from D_i. This indicates virtual points in front of the screen are more sensitive to the viewpoint error than points inside the screen. Curves of P_1 and P_5, as well as P_2 and P_4 are almost overlapped in (left) due to their nearly symmetric layout in (right).

Based on this analysis, we have quantitative results for three tracking systems we use for a 12 inch FTVR prototype. The average joint accuracy for Kinect v2 [60] using its joint tracking SDK is around 8cm [148]. For a maximum angular error of 2°, the viewing distance should be no less than 60 cm assuming a viewer is looking at a virtual object at the center. At this minimum viewing distance, the rendering range should be -8cm to +11cm from the screen. For Polhemus Fastrak with a tracking accuracy of 0.2 cm within 150 cm of the transmitter, it is possible to yield a 0.5° maximum angular error system. Similarly, OptiTrack can achieve tracking accuracy of 0.1 cm, yielding an angular error of 0.25°, making tracking systems with sub-centimeter accuracy suitable for an interactive FTVR system so viewers can get close enough to the display.

Table 4.1: Estimated angular error caused by different tracking systems with suggested viewing distances and rendering regions for a 12 inch prototype. The viewing distance assumes the virtual object is rendered in the center of the sphere. The rendering region assumes viewing at the minimum distance.

Tracking system	Tracking error	Angular error	Viewing distance	Rendering region
Kinect v2	8cm [148]	<2 degree	>60cm	-8cm to +11cm
Fastrak	0.2cm [102]	<0.5 degree	>30cm	-6cm to +30cm
OptiTrack	0.1cm [62]	<0.25 degree	>30cm	-6cm to +30cm

Effect of display error: We used the calibration results of the re-projection error caused by the display calibration in a 12 inch prototype described in Section 4.1. The average 3D Euclidean distance error for pixels on screen is 0.315 mm caused by the projector parameters, while the average distance error is 1.344 mm caused by the sphere pose parameters. This indicates that the error of sphere pose tends to be more influential than the error of projector calibration. It is important to improve the sphere pose estimation for a smaller display error.

In addition, the display error is not spatially homogeneous. Figure 4.8 shows display errors computed at each pixel of a projector with the projector and sphere pose parameters as error sources respectively. The display error reaches peaks on the corners and falls off from the center by a factor up to ten times the error at the center. The fringe pixels have significantly more errors than pixels at the center possibly resulting in a noticeable misalignment in the overlapping area since overlaps happen mostly along the projection fringe. Hence, it is useful to have a pixel-by-pixel optimization after the Euclidean reconstruction to minimize those local errors and alleviate the misalignment. In practice, the local display error in the overlapping area should be no more than 2 mm to avoid making the misalignment noticeable.

If we compare the display error with the viewpoint error, the display error of 1 mm on the display surface only yields a 0.38° angular error with the viewing distance of 30 cm. So the viewpoint error causes significantly more angular error than display error. This is consistent with Holloway's result for HMDs as the viewpoint error is the major cause for visual artifacts like distortions and floating effect [51]. However, it is also important to minimize the display error since it accounts for the ghosting effect in the overlapping

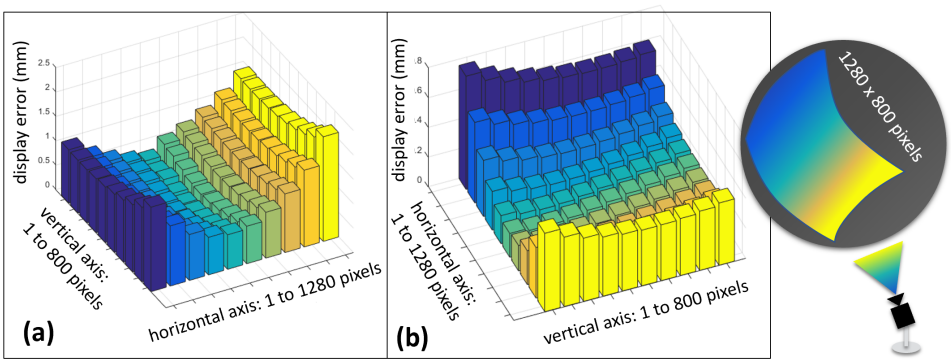


Figure 4.8: Display error varied by pixels caused by (a) projector intrinsics (b) sphere pose as error source. The display error reaches peaks on the corners in the image plane of projectors.

area of adjacent projectors, especially for view-independent applications in which imagery is wall-papered on the entire sphere so that overlapping pixels are always used for rendering.

4.4 Discussions and Limitations

In summary, we found that for the viewpoint error, visual error α increases: 1) as the viewpoint error increases, with its slope influenced by the viewing distance; 2) as the viewer moves closer to the sphere; 3) as the virtual point is further from the screen. For the display error, we found that the parameters of sphere pose plays an important role in recovering the 3D coordinates of pixels on the display. The fringe pixels have significantly more errors than pixels at the center possibly resulting in a noticeable misalignment in the overlapping projector area.

While the error model allows us to establish bounds and requirements for the tracking system and calibration approach, it is important to point out several assumptions and limitations of the error model. Firstly, each error source is analyzed independently. Hence the estimation of visual error is based on the assumption of the independency of error sources, which may not hold for cases when one error source depends on another. For example, for the display error, the estimation of sphere pose may depend on the projector parameters if the calibration approach uses projectors to estimate the sphere pose, which could potentially lead to cascaded error higher than the estimated bounds. Secondly, the error model is based on monocular

viewing while in practice users visualize FTVR displays with both eyes. Future work is required to expand current approach to stereoscopic viewing. Lastly, the error model is based on the assumption of static visualization when viewers remains static with respect to the display. It provides limited insights of the temporal resolution such as the latency of tracking. Future work are required to include the analysis of dynamic visualization when users move around the display.

Being aware of such assumptions, we see the proposed error model useful in providing guidelines for system designs as where to spend the time in order to improve the visual experience. It also gives some insights of what level of errors can be expected for a given set of hardware and software.

4.5 Summary

In this chapter, we present the design and implementation of the Fishbowl VR display by introducing the tracked viewpoint to the calibrated spherical display with the approach described in Chapter 3. We also present the main results of an end-to-end error analysis of the display in terms of the visual error from the viewpoint. By characterizing the nature and sensitivity of the errors, we found: viewpoint error causes significantly more eye angular error than display calibration error; angular error becomes more sensitive to tracking error when the viewer moves closer to the sphere; and angular error is sensitive to the distance between the virtual object and its corresponding pixel on the surface. Taken together, these results provide practical guidelines for building a Fishbowl VR display in choosing the system hardware and software components and can be applied to other configurations of geometric displays.

Chapter 5

Evaluation of Depth and Size Perception

FTVR displays create a compelling 3D spatial effect by rendering to the perspective of the viewer with head-tracking. Combining FTVR with a spherical display has the potential to enhance the 3D experience with unique properties of the spherical screen such as the enclosing shape and borderless views from all angles around the display. The ability to generate a strong 3D effect on a spherical display with head-tracked rendering is promising for increasing user's performance in 3D tasks. An unanswered question is whether these natural affordances of the spherical screen can improve spatial perception in comparison to traditional flat screen.

It is important to provide better spatial perception as it can improve users' performance in 3D tasks which rely on the depth and size perception in 3D applications such as Computer Aided Design (CAD). In this chapter, we described an experiment to determine whether users can perceive the depth and size of virtual objects better on a Fishbowl VR display compared to a flat FTVR display.

5.1 Introduction

Traditional FTVR displays use a high resolution desktop flat display coupled with head-tracking to generate view-dependent imagery. The latest advance of FTVR uses multiple mini-projectors to make it scalable and high resolution with uniformly spaced pixels [157]. Arranged inside a geometric shape, such as a box [127], a cylinder [77] or a sphere [157], it can render 3D scenes by coupling perspective to the user's viewpoint with wide viewing angles. Among different shapes of displays, the spherical shape is of particular interest as it provides seamless and unobstructed views from all angles. Used as a 2D surface display, previous studies have identified several unique properties of the spherical screen when rendering images directly on its surface [10]. In addition to the unobstructed 360-degree field of view, the continuous nature

of the curved shape provides smooth transition both horizontally and vertically [10]. The enclosing nature of the sphere inherently defines the virtual space within the screen. In contrast to flat displays, spherical displays do not have a fixed "sweet spot" of viewing as the normal of the screen is always parallel to the viewing direction regardless of the user position.

These properties are ideal for creating FTVR visualizations. While a number of spherical display prototypes have been reported [10, 16, 133], it is not clear whether the spherical shape factor can effectively improve performance in the 3D space when used for FTVR compared to the traditional planar FTVR display. Hence, in this work, we investigate whether viewers can benefit from the spherical shape factor with additional depth cues such as stereopsis and motion parallax. Particularly, we are interested that when coupled with head-tracking, does the compelling 3D effect provided by Fishbowl VR displays improve viewer's spatial perception.

The spatial perception of the screen itself and the projected imagery on the screen are inherently conflated. Visual cues of rendered objects are directly depicted on the screen surface. The physical representation of the screen shape provides the connection between the real world and the virtual environment (VE). The screen has a geometric indication that conveys the spatial locations of virtual objects situated in the real world. This implication may influence a viewer's spatial perception. Depth perception, for example, is of particular interest. When displaying virtual objects, the screen inherently serves as a depth reference that indicates the relative distance between virtual objects and the viewer. The enclosing nature of the spherical screen might anchor the depth perception differently from the planar screen. In a similar way, size perception is another example that could be affected by the shape of the screen. Depth and size perception are intimately related as objects appear smaller when they are further away. Size has been used as an indirect measure of the depth [73] under the assumption of the *size constancy*. Size constancy refers to the ability to perceive the same size of an object regardless of its distance from the observer even though the retinal size of the object changes [140]. When perceiving a virtual object, to some degree, we can judge the size either based on its spatial size as if it exists in the 3D space, or the 2D size on the screen [76, 142]. Perhaps the screen factor may affect which mode the we see in the VE.

Despite substantial perception studies in VEs, few empirical studies have been done regarding the shape factor of the screen. To investigate this question, we conducted an empirical user study to assess the spatial perception with the Fishbowl VR display in comparison to a flat FTVR display using a depth ranking task and a size matching task. The result of the study

helps us to understand the potential benefits and challenges of Fishbowl VR displays as well as provides directions for future 3D applications suitable for spherical displays.

5.2 Depth and Size Perception Study

We conducted a user study to evaluate user's perception on the depth and size information when visualizing the 3D stimuli on a Fishbowl VR display in comparison to a flat FTVR display. The purpose of this study is to evaluate the capability of the Fishbowl VR display to provide effective spatial perceptions relative to the traditional flat screen. The spherical shape is of particular interest due to its enclosing shape, borderless screen and consistent curved surface.

5.2.1 Apparatus

We used the 12 inch Fishbowl VR prototype described in Section 4.1 to conduct the experiment. To minimize differences in the apparatus, we utilized a projector (Asus P2B [53]) with the same 1024x768 pixel resolution and 120 Hz frame rate (60 Hz each eye) to back-project on a planar screen to build a flat FTVR display. It has the physical screen size of 36cm x 27cm, which has a similar visible screen area with the spherical screen of 30 cm diameter. Both the spherical and flat FTVR display are coupled with Polhemus tracking systems which track viewpoints at the update rate of 60 Hz with the accuracy of 1 mm. They both provide stereopsis and perspective-correction so users can physically move their viewpoint while visualizing stereo images. The stereo images are generated by 120 Hz projectors and synchronized with shutter glasses. The flat FTVR display experimental setup is shown in Figure 5.1.

A Unity [135] application was developed for the experiment to render the stimuli as describe below and record the answers as well as the head movements of participants. A virtual depth axis with 12 tick marks and a virtual pool ball was rendered in the Unity application. We used a standard international billiard ball with the diameter of 57.15 mm as shown in Figure 5.2. We used a post-experiment questionnaire and a demographic questionnaire in the study which can be found in Appendix C.1.

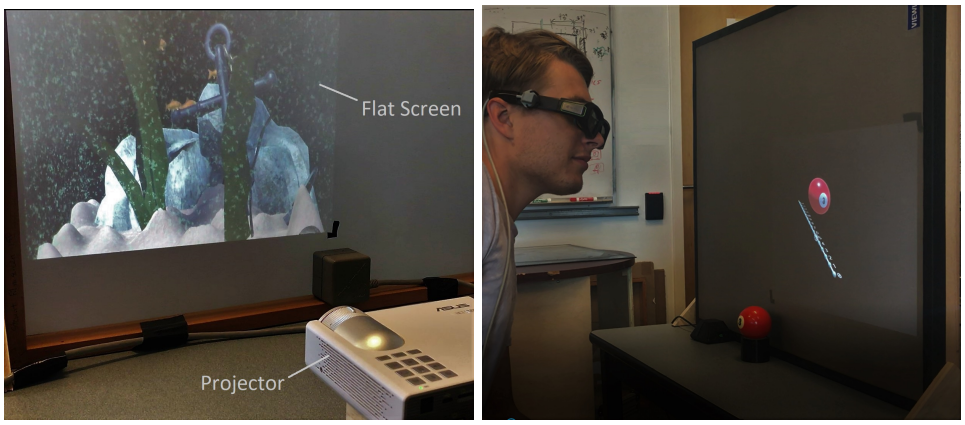


Figure 5.1: Flat FTVR display experimental setup in this study. (left) Mini-projector back-projects onto a planar screen with the projection size of 36cm x 27cm. (right) Flat FTVR display shows stimulus to a user in the study. The projector has the same resolution and brightness as ones used in the Fishbowl VR display.

5.2.2 Tasks

As our goal is to investigate whether a spherical display surface can improve spatial perception with the FTVR display, we measured the exocentric depth perception as well as size perception. Depth and size perception are tightly coupled by nature. They are important as various 3D interactions such as selection, translation and scaling depend on them. We evaluated depth and size perception using a depth-ranking task [45] and a size-matching task [76] respectively. We used these two tasks due to the effectiveness demonstrated from previous researchers' results as discussed in Section 2.2.4. Our present study seeks to provide evidence on spatial perception with a better understanding of the nature of spherical displays. It also provides some insight on size perception with FTVR displays as there has been few size perception studies with FTVR displays.

Task 1: Depth Ranking Participants were required to rate the depth of a pool ball floating in a virtual scene. The task is similar to the tasks previously used by Grossman [45] and Benko [9] in evaluating the volumetric display and spatial augmented reality respectively. In our implementation, a depth axis is presented with 12 tick marks uniformly distributed on the axis labeled as 0-11 with $2cm$ spacing. The depth axis and tick marks are

virtually placed behind the screen with the first tick mark located at the front screen. A pool ball was drawn floating above the axis as shown in Figure 5.2.

The task of participants was to rate the depth of the ball by selecting a tick mark at the same depth. In our experiment, stimuli balls are presented in three possible regions: near (tick 2-3), mid (tick 5-7) and far (tick 9-10). For each trial, the ball will appear in any of the three regions.

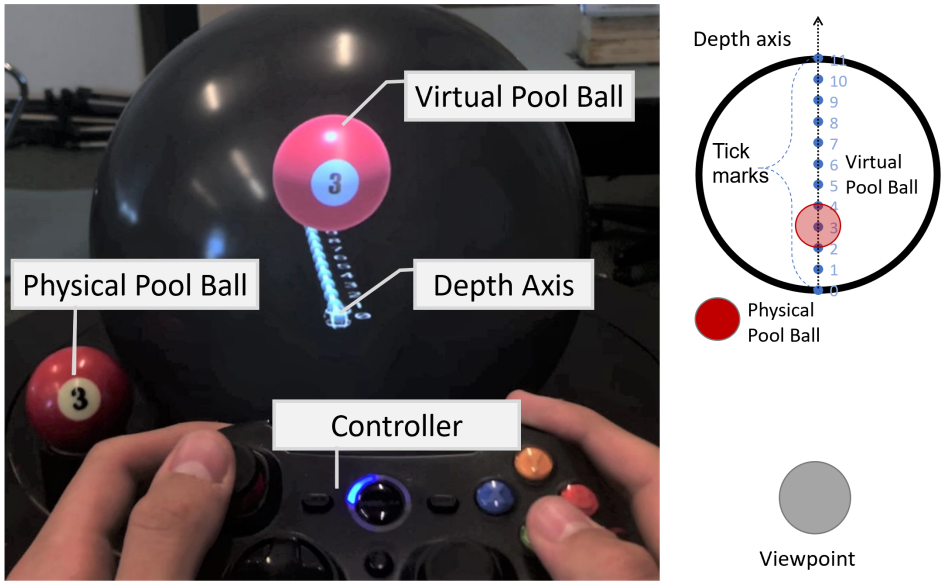


Figure 5.2: (left) Stimulus used in the experiment, including a virtual pool ball, a depth axis and a physical pool ball. The physical pool ball is a standard international billiard ball with the diameter of 57.15mm. In the depth ranking task, participants were required to rate the depth of the virtual ball relative to the depth axis. In the size matching task, they adjusted the size of the virtual ball to match with the physical ball. (right) A top-down view of the stimulus used in the experiment.

Task 2: Size Matching As shown in Figure 5.2, participants were presented with both a real pool ball and a virtual pool ball in the display. The participants' task was to adjust the size of the virtual ball until they perceived the size of the virtual ball to be identical to the real ball. This task is similar to the tasks previously used by Kenyon [76] and Kelly [123]. In our implementation, the real ball with the diameter of 57.15mm was placed

3cm in front of the display. Participants can touch and grasp the real ball to perceive its size while adjusting the virtual ball to match to the real ball using a controller. Similar to the depth ranking task, the virtual pool ball will appear in one of the three regions (near, mid and far) with different initial sizes.

We chose a pool ball as a visual stimuli due to its isotropic shape. This is to encourage participants to perceive and match the size based on the volume in 3D rather than a 1D feature such as the length of an edge.

5.2.3 Participants

Sixteen participants (11 male and 5 female) from University of British Columbia were recruited to participate in the study with compensation. Ethic was approved by the University of British Columbia Behavioural Research Ethics Board (H08-03005). The average age of all participants was 29 years old from 18 to 35 years old. In a post-experiment demographic questionnaire, we asked participants to rate their expertise with 3D interfaces such as 3D video games and 3D design on a scale from 1 ("novice") to 4 ("expert"). Their answers showed an average score of 1.75 ranging from 1 to 4. We also asked participants to provide their previous experience and usage with Virtual Reality with a scale from 1 ("never") to 4 ("regularly"). The average score of all participants was 1.88 between 1 and 4.

5.2.4 Procedure

Each participant was first given a brief introduction to the two displays. They started by filling the consent form. Then they were instructed to sit facing the display with the viewing distance of 0.7m in both display conditions. Participants could move their viewpoint by leaning the upper body to the left/right facing the screen with the angle approximately 150 degrees. Their movement was limited to the volume by requiring them to stay seated at a fixed spot facing the display in order to prevent them from taking viewpoints which would trivialize the depth ranking task. They physically moved from one display to another after completing one display condition. In both tasks, participants received verbal instructions to understand the procedure. They practiced three trials prior to the formal trials per task per display condition.

In the depth ranking task, they conducted 12 trials in a row per display condition, with four of the near, mid and far viewing distances respectively. The sequence of the trials is randomized. All tested depths were within 1

meter. Participants controlled a cursor in the scene to choose the tick mark using a controller. They pressed buttons on the controller to confirm the result and proceed to the next trial. In total, each participant conducted 24 formal trials plus 6 practice trials. It took about 20 min to complete the depth ranking task.

In the size matching task, they also conducted 12 trials per display condition, with four of the near, mid and far viewing distances respectively. Each viewing distance includes four initial sizes, two larger and two smaller ones than the real ball with a random scale between 1.08 and 1.26 for large balls, as well as the scale between 0.73 and 0.91 for small balls, respectively. Participants adjusted the size of the virtual ball using a controller and pressed buttons to confirm the result once they perceived the size of the virtual ball was identical to the real ball. For the size matching task, each participant conducted 24 formal trials plus 6 practice trials across two display conditions. It took about 20 min to complete the size matching task.

At the end of each display condition, participants were presented with a questionnaire. For each task, we presented three Likert-scale questions to participants and asked them to rate each with a number in the range -2 ("totally disagree") to 2 ("totally agree"). The three questions addressed the easiness, intuitiveness and confidence when performing the task with the display. After completing each task (2) per display condition (2), participants answered the three questions(3), resulting in a total of twelve questions (2x2x3). Once they completed the entire experiment, they answered a demographic questionnaire and completed the previous post-experiment questionnaire with additional two Likert-scale questions on the overall realism and enjoyment per display condition. Finally, they were asked to rate their preferences and specify reasons between displays in both tasks. Each participant performed all parts of the experiment in one session. The entire session took about 50 min.

5.2.5 Experiment Design

The study used within-subjects design to evaluate performance across the two display conditions using two tasks. The two display conditions represent the two display devices of a flat FTVR and a Fishbowl VR display. The order of display conditions was counterbalanced to reduce the effects of ordering. In each display condition, participant performed the experiment by completing the two tasks and then filled a post-experiment questionnaire. The two tasks always appeared in the same order. We described the independent and dependent variables in this section for the two tasks.

Task 1: Depth Ranking We investigated two independent variables as *Display* and *Distance* with two levels (Sphere/Flat) for *Display* and three levels (Near/Mid/Far) for *Distance* respectively. Subject performance was evaluated quantitatively based on the dependent variable $ErrorMagnitude$, defined as the average absolute distance between the reported answer and the correct answer. *Distance* is included as a factor since we expected the $ErrorMagnitude$ increases as the viewing distance increases.

Task 2: Size Matching We investigated three independent variables as *Display*, *Distance* and $InitialSize$ with two levels (Sphere/Flat), three levels (Near/Mid/Far) and two levels (Large/Small) respectively. We include $InitialSize$ as a factor because size judgments showed an anchoring effect, by which the adjusted ball size on a given trial tended to be biased toward the initial ball size set in the beginning of that trial [123]. Similar to previous size perception studies [76, 91], subject performance was evaluated based on the measure $SizeRatio$ and $AbsoluteError$. According to [76], $SizeRatio$ represents the relative size of the virtual ball compared to the size of the real ball:

$$SizeRatio = \frac{BallSizeSetByParticipant}{CorrectRealBallSize}$$

The numerator is the adjusted ball size reported by participants as identical to the real ball size, which varied in each trial. The denominator is the size of real ball fixed at 57.15mm. Ideally, $SizeRatio$ would be 1 if participants adjust the ball size without any error to match the real ball.

$SizeRatio$ can indicate the under or over-estimation of the size. It is effective to investigate the influence of $InitialSize$. However, it is not a good measurement of the accuracy since $SizeRatio$ can be above or below 1 so that the average $SizeRatio$ can be 1 while the absolute mean error is much greater than zero. As a result, $AbsoluteError$ is another dependent variable calculated to examine the deviation between the ideal result and $SizeRatio$:

$$AbsoluteError = |1 - SizeRatio|$$

For both tasks, we tracked the head movement and computed the *normalized head movement* , defined as the amount of head motion in meters per second per trial. As the viewing duration were different for trials and participants, we normalized the amount of head movement by dividing it with the time spent per trial.

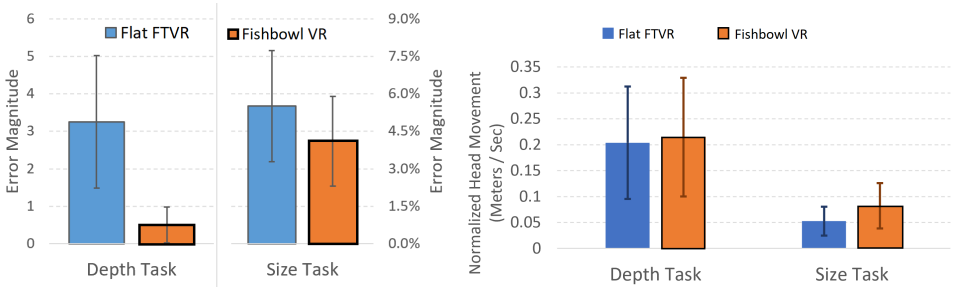


Figure 5.3: (left) Error magnitudes with means and 95% confidence interval (CI) for the depth ranking task and size matching task. (right) Normalized head movement with means and 95% CI for the two tasks, computed as the amount of head movement divided by the time spent per trial.

5.2.6 Result

Task 1: Depth Ranking

Repeated measures two-way ANOVA (2 *Display* x 3 *Distance*) was carried out on the *ErrorMagnitude*. Results revealed an effect of *Display* ($F(1,15) = 11.00, p < 0.01$), but not of *Distance* ($F(2,30) = 0.287, p = 0.753$) nor was the interaction ($F(2,30) = 0.448, p = 0.643$). The Fishbowl VR display ($M = 0.50, SE = 0.22$) has significantly lower *ErrorMagnitude* than the flat FTVR display ($M = 3.25, SE = 0.83$). These results are shown in Figure 5.3 (left). A complete table of results can be found in Appendix D Figure D.1.

ANCOVA on the participants' expertise on VR and 3D interfaces was carried out on the *ErrorMagnitude* with the *Display* as independent factor and their expertise on VR and 3D interfaces as covariate. The results did not show significant difference on the covariate of VR expertise ($F(1,29) = 0.297, p = 0.590$) or 3D interface expertise ($F(1,29) = 1.89, p = 0.180$) on *ErrorMagnitude*.

Task 2: Size Matching

Repeated measures three-way ANOVA (2 *Display* x 3 *Distance* x 2 *InitialSize*) was carried out on the *AbsoluteError* with factors of *Display*, *Distance* and *InitialSize*. Results revealed an effect of *Display* ($F(1,15) = 5.132, p < 0.05$), but not of *Distance* ($F(2,30) = 2.093, p = 0.141$), *InitialSize* ($F(1,15) = 0.546, p = 0.472$), nor any interactions. The Fishbowl VR display ($M = 0.041, SE = 0.008$) has significantly lower *AbsoluteError* than the flat

FTVR display ($M = 0.055, SE = 0.010$). These results are shown in Figure 5.3 (left). A complete table of results can be found in Appendix D Figure D.2.

Repeated measures three-way ANOVA (2 *Display* x 3 *Distance* x 2 *InitialSize*) was carried out on the *SizeRatio* with factors of *Display*, *Distance* and *InitialSize*. Main effects of *Display*, *Distance* and *InitialSize* were found as *Display* ($F(1, 15) = 4.88, p < 0.05$), *Distance* ($F(2, 30) = 5.13, p < 0.05$) and *InitialSize* ($F(1, 15) = 36.64, p < 0.001$). Results revealed two-way interaction effects between *Display* and *Distance* ($F(2, 30) = 5.13, p < 0.05$), *Display* and *InitialSize* ($F(1, 15) = 38.71, p < 0.001$), as well as *Distance* and *InitialSize* ($F(2, 30) = 12.97, p < 0.001$), but not on the three-way interaction ($F(2, 30) = 1.47, p = 0.245$). A complete table of results can be found in Appendix D Figure D.3.

Post-hoc pairwise t-test with Bonferroni correction [2] for the interaction *Display* x *Distance* shows an effect of *Distance* when *Display* is Flat between *Distance* Far and Near ($t = 4.23, p < 0.01$), as well as Far and Mid ($t = 2.93, p < 0.05$), but not when *Display* is Sphere. For the interaction *Display* x *InitialSize*, there is an effect of *Display* when the *InitialSize* is Small ($t = 3.834, p < 0.01$), but not when the *InitialSize* is Large. *InitialSize* has an effect in both Flat display ($t = 7.835, p < 0.001$) and Sphere display ($t = 2.926, p < .05$). For the interaction *Distance* x *InitialSize*, there is an effect of *InitialSize* when *Distance* is Far ($t = 7.468, p < 0.001$) and Mid ($t = 4.859, p < 0.001$), but not when *Distance* is Near. *Distance* has an effect when *InitialSize* is Large ($t = 5.473, p < 0.001$) but not when *InitialSize* is Small.

ANCOVA on the participants' expertise on VR and 3D interfaces was carried out on the *AbsoluteError* and *SizeRatio* with the *Display* as independent factor and their expertise on VR and 3D interfaces as covariate. The results did not show significant difference on the covariate of VR expertise ($F(1, 29) = 0.659, p = 0.423$) on *AbsoluteError* or ($F(1, 29) = 0.829, p = 0.370$) on *SizeRatio*, as well as 3D interface expertise ($F(1, 29) = 0.0369, p = 0.849$) on *AbsoluteError* or ($F(1, 29) = 2.223, p = 0.147$) on *SizeRatio*.

[2] As the two-way interaction effect *Distance* x *InitialSize* was not relevant to differences between display conditions, it was not analyzed further in the post-hoc simple effect anaylsis.

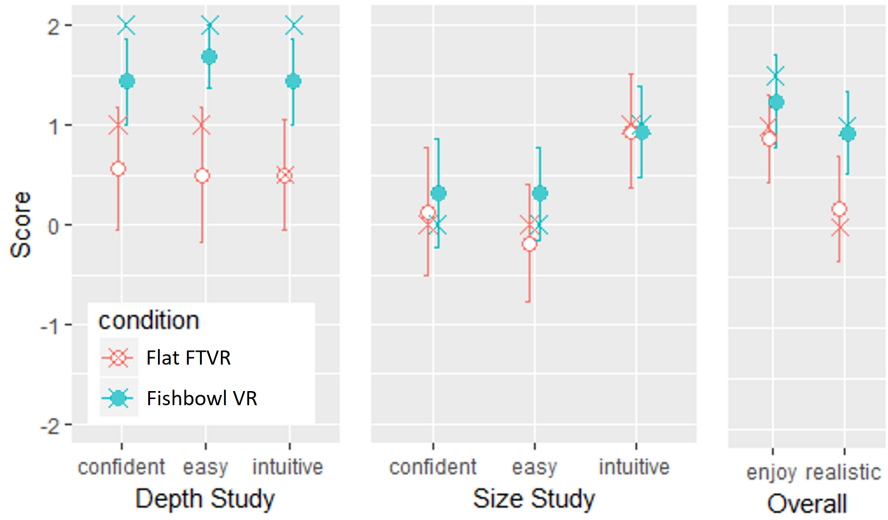


Figure 5.4: Participants' ratings with means (circle), medians (cross) and 95% CI, from -2 "totally disagree" to 2 "totally agree". They reported ratings on the confidence, easiness and intuitiveness for the depth ranking and size matching task, as well as the overall impression regarding the enjoyment and realism per display condition.

Head Movement Analysis

We conducted pairwise t-test on the *normalized head movement* between the Fishbowl VR and flat FTVR display. For the depth ranking task, a t-test indicated that there is no significant difference on the *normalized head movement* ($t = -1.02, p = 0.322$). For the size matching task, the Fishbowl VR display has significantly more head movement than the flat FTVR display ($t = -2.76, p < 0.05$) as shown in Figure 5.3 (right). A complete table of results can be found in Appendix D Figure D.4 and D.5.

Post-Experiment Questionnaire

A Wilcoxon Signed Rank Test was performed on the Likert-scale questions. There were significant differences on the easiness ($W = 78, p < 0.01$), intuitiveness ($W = 62, p < 0.01$) and confidence ($W = 45, p < 0.01$) for the depth-ranking task, but not for the size-matching task. There were also significant differences on the overall impression of realism ($W = 72.5, p < 0.01$), and enjoyment ($W = 15, p < 0.05$) between display conditions. The results

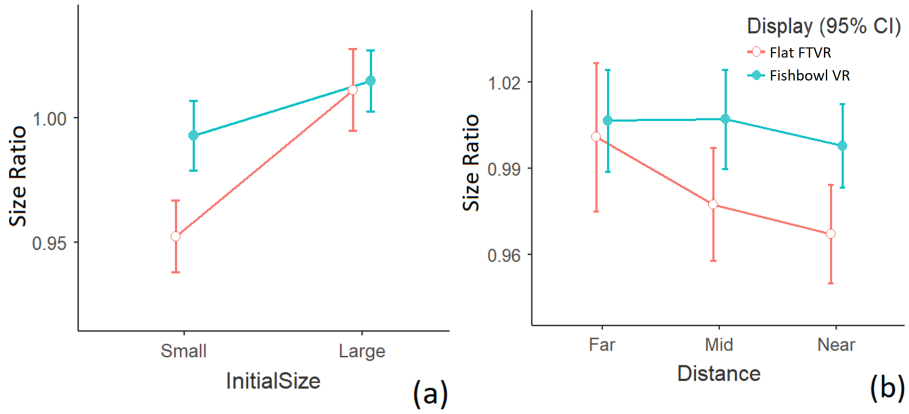


Figure 5.5: The two-way interaction effects between (a) *Display* x *InitialSize*, as well as (b) *Display* x *Distance*. The size perception is less influenced by *Distance* as well as *InitialSize* on the Fishbowl VR display than on the flat FTVR display.

of questionnaire are summarized in Figure 5.4. For the preference data, 75% participants believed they performed better on the Fishbowl VR display for both tasks. There were 93% and 87.5% participants preferring the Fishbowl VR display for the depth-ranking and size-matching task respectively. A Chi-Square analysis shows significant differences on the performance ($\chi^2 = 14, p < 0.001$) and preference ($\chi^2 = 12.25, p < 0.001$) for the depth-ranking task, as well as the performance ($\chi^2 = 12.875, p < 0.01$) and preference ($\chi^2 = 11.267, p < 0.001$) for the size-matching task. A complete table of results can be found in Appendix D Figure D.6.

Crossover between Depth and Size

Because depth and size perception are tightly coupled by nature, we seek to investigate whether the under/over perception occurs on both size and depth judgments in two display conditions. On the flat FTVR display, 86.5% errors were underestimating the depth and 83% errors were overestimating the size. On the Fishbowl VR display, 50% errors are caused by underestimating the depth and 67% errors are caused by underestimating the size. In short, participants tend to underestimate the depth and overestimate the size on the flat FTVR display; on the Fishbowl VR display, there is no clear sign as whether they under or over estimated the depth and size. As as result, we did not find valuable crossovers between the two tasks.

5.3 Discussion and Limitations

The Fishbowl VR display shows superiority in both tasks compared to the flat display. In the depth ranking task, participants performed much better on the Fishbowl VR display with the *ErrorMagnitude* of 0.5 in comparison to 3.25 on the flat display, as shown in Figure 5.3(left), which yielded a depth accuracy of $1cm$ on the spherical display and $6.5cm$ on the flat display. The difference of performance is consistent with the results of the questionnaire. Participants reported significantly higher scores in the spherical condition on confidence, easiness and intuitiveness as shown in Figure 5.4.

In the size matching task, although the difference of performance is statistically significant between display conditions, the improvement was marginal with the *AbsoluteError* of 4.1% in the spherical condition and 5.5% in the flat condition, which yielded a size accuracy of $2.34mm$ on the spherical display and $3.14mm$ on the flat display. Inconsistent with the performance data, participants rated similar scores over the two displays in the questionnaire. No significant difference has been found on confidence, easiness or intuitiveness. This indicates the spherical shape did not show significant superiority on the accuracy in the size study as much as it showed in the depth study. However, we found an interesting interaction effect on the dependent variable *SizeRatio*, implying that the strength of the spherical shape in the size study may not be reflected as higher accuracy but rather as better consistency, which we'll discuss further in this section.

5.3.1 Size Constancy

While the difference of *AbsoluteError* in the size study is relatively small, the interaction effect on the *SizeRatio* between *Display* and *Distance* shows that *Distance* influenced user performances differently in the two display conditions. The different slopes in Figure 5.5 (b) indicates that the size perception is less influenced by *Distance* on the spherical display than on the flat display. When matching the balls with different *Distance*s, participants had better task precision in the spherical condition.

This interaction effect has a better interpretation as *size-constancy* [76]. Failure to preserve size-constancy makes an observer judge size based on visual angle so that the perceived size of an object shrinks with increasing viewing distance. In our study, if size-constancy is preserved and dominates across different distances, the slope is zero with the *SizeRatio* of 1 for all stimulus in Figure 5.6. To analyze size-constancy, we fit the *SizeRatio* data with a regression line over different distances for each display. As shown

in Figure 5.6, the positive linear relationship between *SizeRatio* and the viewing distance indicates that participants have a greater tendency of estimating based on the visual angle rather than preserving size-constancy. Paired t-test on slopes based on the standard error of regression models[3] shows the regression slope with the flat display is significantly larger than the slope with the spherical display ($t = 2.85, p < 0.05$). This indicates that perceptual judgments of the size on the spherical display better preserved the size-constancy in comparison to the judgment on the flat screen. Participants have a greater tendency to base their judgments on the visual angle when using flat displays.

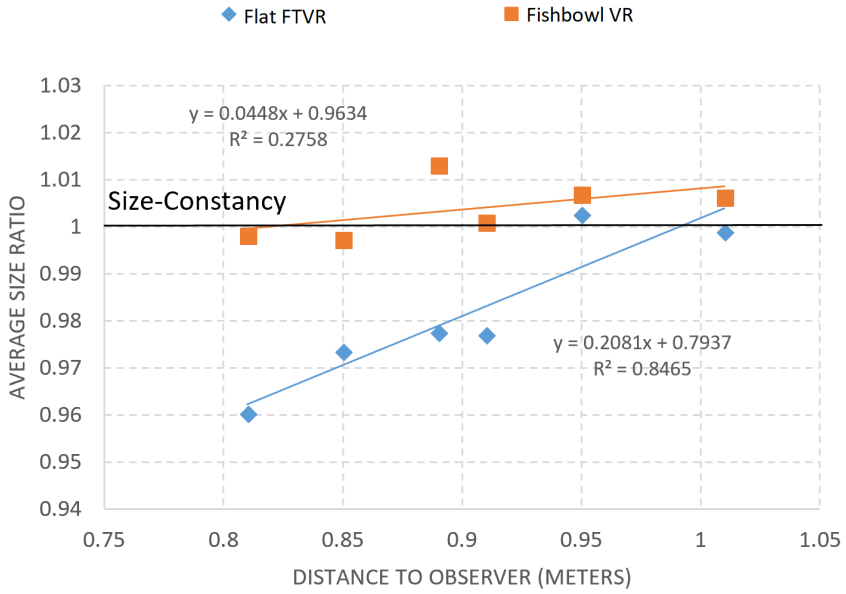


Figure 5.6: Linear regression of *SizeRatio* versus the viewing distance for all participants in two display conditions with the slope of regression indicating size-constancy. If the size-constancy is completely met, *SizeRatio* will be a constant of 1 with the slope of zero. The rise in the size-ratio with the increasing distance indicates that participants are performing more like visual angle than size-constancy.

Another interaction effect *Display* x *InitialSize* also shows interesting performance consistency as a hysteresis effect of the *InitialSize*. As illustrated in Figure 5.5 (a), the distinct slopes show the hysteresis effect is different in the two display conditions. The size perception on the spherical

display appears to be less influenced by the *InitialSize* compared to the flat display. Consequently, when matching the balls with different *InitialSize*s, participants had better task precision in the spherical display condition.

When similar size-judgment tasks were conducted in the real world, a previous study found that participants were less accurate at perceiving the size in virtual world, and the effect of the *Distance* factor was greater in the virtual world than in the real world [128]. Notably, in our study the spherical screen provides task performance closer to the performance in the real environment with better accuracy and consistency compared to a flat screen. Perhaps this indicates that the spherical screen may provide better visualization in a way closer to the real world. Consistent with the task performance, we found that the rate of the realism on the spherical display is significantly higher than the flat display ($W = 72.5, p < 0.01$) as shown in Figure 5.4. Eight out of sixteen participants further commented that the pool ball looks more 3D-like and realistic on the spherical display when asked about reasons of their preferences on displays. The level of realism may influence the visual cues that participants take to infer the position and size of stimulus. In our study, there are multiple visual cues, including 2D cues, such as the projected ball's 2D size on the screen surface, as well as 3D cues, such as stereopsis and motion parallax. When a user performs the tasks purely based on 2D cues, the performance will be greatly influenced by *Distance* and *InitialSize*. However, if they use more 3D cues, this will lead to a more consistent performance across different *Distance* and *InitialSize*. When participants felt the objects appear more realistic when viewing the stimulus on a particular display, it is more likely they counted more on the 3D cues rather than 2D cues. This is also referred by Ware as the "duality of depth perception" when people perceive pictures of objects [142], which is further investigated in Chapter 6. For 3D displays, it is necessary to make sure that the screen effectively conveys spatial information so that 2D cues will not overrule 3D cues.

5.3.2 Visibility

Due to the nature of the spherical screen, viewers are not able to see the entire curved screen simultaneously. So we made the screen size of the flat FTVR display approximately the same as the visible area of the spherical screen. This is also to reduce the influence of screen size on the perception of stimulus size. When rendering an object with the same distance to the viewer, the object can be cut-off by the screen edge on the flat display if the viewpoint moves to the side, but not on spherical display as it is borderless.

As illustrated in Figure 5.7, the further away the object is to the viewer, the smaller the Range of Visibility (RoV) is before the object gets cut-off. This results in a distance-dependent RoV on the flat display but a constant RoV on the spherical display. In our study, viewpoint movements are constrained since participants are required to sit on a fixed chair when viewing the stimuli. The angle of head movements facing the screen is about 150 degree in both display conditions so they could not see the side view. However, the limited RoV for objects far away may still discourage participants to move their viewpoints so the motion parallax cue has not been exploited on the flat display.

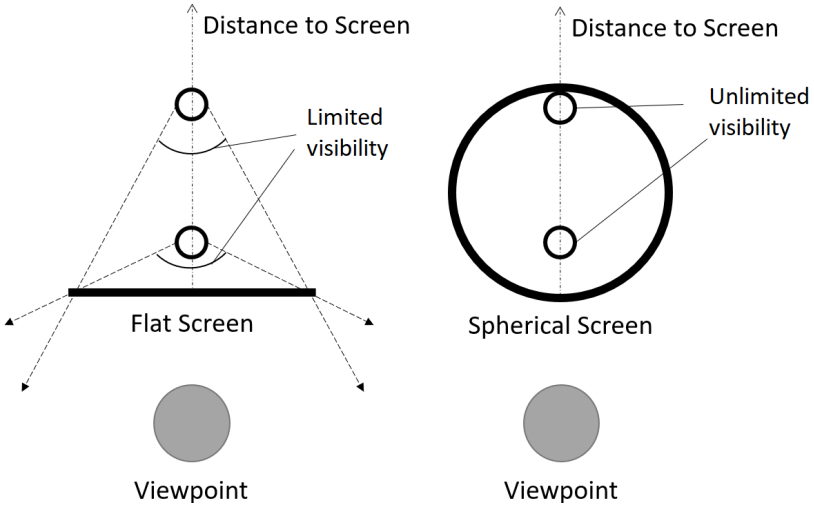


Figure 5.7: Limited visibility on the flat FTVR display for displaying objects behind the screen. The object is cut-off by the screen edge of flat FTVR display if the viewpoint moves to the side, but not on the Fishbowl VR display. The further away the object is to the viewer, the smaller the visibility is before the object gets cut-off.

In particular, the RoV has a large impact on the performance of the depth ranking task. As shown in Figure 5.3 (right), the head movement was much more frequent in the depth-ranking task than the size-matching task. It is a task evaluating the exocentric depth in which participants relied heavily on head movements. If the ball gets cut-off by the screen edge, the information is then lost, making it harder to determine the correct tick. Eight out of sixteen participants commented that they got better perspec-

tives on the spherical display when asked for reasons of their preferences between displays in the questionnaire. Participants may obtain more information on the relative distance by making use of the spherical borderless feature. It would be interesting to further explore this issue and compare results with large flat screens so that the stimuli will not be cut-off by the screen edge.

5.3.3 Head Movement

As shown in Figure 5.8, head movements were mostly side-to-side in both tasks to provide better side views. The movements appear to be more constrained vertically in the flat condition than the spherical condition, likely due to the spherical continuously varying surface which provides visible contents changed by the viewpoint heights. Their movements appeared to form a curved path around the virtual ball to keep a consistent viewing distance with respect to the stimuli target. In the spherical condition, this results in a nearly constant distance to the screen surface while the screen distance varies in the flat condition due to the nature of the screen shape. This difference might influence users' performance if we consider the vergence/focus conflict: viewers converge toward the virtual object but focus eyes on the display surface. As discussed by Bruder [20], the conflict might become more evident if viewers get closer to the screen surface while moving their head. In our study, while participants are constrained to sit, we did not limit the type of their head motion. So if they naturally moved their head following a curvature when trying to see different sides of the object, they could keep a relatively constant screen distance on the spherical display. In this case, the spherical screen shape seems to reduce this conflict, though future study is required to investigate on that.

In the depth-ranking task, participants relied heavily on head movements to find the closest tick to the stimuli ball in both display conditions. While participants actively moved viewpoints in both conditions as shown in Figure 5.8 (a)(b), we did not find a significant difference on the *normalized head movement* shown in Figure 5.3(right). As their movements were constrained on a chair, it is likely that they exploited the possible movements while staying on the chair in both conditions, resulting in similar amount of head movements in two display conditions.

In the size-matching task, participants showed less head movements in both conditions. While their movements were equally constrained in both conditions, they moved more frequently in the spherical condition than the flat condition ($t = -2.76, p < 0.05$) as shown in Figure 5.8 (c)(d). In our

observation, participants were more likely to stay at the same viewpoint in the flat condition, while they frequently moved viewpoints to compare the size between the real ball and virtual ball in the spherical condition, implying that they took the motion parallax cue more with the spherical screen. This is also consistent with Ponto's shape-matching study in which they found that subjects produced smaller errors when they were allowed to move freely in the environment [103]. As moving viewpoints is a natural behavior in the real world, this indicates participants felt more natural and realistic when viewing the 3D scene on the spherical screen.

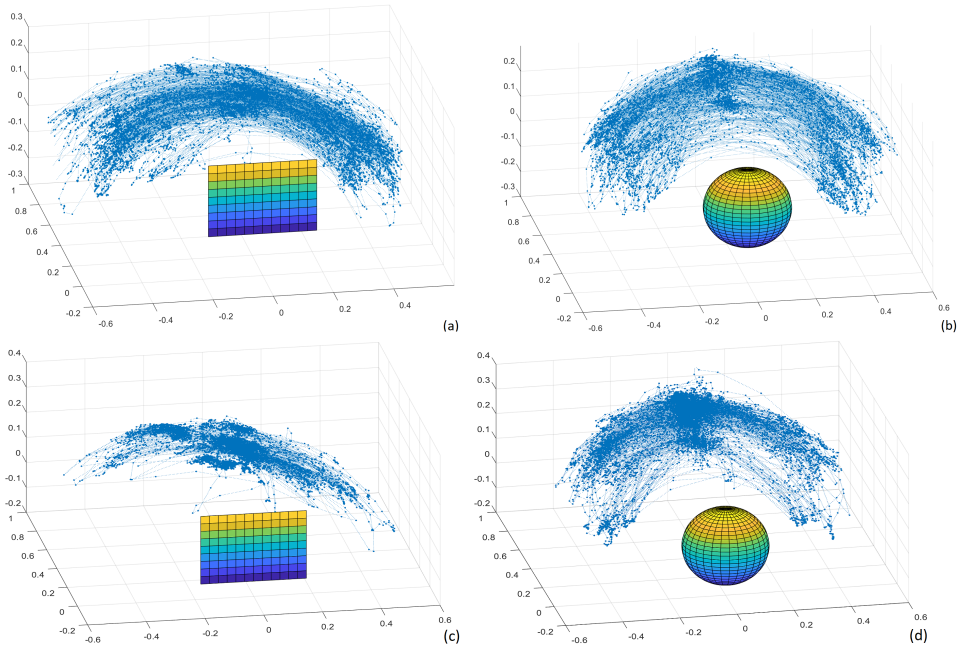


Figure 5.8: Head movement visualization of all participants in the depth study with the (a) flat FTVR display and (b) Fishbowl VR display, as well as in the size study with the (c) flat FTVR display and (d) Fishbowl VR display. Blue dots represent the viewpoint positions where participants were viewing the stimulus throughout all trials.

While participants were instructed to sit on the chair, we did not observe them trying to stand up and move around the spherical display in the size study. As they were only using part of the spherical display space shown from their head movement, it also implies that a whole sphere screen may not be necessary for the size-matching task. A hemisphere screen facing the

viewer could potentially be sufficient to support the range of head motion, though future work is required to find the minimum area of the spherical screen without compromising the task performance.

The head movements finding diverged from the Kenyon's size-constancy study [76] in which they found side-to-side head movements were small or absent. One possible reason is the difference between the experiment stimulus used. In their experiment, they used a Coke bottle which had a vertical size feature. The side-to-side head movement is less helpful when participants tried to adjust the height of the bottle. This also indicates that the results of size perception experiments might be sensitive to different stimulus, which will be discussed in Section 5.3.4.

5.3.4 Indications to FTVR Studies and Applications

As there have been few size perception studies with FTVR displays, our work seeks to provide insights on the size perception with FTVR displays. Using the same size-matching task, we compared our data to the Kenyon's study on the CAVE display [76] with the Near viewing distance of $1.07m$ in their sparse environment condition. As shown in Table 5.1, both $AbsoluteError$ and size-constancy[3] appear to be better on FTVR displays, indicating FTVR displays might be promising to provide better size perception in the VE.

While the findings are encouraging, we should be aware of several aspects of the experiment which the results could be potentially sensitive to. Identifying them would be helpful for designing future related experiments. As Ponto mentioned in [103], the accuracy of the perceptual calibration could influence the size perception. In our study, we used a perceptual calibration approach to determine viewing parameters of each eye. In practice, we noticed that inaccurate calibration could possibly make stationary objects appear to move when the viewer changes the position, similar to the *floating effect* mentioned in Ponto's study [103]. Hence it is important to employ a reliable and accurate perceptual calibration method for this type of study. Another factor that may influence the result is the visual stimuli. We chose a spherical stimuli in our study due to its isotropic property. It remains to be a question that whether a stimuli with different shapes such as a cube would reproduce similar results. Future controlled experiments are needed to compare the size perception with different stimulus on different shapes of screens. Using the methodology of this study, we plan to expand the shape factor from the spherical to cubic and cylindrical displays. Further compar-

[3]The slope value of 0.36 is after converting the value of 0.11 from *feet* to *meters*.

Table 5.1: Comparison of the size perception on different 3D displays, including the CAVE display from existing work [76], and our work.

Metric	CAVE	Flat FTVR	Fishbowl VR
Absolute-Error	20%	5.5%	4.1%
Size-Constancy Slope	0.36	0.21	0.045

isons with immersive displays such as CAVE and HMD displays would also help us to understand the impact of display factors on the depth and size perception.

Realistic visual perception is crucial in 3D applications such as Computer Aided Design (CAD) systems as they need to accurately depict the size, shape and position of the object. Misperception of the distance and size of virtual models would decrease the performance of tasks ranging from 3D scaling to assembly in CAD. The ability to preserve size-constancy allows users to effectively position, scale and design prototypes in the virtual world. It is not just about making correct size judgments between two virtual entities, but also preserving consistent size perception between the real and virtual environment such that the perceived scale of a virtual model is consistent when it gets fabricated in the real world. In comparison to other 3D displays such as HMD, the FTVR display is promising to provide consistent perception as it allows users to see virtual contents situated in the real world. Hence we hope that our work on the distance and size perception would help to pave the way for 3D applications such as CAD in FTVR.

5.3.5 Limitations and Future Work

There are several limitations to the experiment presented in this chapter. We used a flat FTVR display as a baseline to evaluate the perception on the spherical display. It remains to be a question as how the performance would be compared to a real world stimuli as flat FTVR displays provided limited insight. In particular, the size-constancy has been demonstrated in the physical world with consistently reproducible results [76]. We expect the results would be better using a real world stimuli with future experiments evaluating the perception on the Fishbowl VR compared to a baseline based on real world stimulus.

Another limitations is that we measured exocentric depth between two virtual points rather than the egocentric depth. This is different from previous studies [74] in which they measured the egocentric depth. This is a

limitation as the exocentric distance measure may not directly reflect the egocentric distance which participants perceived when adjusting the size. It would be worthwhile further investigating the egocentric depth perception and comparing results with the exocentric depth. In addition, as participants performed the depth task prior to the size task, it is possible that their performance in the size task could be affected by their performance in the depth task, which makes the order of two tasks a potential confound variable. Future studies could consider separate the two tasks into two studies.

While the result indicates that the spherical shape factor improves the depth and size perception with its unique properties such as enclosure shape, it is not clear on how these properties influence the task performance individually. Future work is required to understand the mechanisms that may cause the difference. For example, a cubic FTVR display could be used to evaluate the importance of the borderless feature of the spherical screen while having a similar enclosure shape. In addition to its borderless feature, the spherical screen is a display with finite volume as Benko mentioned in [10]. Rendering virtual objects within the volume presents a metaphor as if they are inside a glass globe or display case. The question can be raised of whether rendering a virtual boundary of the back screen could influence the task performance as one may perceive the display ending on the other side. This may also provide additional motion parallax when moving the viewpoints.

Finally, in our study we provided various 3D cues such as motion parallax and stereopsis. It remains to be seen whether more subtle cues such as the vergence/focus conflict, would impact differently on curved screens compared to flat screens, as the spherical screen seems to present a larger range of physical surface depths to focus on. Therefore, the screen shape may help to limit this conflict depending on different visual stimulus. Future experiments designed to explore the relationships that may exist will help us better understand the impact of display technologies on visual cues presented to the user.

5.4 Summary

In this chapter, we have presented an empirical study evaluating spatial perception on a Fishbowl VR display using a depth-ranking and a size-matching task. We found that the Fishbowl VR display provides better task performance with depth accuracy of $1cm$ which yielded significantly less error compared to the accuracy of $6.5cm$ on the flat FTVR display in

the depth-ranking task. The performance of the size-matching task was also better with an accuracy of $2.3mm$ on the spherical display as compared to $3.1mm$ on the flat display. In addition, their performance is more consistent on the spherical display with better size-constancy. We believe the Fishbowl VR display is promising to improve users' performance in 3D tasks which rely on the depth and size perception in 3D applications such as CAD that may benefit from the spatial perception provided by the Fishbowl VR display.

Chapter 6

Perceptual Duality of FTVR Displays

In this previous chapter, we demonstrated the superiority of Fishbowl VR displays in improving spatial perception by depicting 3D objects on a spherical screen. Because users perceive 3D objects by looking at digital pixels on a 2D screen, there exists a perceptual duality between the on-screen pixels and the 3D percept. In this chapter, we investigated this perceptual duality by evaluating the influence of the on-screen imagery. Understanding the perceptual duality helps us to provide accurate perception of real-world objects depicted in the virtual environment and pave the way for 3D applications.

6.1 Introduction

FTVR displays provide compelling 3D experiences by rendering view dependent imagery on 2D screens. While users perceive a 3D object in space, they are actually looking at pixels on a 2D screen. The users know it is a 2D image, and also know it is a 3D object. These two concurrent understandings represent the perceptual duality between the object's pixels and the object in space, which could potentially cause perceptual inconsistency between the real and virtual world. This potential inconsistency can be illustrated as an example in Figure 6.1. When a 3D house is rendered in a Fishbowl VR display, the on-screen imagery is computed based on the viewer's position. Due to perspective projection, the on-screen imagery becomes smaller as the viewer moves closer, which contradicts to the reality as users expect a larger percept when getting closer to the object. This mismatch may cause perceptual inconsistency and interfere with the 3D experience.

The on-screen imagery is a visual stimuli that provides various visual cues to the viewer, including 3D cues (depth cues), such as motion parallax and binocular stereo, as well as 2D cues (on-screen cues), such as the position and size of the 2D projection on the screen. While the importance of these depth cues has been long appreciated in the practice of FTVR [36, 81, 142],

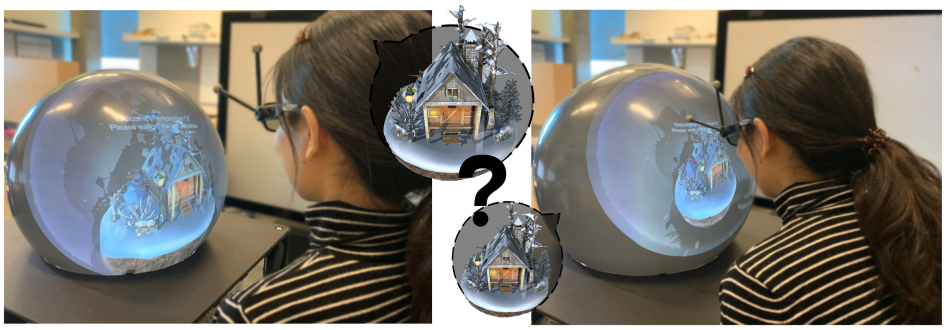


Figure 6.1: A user interacting with the Fishbowl VR display. Due to perspective projection, the on-screen imagery becomes smaller as the user gets closer, which contradicts to the reality as users expect a larger percept when getting closer to the object.

the influence of on-screen visual cues has been underestimated. As the virtual objects are depicted via the imagery, these on-screen cues might influence the perception of virtual objects. In our previous example, the size of the on-screen imagery interferes with the viewer's size perception in Figure 6.1. When judging the house's size, they may judge it based on the actual size of the house, or, the 2D size of its projection on the screen. In vision science, similar ambiguity has been previously referred as "duality of the depth perception in pictures" [44, 46, 142]. They found that adding depth cues in paintings can make one see in 3D rather than in 2D [142]; though their work focuses on static pictures. It is an open question whether these findings could be applied to screen-based 3D displays such as FTVR displays.

To investigate whether the on-screen imagery can influence users' perception, we conducted two experiments that measures users' size perception with different on-screen imagery on a Fishbowl VR display. We focus on the spherical form factor in this study because it has been most widely adopted for FTVR displays [36]. The first experiment evaluated the influence of on-screen imagery with and without the stereo cue. The second study investigate whether different sizes of the on-screen imagery caused by different projection matrices can affect the perceived size of virtual objects.

To the best of our knowledge, it is the first study to evaluate and provide insights on the perceptual duality with FTVR displays. While it is conducted with a Fishbowl VR display, the result applies to most screen-based 3D displays such as a CAVE [29]. Our study establishes a fundamental

limitation for a broad range of "screen-based" 3D displays. All screen-based 3D displays, like FTVR and CAVE, approximate holograms by rendering perspective on the surface with the assumption that if the perspective is geometrically correct, the perception will be correct. But our study shows that under some circumstances, this assumption may not hold and the approximated "hologram" causes perceptual bias with visual artefacts. Understanding the perceptual bias helps us to provide accurate perception and pave the way for 3D applications.

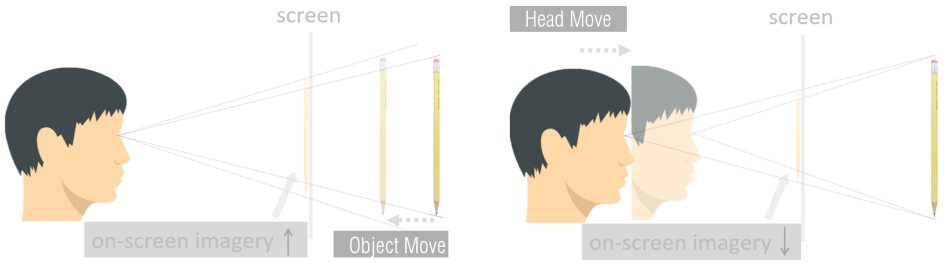


Figure 6.2: Illustration of *HeadMove* and *ObjectMove*. In *ObjectMove*, as object moves towards the user, the on-screen imagery gets larger. In *HeadMove*, as the user moves towards the object, the on-screen imagery gets smaller.

6.2 User Study 1: Influence of the On-screen Imagery

The purpose of this study is to investigate whether different sizes of the on-screen imagery can affect perceived size of virtual objects. As the perceived object size can be greatly affected by the visual angle on the retinal image [76], we defined two types of movements (*HeadMove* and *ObjectMove*) as illustrated in Figure 6.2 to provide consistent retinal images across conditions with different on-screen size as described below.

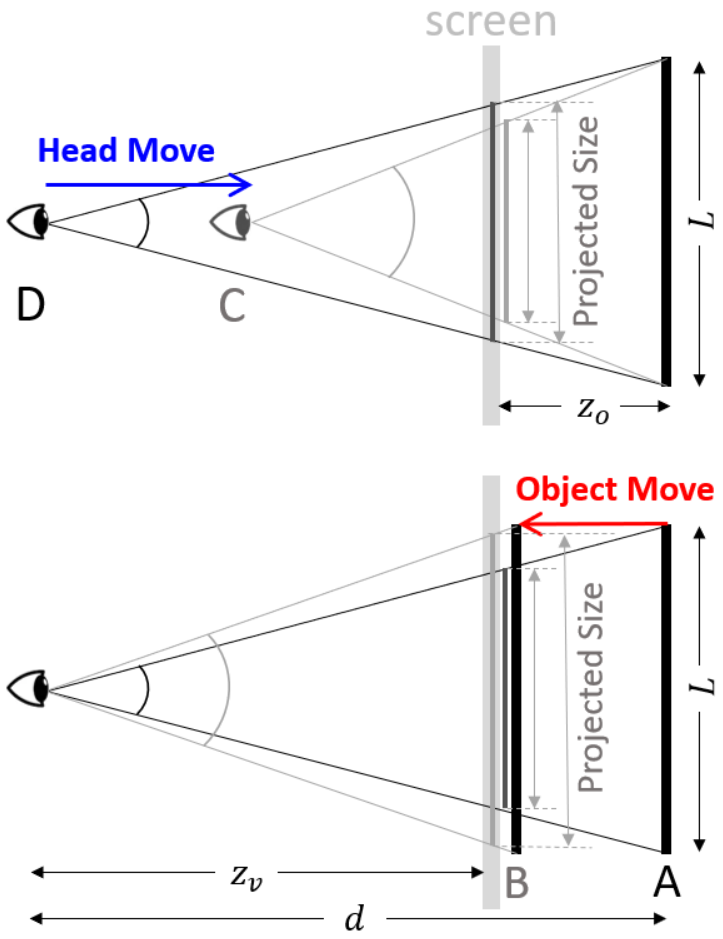


Figure 6.3: Perspective projection of a traditional planar FTVR display. (top) The projected size on the screen decreases as the viewpoint moves towards the screen from D to C in *HeadMove*. (bottom) The projected size increases as the object moves towards the screen from A to B in *ObjectMove*. The visual angle on the retinal image increases in the same way in both cases, thus a viewer moving toward an object versus object moving towards the viewer has the same impact on the retinal image.

6.2.1 Projected Size and Movement

As shown in Figure 6.3, the on-screen imagery is computed based on the perspective projection between the virtual object and viewpoint [4] . To keep consistent retinal images, we define two types of movements: *HeadMove* and *ObjectMove*. We show that *HeadMove* and *ObjectMove* provides distinct on-screen imagery while maintaining the same visual angle.

In *HeadMove* (Figure 6.3 (top)), the viewpoint moves closer to the virtual object via forward head movements toward the screen. The projected size on the screen can be computed as:

$$ProjectedSize = L - \frac{Lz_o}{d} \tag{6.1}$$

where d is the viewing distance between object and viewpoint, z_o is the distance between object and screen, and L is the virtual object's size. The projected size decreases as d decreases, shown as the blue line in Figure 6.4 (left). In *ObjectMove* (Figure 6.3 (bottom)), the virtual object moves closer to the viewpoint. Similarly, it can be computed as:

$$ProjectedSize = \frac{Lz_v}{d} \tag{6.2}$$

where L is the virtual object's size, z_v is the distance between the viewpoint and screen. Contrary to *HeadMove*, the on-screen projected size increases as d decreases, shown as the red dash line in Figure 6.4 (left). In both *HeadMove* and *ObjectMove*, the visual angle α can be computed as:

$$\alpha = 2arctan\frac{L}{2d} \tag{6.3}$$

As the visual angle α only depends on d and L, it is independent of movement types; thus, the retinal images are the same across *HeadMove* and *ObjectMove* as shown in Figure 6.4 (right). The primary difference between two movements is the on-screen size of the 2D imagery. Note that the changes of the on-screen are in opposite for *HeadMove* and *ObjectMove*: it shrinks in *HeadMove* and grows in *ObjectMove*. In this study, we want to test whether the oppositely varied size of the on-screen imagery results in different size interpretations with equivalent retinal images.

[4] The implementation of the perspective-corrected rendering in FTVR can be found in Section 4.2.

While the example shown in Figure 6.3 uses a planar screen, the same projected phenomena occurs for arbitrary screen shapes, as each pixel on the screen follows the same rule of the perspective projection in the view frustum using the rendering approach described earlier in Figure 4.4.

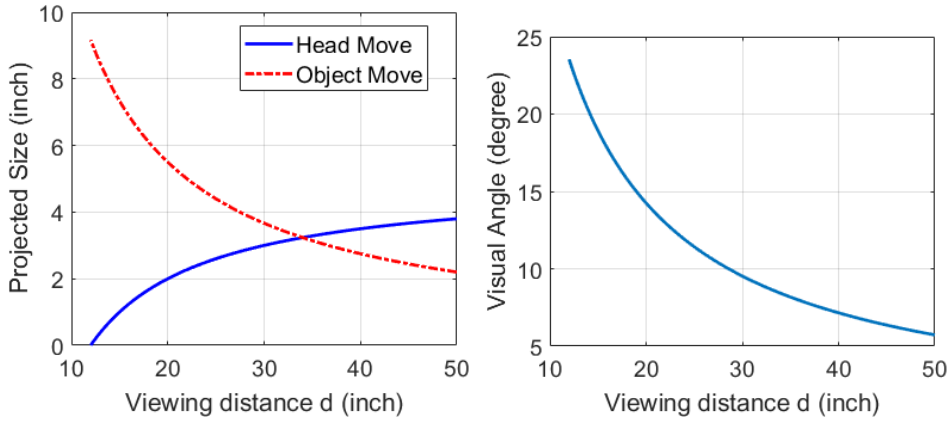


Figure 6.4: Diagram of equation 6.1, 6.2 and 6.3. (left) Projected size on the screen and (right) visual angle on the retinal image as functions of the viewing distance d. In *HeadMove*, the projected size decreases as d decreases, while in *ObjectMove*, the projected size increases as d decreases. In both *HeadMove* and *ObjectMove*, the visual angle increases independent of movement types as d decreases. The values of parameters in equation 6.1, 6.2 and 6.3 are based on our experimental setup.

6.2.2 Task

Participants visualize a virtual ball while getting closer to the ball, via either *HeadMove* or *ObjectMove*. The task for participants is to judge whether the size of the ball has changed or not by making a three-alternative forced choice via answering the question: "Is the size of the ball getting smaller, larger or unchanged?". Early pilots of this experiment showed that presenting the virtual ball without modifying its size would trivialize the task; to make the task nontrivial, the size of the virtual ball is adjusted by making it smaller, larger or unchanged so that ball's size was varied at the same time as the head/object was moving. If participants' perception is influenced by the on-screen imagery, their answers will be biased towards one side.

6.2.3 Experimental Design

We followed a 2x2 within-subjects design with two independent variables as:

- **C1** the movement condition, which could be *HeadMove* or *ObjectMove*. In *HeadMove*, participants move the head towards the object while in *ObjectMove* they move the object towards them.
- **C2** the viewing condition, which could be *Stereo* or *NonStereo*. In *Stereo*, participants visualize stereoscopic imagery while in *NonStereo* they visualize monocular imagery set to the mid-point of two eyes as suggested by [36].

User performance was evaluated based on the measure *BiasError*, defined as the difference of scores between the reported and expected answer, with the score of Small, Same and Large equal to -1, 0 and 1 respectively. Hence, a positive value of *BiasError* indicates overestimation of size while a negative value means underestimation.

6.2.4 Hypothesis

We hypothesized that:

- **H1-1:** there is a difference of size perception on *BiasError* between the *HeadMove* and *ObjectMove*;
- **H1-2:** *Stereo* will have lower *BiasError* than *NonStereo* because of its additional depth cue.

These hypothesis were based on the combination of previous research on the duality of size perception in pictures [35, 142] and the observations made using FTVR displays in lab.

6.2.5 Stimuli

Similar to other size perception studies [73, 123, 152], we chose a spherical stimuli due to its isotropic shape. We use the same texture and shadow pictorial cues across conditions to help users perceive ball depth. A shadow is dropped to appear on a plane overlapping the physical black surface holding the display as shown in Figure 6.5. It is necessary to render a shadow to indicate the ball's position. In particular, without stereopsis, this is the primary visual cue that indicates ball depth. We chose a wooden texture so

users do not have an obvious size feature. This would encourage users to perceive the size based on the volume in 3D rather than 1D or 2D features like the length of a checkerboard pattern. In real life, people do see wooden balls but have no prior knowledge about their exact size. We assume this would help to minimize prior size bias.

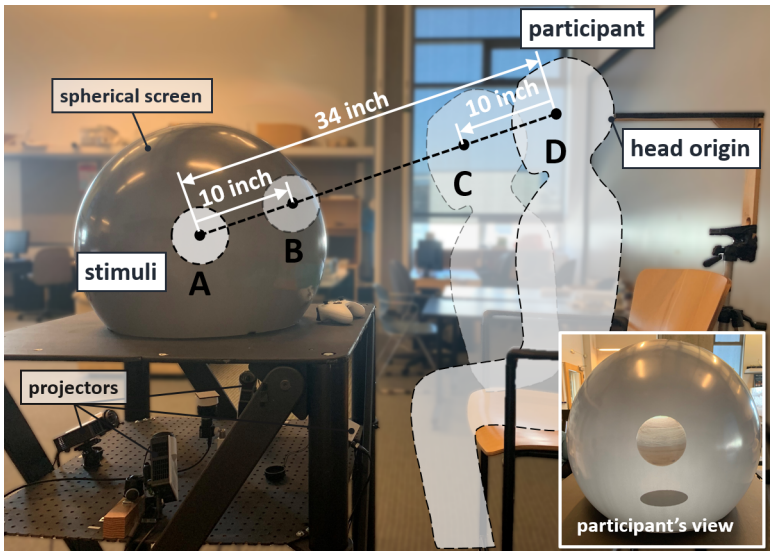


Figure 6.5: Experimental setup of Study 1. In *HeadMove*, participants move their head from D to C when the stimuli stays at A. In *ObjectMove*, participants move the stimuli from A to B when the head stays at the origin D. We use a spherical display (24" diameter) with projectors rear-projecting through a projection hole at the bottom of the screen. We track the head position to render view-dependent imagery shown in the right corner as an example of *NonStereo*, and ensure the movement magnitude is 10 inches in both *HeadMove* and *ObjectMove*.

6.2.6 Participants

Seventeen participants (12 male and 5 female) from the University of British Columbia were recruited to participate in the study with compensation. Ethic was approved by the University of British Columbia Behavioural Research Ethics Board (H08-03005). All participants passed the stereo acuity test. The average age of all participants was 29 years old from 18 to 35 years old. All participants completed written informed consent. We also asked participants to provide their previous experience and usage with Vir-

tual Reality displays with a scale from 1 ("never") to 4 ("regularly"). The average score of all participants was 2.4.

6.2.7 Procedure

Participants started by filling out the consent form after verbal explanations of the study. We measured the interpupillary distance (IPD) of each participant with a ruler tape and calibrated the viewpoints based on the IPD [138]. Prior to the study, they underwent a stereo acuity test to confirm the eligibility [42]. Then they were seated on a fixed chair in front of the spherical display. They were instructed to place their head in a position where the head could gently touch a wooden bar rigidly attached on the chair to ensure the viewing distance d of 34 inches as shown in Figure 6.5. To ensure the consistency of the moving velocities in *HeadMove* and *ObjectMove*, participants were instructed to perform forward movements toward the screen paced by an audible electric metronome at 1.5 Hz similar to [18, 91]. We measured the velocity of their head movements before the study and set the measured velocity to all conditions.

As shown in Figure 6.5, in *HeadMove*, participants were required to judge the size change of the ball placed at A, while moving their heads forward towards the display from D to C. They moved their head 10 inches with the velocity paced by the metronome. Participants were presented with two successive stimuli per trial, which allows them to confirm their answer before reporting it. They chose among three alternatives of smaller, unchanged and larger using a controller to report their answer. In *ObjectMove*, participants were required to judge the size change of the ball while pressing a button on the controller to move the ball towards them. They moved the ball 10 inches from A to B at the pre-measured velocity paced by the metronome while keeping their heads stationary at the origin D. Likewise in *HeadMove*, they were presented with two successive stimuli per trial and reported the answer among three alternatives using a controller.

In both conditions, the movement of head/object caused the viewing distance d to decrease from 34 inches to 24 inches. This is to ensure the visual angle changes in the same way across conditions. Each participant conducted 12 data trials plus 3 practice trials per condition, resulting in 48 (12x4) data observations for *BiasError*. The 12 data trials always contained 4 larger, 4 smaller and 4 unchanged stimuli in a random sequence at the resizing ratios of either 0% (4 unchanged), 15% (2 larger, 2 smaller) or 30% (2 larger, 2 smaller). The initial diameter of the ball was randomized between 4 and 6 inches. At the end each condition, we presented a questionnaire

with two likert scale questions (confidence and realism) to participants and asked them to rate each with a number in the range -2 ("totally disagree") to 2 ("totally agree").

Early pilots of this experiment showed that repeated toggling between *HeadMove* and *ObjectMove* was disorienting; to minimize this disorientation, participants only toggled between *HeadMove* and *ObjectMove* once and switched viewing conditions within each movement condition. Hence we counter-balanced the movement conditions (2), and also counter-balanced the viewing conditions (2) within each movement condition (2), resulting in 8 (2x2x2) different sequences.

6.2.8 Apparatus

We used the 24 inch prototype described in Section 4.1 and Figure 4.2 (right) to conduct the study. As shown in Figure 6.1, the user is tracked using OptiTrack [62] optical tracking system with passive markers attached to the stereo glasses. We use Unity game engine [135] to create our 3D content of the study with a two-pass rendering approach to generate perspective-corrected imagery based on tracked viewpoints [37]. We calibrated the spherical display using an automatic calibration approach [154] described in Chapter 3 with the on-surface error of 1-2 millimeter. We used a pattern-based viewpoint calibration [138] to register user-specific viewpoint with respect to the display (Figure 4.3 (right)) with the average angular error of less than one degree. The total latency is between 10-20 msec [37]. We used a post-experiment questionnaire and a demographic questionnaire in the study as attached in Appendix C.2.

6.2.9 Result

Data did not meet the normality assumption of ANOVA. A Friedman ranked sum test was performed and revealed a significant difference across conditions in *BiasError* ($\chi^2(3) = 32.0,\ p < .001$). Pairwise post-hoc Wilcoxon signed rank test for multiple comparisons with Bonferroni correction shows *BiasError* for *HeadMove* ($M = -0.23, SE = 0.043$) is significantly lower ($W = 151,\ p < .001$) than *ObjectMove* ($M = 0.48, SE = 0.098$) when viewing in *NonStereo*; *BiasError* for *HeadMove* ($M = -0.088, SE = 0.025$) is significantly lower ($W = 119,\ p < .05$) than *ObjectMove* ($M = 0.24, SE = 0.090$) when viewing in *Stereo*. *BiasError* for *Stereo* ($M = 0.24, SE = 0.090$) is significantly lower ($W = 94,\ p < .05$) than *NonStereo* ($M = 0.48, SE = 0.098$) when performing *ObjectMove*, but not when performing *HeadMove*

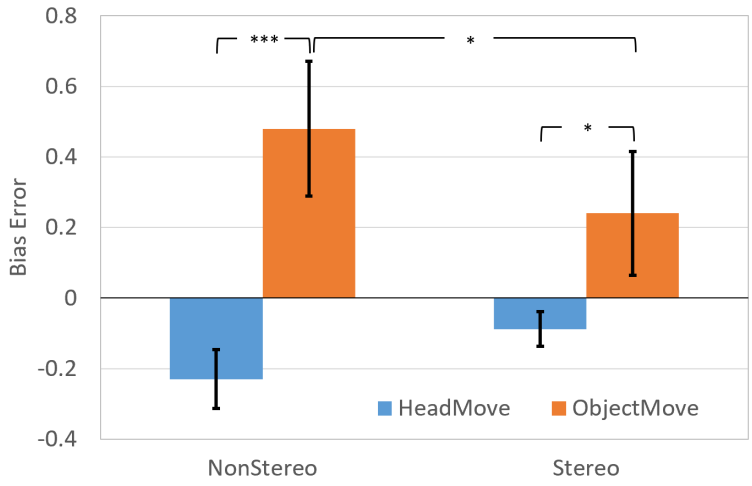


Figure 6.6: *BiasError* with means and 95% confidence intervals. *BiasError* of *HeadMove* are below zero while *BiasError* of *ObjectMove* are above zero, showing a trend of underestimation in *HeadMove* and overestimation in *ObjectMove*. Significance values are reported in brackets for $p < .05(*)$, $p < .01(**)$, and $p < .001(***)$.

($W = 20.5,\ p = .108$). The mean *BiasError* with 95% confidence intervals (CI) is shown in Figure 6.6. A complete table of results can be found in Appendix D.7.

An one-sample Wilcoxon signed rank test was performed on *BiasError* indicated a significant underestimation of size ($W(\mu < 0) = 136,\ p < 0.001$) when performing *HeadMove* and overestimation ($W(\mu > 0) = 140,\ p < 0.001$) when performing *ObjectMove*. The mean underestimation rate in *HeadMove* and overestimation rate in *ObjectMove* is 83.3% when viewing in *NonStereo*, and reduce to 64.7% when viewing in *Stereo*.

A Friedman ranked sum test was performed on the Likert-scale questions of the confidence and realism. Results revealed a significant difference across conditions in the confidence ($\chi^2(3) = 10.9,\ p < 0.05$) and realism ($\chi^2(3) = 8.86,\ p < 0.05$). Post-hoc Wilcoxon signed rank test for multiple comparisons with Bonferroni correction did not show significant difference between any pairs. Results of the mean, median and 95% CI are shown in Figure 6.7. A complete table of results can be found in Appendix D.8 and D.9.

ANCOVA on the participants' expertise on VR was carried out on the

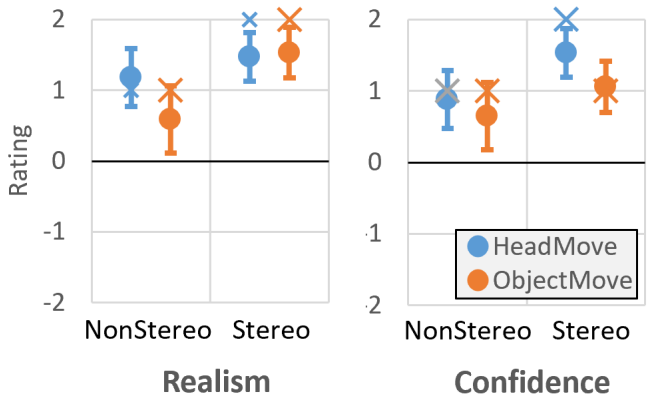


Figure 6.7: Participants' ratings with means (circle), medians (cross) and 95% confidence interval from -2 "totally disagree" to 2 "totally agree" to the questions of to how confident and real they felt about the reported result and stimuli in Study 1.

BiasError with the movement and viewing conditions as independent factors and VR expertise as covariate. The results did not show significant difference on *BiasError* ($F(1, 63) = 3.31,\ p = 0.074$) influenced by the the covariate of VR expertise.

6.2.10 Discussion

Results of One-sample Wilcoxon signed rank test show that participants systematically underestimated size in *HeadMove* and overestimated size in *ObjectMove* when retina images are the same across conditions. Post-hoc analysis of the Friedman test show that the perceived size in *HeadMove* is significantly smaller than *ObjectMove*. These results indicate that **C1** (movement) affects the perceived object size between *HeadMove* and *ObjectMove*. Hence we reject the null hypotheses of `H1-1` and accept it. Results of post-hoc analysis also show that overestimation of size is less severe in *Stereo* than *NonStereo*, indicating that adding the stereo cue significantly mitigated the perceptual bias. Given the result, we reject the null hypotheses of `H1-2` and accept it. The result of *Stereo* is consistent with Stefanucci's study [128] in which they found the size underestimation was alleviated by the addition of stereo as a depth cue in the display.

While this perceptual bias was alleviated by the stereo cue, many FTVR displays do not support stereoscopic viewing [36] especially for multiple

users due to limitations of hardware [153]. These FTVR displays require approaches to reduce this perceptual bias when viewing without *Stereo*. As an important pictorial cue, linear perspective has been closely associated with size perception, as an object getting further appears smaller to us. As shown in Figure 6.3, the perspective projection makes the on-screen imagery varied along with the depth of viewpoint and virtual object. A natural question that arises is whether using different projection matrices can affect the under/overestimation of perceived size. Hence, we designed another study aiming at (i) studying the influence of projection matrices on the observed effect and (ii) confirming the influence of the on-screen imagery on the perceived size.

6.3 User Study 2: Influence of the Projection Matrix

We investigated whether different sizes of the on-screen imagery caused by different projection matrices can affect perceived size of virtual objects. We used two types of projection matrices (*Persp* and *WeakPersp*) to create on-screen imagery with different sizes when viewing in *NonStereo* as described below.

Projection Matrix

In *Persp*, the on-screen imagery is generated using perspective projection based on the positions of viewpoint, screen and virtual object as shown in Figure 6.3. In *WeakPersp*, we used an average constant distance z_{avg} between the viewpoint and screen, so that the on-screen imagery is generated based on the constant distance independent of the viewpoint-screen depth as shown in Figure 6.8(left). When the viewpoint moves towards the screen in *HeadMove*, the projected size on the screen is a constant:

$$ProjectedSize = \frac{Lz_{avg}}{z_o + z_{avg}} \tag{6.4}$$

where z_{avg} is the constant average distance between screen and viewpoint, z_o is the distance between object and screen, and L is the virtual object's size. The on-screen projected size is a constant as shown in Figure 6.8 (right). In *ObjectMove*, when the object moves closer to the viewpoint, the on-screen projected size increases as z_o decreases as shown in Figure 6.8 (right).

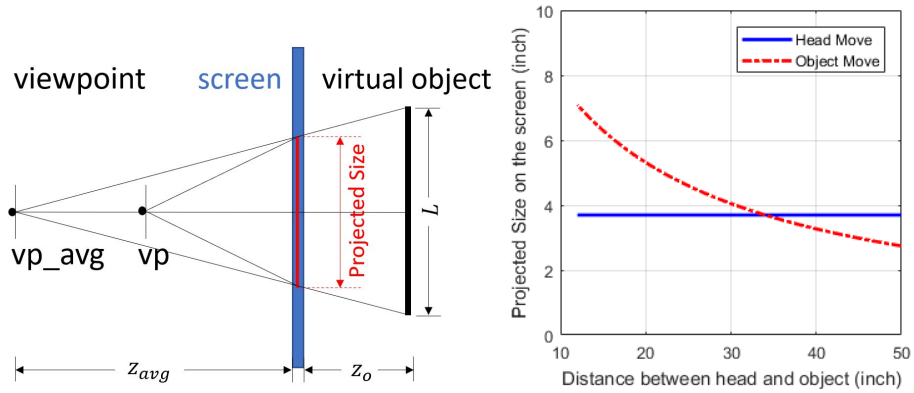


Figure 6.8: Weak perspective projection of a traditional planar FTVR display. (left) The on-screen imagery is generated based on a constant distance z_{avg} between the viewpoint and screen, so that the projected size is independent of the viewpoint-screen depth. (right) Projected size on the screen as a function of viewing distance. In *HeadMove*, the projected size is a constant independent of the viewing distance, while in *ObjectMove*, the projected size changed at a smoother gradient compared to the perspective projection in Figure 6.3(b).

We chose *WeakPersp* due to the following reasons. First, *WeakPersp* provides a smoother transition on the projected size than *Persp* when the viewpoint gets closer to the object. Comparing Figure 6.3(b) with Figure 6.8(right), *WeakPersp* generates on-screen imagery with a constant size in *HeadMove*, while *Persp* shrinks the projected size, which contradicts to the reality as users expect a larger percept when getting closer. Hence, rendering on-screen imagery with a constant size can potentially preserve the perceptual consistency between the virtual and real world. In the case of *ObjectMove*, the gradient of projected size in *WeakPersp* is smaller than the gradient in *Persp*. While the projected size will still become larger, the smaller gradient provides a smoother transition so there will be less overestimation on the perceived size as observed in Study 1. Second, since the on-screen imagery shrinks in *Persp* when moving towards the screen, the same virtual object will be rendered with fewer pixels, causing a decrease in the perceived resolution, which is also against the common sense as closer objects should appear more clear. *WeakPersp* warps the viewing distance at a constant value, which prevents the resolution from decreasing when the viewer moves towards the screen. It is worth noting that *WeakPersp* still

provides perspective corrected imagery as the imagery changes when users perform lateral movements, but the imagery remains unchanged for forward or backward movements.

We evaluate the effect of projection matrices using the same size judgement task as described in Section 6.2. As different projection matrices may also influence users' general spatial impression and subjective preference. We also included a forced-choice viewing preference task to assess subjective 3D impression.

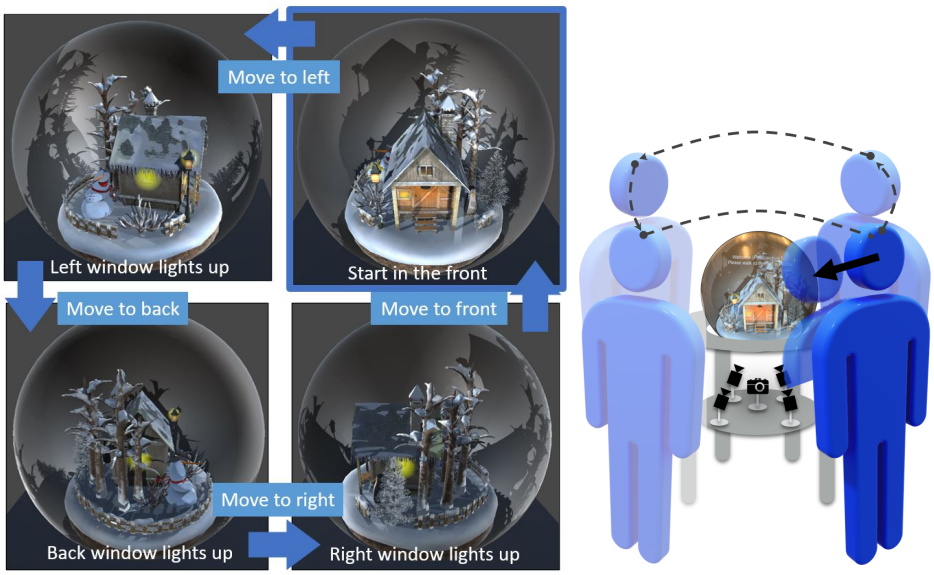


Figure 6.9: Subjective impression task: participants walked around the display clockwise and performed forward head movement towards the screen for closer inspection before selecting their preference between a pair of projection conditions (*Persp* and *WeakPersp*).

6.3.1 Task 1: Subjective Impression

Participants visualized a stationary 3D scene with a house model while walking around it. They were instructed to pay close attention to how 3D the scene appeared and how natural the objects looked like as in the real world. As shown in Figure 6.9, participants were instructed to stand facing the front of the house, and walk to each side of the house clockwise. At each side of the house, they were asked to inspect the house and move closer until the house was lit up shown from a window on the side of the house. Once

the window lit up, they walked to the next side until completing a full circle. We therefore enforced forward and lateral head movement to simulate the viewing experience for 3D visualization when users browse the 3D data by walking around, and inspect points of interest by moving closer to the display. The independent variable is **C3** (projection) with two levels of *Persp* and *WeakPersp*. We measured subjects' rating and preference on **C3** via a post-questionnaire (Appendix C.3).

6.3.2 Task 2: Size Judgement

We used the same size judgement task in Study 1. Participants visualized a virtual ball while getting closer to the ball, via either *HeadMove* or *ObjectMove*. The task for participants was to judge whether the size of the ball has changed or not by making a three-alternative forced choice as described in Section 6.2.2. We followed a 2x2 within-subject design with two independent variables as **C3** (projection) and **C1** (movement). Likewise in Task 1, **C3** has two levels of *Persp* and *WeakPersp*, and **C1** has two levels of *HeadMove* and *ObjectMove*. Similar to Study 1, in *HeadMove*, participants moved the head towards the object while in *ObjectMove* they moved the object towards them. Participants' performance was evaluated based on the measure *BiasError* as in Study 1 described in Section 6.2.3. A positive value of *BiasError* indicates overestimation of size while a negative value means underestimation.

6.3.3 Hypothesis

We hypothesized that:

- **H2-1:** *WeakPersp* will be preferred over *Persp* since the on-screen imagery of *WeakPersp* does not change when subjects get closer to the virtual stimuli.
- **H2-2:** *WeakPersp* will have lower *BiasError* than *Persp* since *WeakPersp* produces more stable on-screen imagery when the viewing distance is decreasing.

6.3.4 Participants

Twelve participants (7 male and 5 female) from University of British Columbia were recruited to participate in the study with compensation. Ethic was approved by the University of British Columbia Behavioural Research Ethics

Board (H08-03005). The average age of all participants was 27 years old from 18 to 35 years old. We also asked participants to provide their previous experience and usage with Virtual Reality displays with a scale from 1 ("never") to 4 ("regularly"). The average score of all participants was 2.1.

6.3.5 Procedure

All participants completed written informed consent. They performed Task 1 before Task 2. The order of tasks was the same for all participants, whereas the projection and movement conditions were counter-balanced within each task. In Task 1, participants walked around and visualized the display in *Persp* and *WeakPersp* with a maximum viewing distance of 50 inch to the display. The constant viewing distance of *WeakPersp* is chosen to be 50 inch to maximize perceived resolution. The light on each side of the house would be turned on when participants moved forward and reached 20 inch from the display. After each projection condition, they filled in a post-questionnaire (Appendix C.3) with five Likert-scale questions to rate each with a number in the range -2 ("not at all") to 2 ("very much"). The five questions were chosen and modified from Parent's questionnaire which evaluated the physical presence of virtual art exhibits [100]. At the end of Task 1, participants were asked to provide their preferences and specify reasons between projection conditions. Task 2 has the same procedure as in Study 1. Participants were seated on a fixed chair in front of the display with the procedure described in Section 6.2.7. Likewise in Study 1, each participant conducted 12 data trials plus 3 practice trials per condition, resulting in 48 (12x4) data observations for *BiasError*.

We used the same apparatus as in Study 1 for both tasks, except for the questionnaires. A post-experiment questionnaire assessing the subjective impression and a demographic questionnaire used in Study 2 can be found in Appendix C.3.

6.3.6 Result

Task 1 - Subjective Impression: Chi-Square analysis shows significant differences on the preference of **C3**(projection) ($\chi^2 = 8.33, p < 0.01$) with 11 out of 12 participants preferring *WeakPersp* over *Persp*. Wilcoxon Matched Pairs Signed Rank Test was performed on the likert scale questions. There were significant differences on the Q1-consistent ($W = 28.0,\ p < .05$), Q2-reachable ($W = 55.5,\ p < .05$), Q3-geometry ($W = 50.0,\ p < .05$), and Q4-realistic ($W = 32.5,\ p < .05$), but not Q5-presence ($W = 37,\ p = .09$).

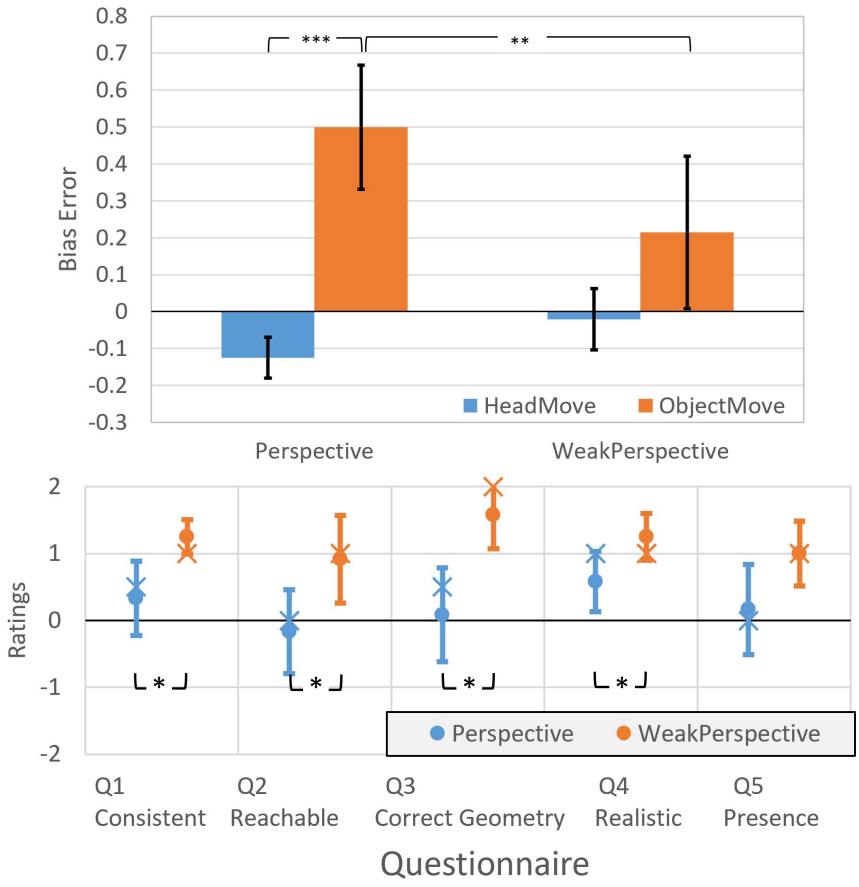


Figure 6.10: (top) *BiasError* with means and 95% confidence intervals (CI). The under and overestimation rates in *HeadMove* and *ObjectMove* has been reduced with *WeakPersp* compared to *Persp*. (bottom) Participants' rates with means (circle), medians (cross) and 95% CI from -2 "not at all" to 2 "very much" to the five questions in the post-questionnaire. The five questions include: Q1-consistency ("visualization is consistent with real-world experience"), Q2-reachable ("feel can reach and grasp a virtual object"), Q3-geometry ("virtual objects appear geometrically correct"), Q4-realistic ("virtual objects appear realistic") and Q5-presence ("feel virtual objects exist in the real environment"). Significance values are reported in brackets for $p < .05(*)$, $p < .01(**)$, and $p < .001(***)$.

The results of questionnaire are summarized in Figure 6.10. A complete table of results can be found in Appendix D.11.

Task 2 - Size Judgement: Repeated measures two-way ANOVA was carried out on the *BiasError* with **C1**(movement) and **C3**(projection). Results revealed a two-way interaction effect between **C1** and **C3** ($F(1, 11) = 10.1,\ p < 0.01$). Main effect of **C1** was also found ($F(1, 11) = 36.64,\ p < .001$), but not on **C3** ($F(1, 11) = 4.14,\ p = 0.067$).

Post-hoc pairwise t-test with Bonferroni correction for the interaction between **C1** and **C3** shows *BiasError* for *HeadMove* ($M = -0.125, SE = 0.028$) is significantly lower ($t = 6.66,\ p < .001$) than *ObjectMove* ($M = 0.5, SE = 0.086$) when viewing in *Persp*, but not in *WeakPersp* ($t = 2.52,\ p = 0.08$). *BiasError* in *WeakPersp* ($M = 0.215, SE = 0.105$) is significantly lower ($t = 3.77,\ p < .01$) than *Persp* ($M = 0.5, SE = 0.086$) when performing *ObjectMove*, but not when performing *HeadMove* ($t = 1.38,\ p = 0.732$). The mean *BiasError* with 95% CI is shown in Figure 6.10. A complete table of results can be found in Appendix D.10.

One-sample t-test was performed on *BiasError* indicated a significant underestimation of size ($t(\mu < 0) = -4.450,\ p < 0.001$) when performing *HeadMove* in *Persp*, but not in *WeakPersp* ($t(\mu < 0) = -0.491,\ p = 0.633$). It also indicated a significant overestimation of size($t(\mu > 0) = 5.826,\ p < 0.001$) when performing *ObjectMove* in *Persp*, but not in *WeakPersp* ($t(\mu > 0) = 2.044,\ p = 0.066$). In *Persp*, the mean underestimation rate of *HeadMove* is 83.3%, and the overestimation rate of *ObjectMove* is 91.7%, which reduced to 50.0% and 58.3% respectively in *WeakPersp*.

ANCOVA on the participants' expertise on VR was carried out on the *BiasError* with the movement and viewing conditions as independent factors and VR expertise as covariate. The results showed significant difference on *BiasError* ($F(1, 43) = 4.82,\ p < 0.05$) influenced by the covariate of VR expertise.

6.3.7 Discussion

Results show that participants systematically underestimated size in *HeadMove* and overestimated size in *ObjectMove* when viewing in *Persp*, while no significant under or overestimation is found when viewing in *WeakPersp*. Results of two-way anova show that the perceptual bias in *Persp* is significantly larger than *WeakPersp* when performing *ObjectMove*. These results indicate that different projection matrices affect the perceived object

size in 3D space. Hence we reject the null hypothesis of `H2-1` and accept it. When performing *HeadMove*, although there is difference on the perceived size between *WeakPersp* ($M = -0.021, SE = 0.042$) and *Persp* ($M = -0.125, SE = 0.028$), the result is not significant. However, results of Task 1 show that participants overwhelmingly preferred *WeakPersp* over *Persp* when performing *HeadMove* and visualizing 3D scene. Given the result, we reject the null hypothesis of `H2-2` and accept it.

6.4 General Discussion

The results of our studies show a perceptual bias in size perception caused by on-screen imagery. Adding stereo cue can mitigate the observed bias. When viewing without the stereo cue, weak perspective projection reduced the under/overestimation rate. Users also have a strong subjective preference for *WeakPersp* over *Persp*.

6.4.1 Influence of the On-screen Imagery

Results of both studies show that participants systematically underestimated size in *HeadMove* and overestimated size in *ObjectMove* when retina images are the same across conditions. This indicates the on-screen size of the 2D imagery affects the perceived object size in 3D space. In particular, when participants moved closer to the object in *HeadMove*, they had a tendency to report objects as smaller. This contradicts reality as we usually expect closer objects look larger due to a larger visual angle. In addition, the bias appears to be stronger in *ObjectMove* than *HeadMove* in both studies. One potential explanation is that the absolute gradient of the projected size along the viewing distance is steeper in *ObjectMove* than *HeadMove* as shown in Figure 6.3(b) and Figure 6.8(right). To encourage perception of 3D features, users were instructed to attend to the change of the ball. Note that if they focused on the change of the shadow or the distance between shadow and ball instead, their performance will still be consistent because they are visualizing the scale of the entire scene and any geometries in the scene are subject to the projection model. However, it may impact the extent of the effect, which may account for the performance variance shown in Figure 6.6.

Knowing the effect of the on-screen imagery helps us to understand spatial perception in the screen-based 3D displays. One of the perceptual errors is the size underestimation of virtual objects [73, 91, 128]. Virtual objects are usually located behind the screen in the screen-based 3D displays

[126, 128, 152], rendering an on-screen imagery smaller than the actual size as shown in Figure 6.3. If the size perception regresses towards the on-screen imagery similar to the perceptual bias observed with the real object [35], users will have a tendency to report a modified value for the perceived object size, biased towards the on-screen size, resulting in the underestimation of the reported value. Hence the on-screen imagery could be a possible source for the size underestimation. On the other hand, if virtual objects are rendered in front of the screen, the on-screen imagery will be larger than the actual size. Therefore, we expect users will overestimate the size, which requires future experiments to investigate the effect of the on-screen when rendering on different side of the screen.

The influence of the on-screen imagery might not be restricted to size perception. One of the visual artifacts in screen-based 3D displays is the *floating effect* [36], sometimes also known as *swimming effect* [51]. Users observed stationary virtual objects to move along while they walked around in the environment [103] . As shown in Figure 6.11, as the viewer moves laterally while visualizing a ball in the center, the projection on the screen also moves based on the viewer's position. Likewise in our size judgement task, if the on-screen optical flow interferes with users' percept of the virtual object, they may have a tendency to report that the object moves along with their head movement. With system latency and registration error, such local changes may become more noticeable. Hence, the on-screen imagery could be a potential cause for the *floating effect*.

Another problem which has been well-known is the *vergence/focus conflict*. It describes a problem as viewers converge toward the virtual object but focus eyes on the on-screen imagery, which has been found to cause eyestrain [142]. We see the perceptual duality and *vergence/focus conflict* as two separate problems arisen from the same origin: users directly see the pixels rather than the object. The depth separation of the on-screen imagery and the 3D object causes visual artifacts and problems, which unfortunately present in all synthesized 3D displays which use pixels to depict a 3D scene.

6.4.2 Size Perception Accuracy

As *BiasError* does not directly reflect the accuracy [5], to better understand the effect of *WeakPersp* and *Stereo* on the performance, we also computed the *AbsoluteError*, defined as the average absolute difference between the reported answer and the correct answer. In study 2, while

[5] *BiasError* can be negative or positive so that the average *BiasError* can be zero while the absolute mean error is much greater than zero.

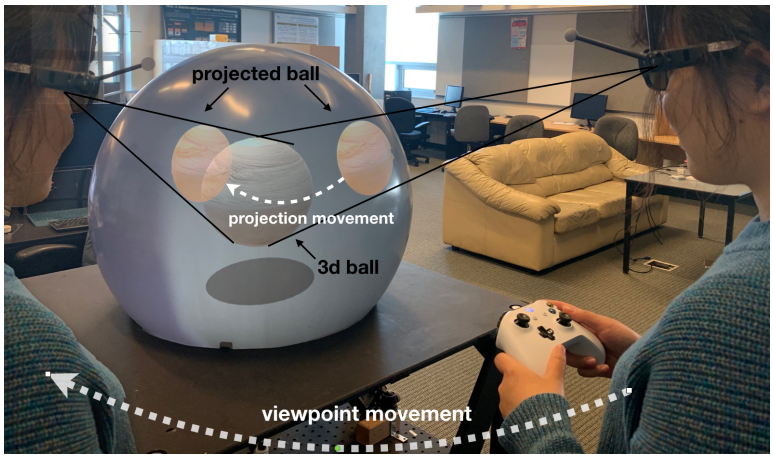


Figure 6.11: Illustration of the *floating effect*: users observe stationary virtual objects to move along while they walk around in the environment, which could possibly be caused by the movement of on-screen imagery.

WeakPersp reduces perceptual bias on perceived size, it did not significantly improve the accuracy of perceived size, as there is no significant difference ($F(1, 11) = 1.53,\ p = 0.24$) between *WeakPersp* ($M = 0.354,\ SE = 0.050$) and *Persp* ($M = 0.403,\ SE = 0.056$) on *AbsoluteError*. In Study 1, *AbsoluteError* is significantly lower ($F(1, 16) = 21.1,\ p < .001$) with *Stereo* ($M = 0.306,\ SE = 0.046$) than *NonStereo* ($M = 0.466,\ SE = 0.043$) , suggesting the addition of the stereo cue improved the size perception with less systematic bias but also better accuracy. The result is consistent with Ware's discussions on the duality of depth perception in picture: the amount and effectiveness of the depth cues can help viewers to judge the size of a depicted object in a 3D space rather than on the picture plane [142]. This is similar to the scenario when people view a painting in the real world. They can choose to see the depicted object as a 3D structure, or as a 2D picture surface. The painter creates spatial vividness by adding various pictorial depth cues in their work to strengthen the illusion. Similarly, the stereo cue in our study is also an additional depth cue, making it easier to see in a '3D mode', rather than in a '2D mode', reducing sensitivity to the on-screen cues. Hence, it is suggested that future designs of FTVR displays should include stereopsis to alleviate the influence of on-screen cues.

6.4.3 *HeadMove* vs *ObjectMove*

HeadMove and *ObjectMove* are designed to provide equivalent retinal images across conditions while rendering oppositely varied size of the on-screen imagery. Naturally, in this study, we cannot decouple the on-screen imagery from the movement itself, making it ambiguous to interpret the result as whether the bias is caused by different on-screen cues or different movement types. In reality, when our eyes get closer to an object, via either head or object movement, we always feel closer objects look larger due to the increased visual angle. Hence it seems unlikely that the movement types would be responsible for the opposite *BiasError* observed in our study. We also computed *AbsoluteError* for *HeadMove* and *ObjectMove*. In study 1, *AbsoluteError* of *ObjectMove* $(M = 0.578,\ SE = 0.038)$ is significantly higher $(F(1, 16) = 45.8,\ p < .001)$ than *HeadMove* $(M = 0.194,\ SE = 0.027)$, so is in study 2 $(F(1, 11) = 37.25,\ p < .001)$. In the study, we provided limited pictorial cues with a plain background to keep the visual stimuli simple. One possible explanation for the performance difference is that the plain background did not provide sufficient depth information when the object moved towards participants in *ObjectMove*, compared to *HeadMove*, in which they might have a better depth perception via self movement. As depth and size perception are closely related, the lack of depth cues might be the cause of the difference in *AbsoluteError*. Additionally, *HeadMove* provides extra proprioception cue compared to the object movement. Participants could see the changes of real-world objects as they approached the screen, while the only visual cue in *ObjectMove* is the virtual object. Therefore, it is suggested to consider the head movement for better size perception.

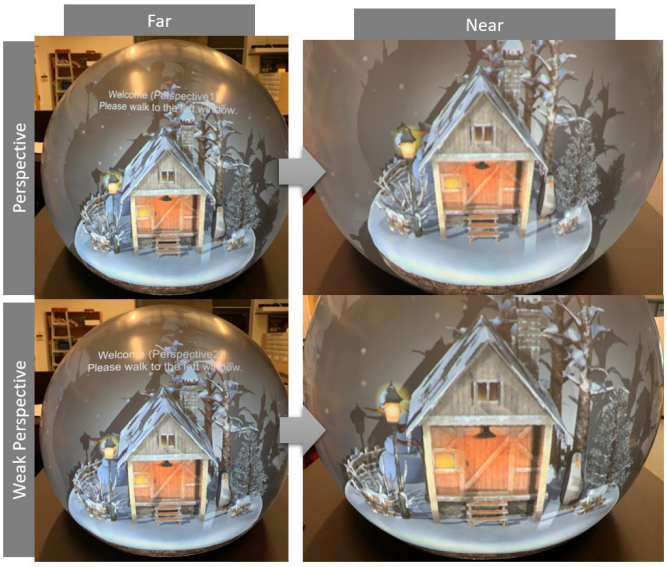


Figure 6.12: Participant's view with (top) *Persp* and (bottom) *WeakPersp* when moving from (left) far to (right) near. While there is only subtle difference between *Persp* and *WeakPersp* when viewing at far, the imagery of *WeakPersp* is enlarged and distorted compared to *Persp* when viewing at near.

6.4.4 Choice of Projection Matrix

In Study 2, we used the weak perspective projection to warp the viewing distance at a maximized constant value for a higher resolution, resulting in an enlarged projected imagery compared to *Persp*. However, the imagery is distorted due to the warping so that straight lines become noticeable curved when viewing at a close distance as shown in Figure 6.12. In addition, the enlarged imagery inflates the perceived absolute scale of virtual objects. Particularly, we found that participants with higher VR expertise tend to overestimate the size with positive *BiasError* as shown in Figure 6.13. One potential explanation is that advanced VR users are more sensitive to the enlarged imagery caused by *WeakPersp*, as *Persp* is the default projection matrix commonly used in VR applications.

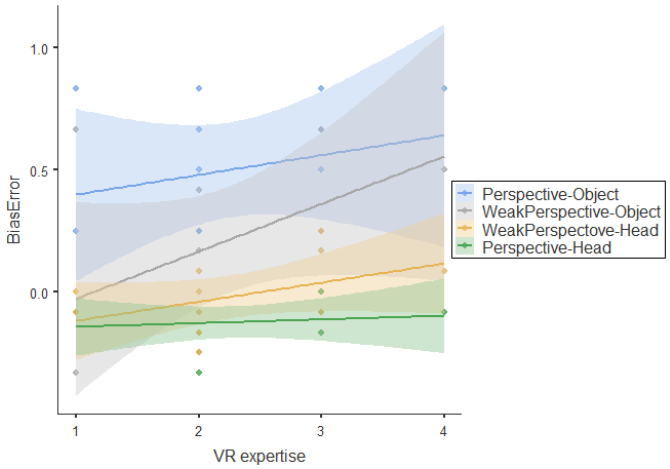


Figure 6.13: *BiasError* influenced by participants' expertise on the usage of VR. Participants with higher VR expertise are more likely to overestimate the size in *WeakPersp* than *Persp*, potentially caused by the enlarged projected imagery in *WeakPersp*, which is less commonly used in VR scenes.

Despite of these visual artifacts, *WeakPersp* was still favored by 91.67% participants. When asked about the reason, 5 out of 12 participants mentioned that the scene looked larger when getting closer, which better matched with the real world experience. Two participants further explained that they felt the scene in *Persp* was "moving away" from them since it shrunk as they moved closer. This finding is interesting, if we consider the fact that *Persp* produces on-screen imagery which is geometrically correct based on the viewer's position, while the imagery is actually warped in *WeakPersp*. The result of questionnaire regarding Q3-geometry shows opposite rating: participants believed *WeakPersp* provided correct geometry projection with respect to the viewpoint as shown in Figure 6.10. This leads us to believe that the warped scene in *WeakPersp* better matched viewer's expectation when moving closer to the screen, even if it is not geometrically correct. A potential explanation might be related to users' prior experience with spherical objects such as fish bowls and snow globes, which usually have visual distortions due to the light refraction caused by the spherical glass and water. It is possible that similar distortion on the spherical displays is expected, which requires further experiments to investigate.

While *WeakPersp* was favored by participants, it may not be the optimal projection matrix. Our studies investigated *WeakPersp* and *Stereo* separately in two studies. In Study 2, we rendered in *WeakPersp* without

Stereo cue. When they are coupled, it is worth noting that the disparity of *Stereo* will not change due to the nature of *WeakPersp*. In other word, when users get closer to the screen, the disparity deviates from the correct value which could potentially cause problems. More experiments are required to investigate the potential issues and understand the effect of projection matrix. In our study, it shows *WeakPersp* is preferred than *Persp* when viewing in *NonStereo* for a simple walk-around visualization. This indicates *WeakPersp* could be considered for *NonStereo* casual applications such as attention-drawing showcase visualization which does not require users to wear glasses.

6.4.5 Design Recommendations for FTVR Displays

As our study found that the on-screen imagery influenced users' size perception, we summarize the following design recommendations for FTVR displays. First, it is suggested to include the stereo cue to reduce the observed perceptual bias. Similar to the stereo cue, other depth cues such as pictorial cues or motion parallax cue may help to reduce the bias, while it is worthwhile investigating these cues in the future study. Second, when viewing without the stereo cue, projection matrices which produce a stable on-screen imagery, such as weak perspective projection, can mitigate the perceptual bias and make sure users perceive 3D scenes in a way as expected. Third, the head movement should be considered to provide accurate size perception in 3D applications which need to accurately depict the size and shape of the object, such as computer aided design and virtual surgery.

6.4.6 Limitations and Future Work

There are several limitations to the experiments presented in this chapter. Our study was performed on a Fishbowl VR display. We chose a spherical display as it is the most common form-factor for recent FTVR displays. We chose a spherical stimuli as it has been widely used in size perception studies [73]. However, awareness should be raised on the potential interaction between the shape of the screen and the stimuli. Our study evaluated the size perception of a spherical stimuli on a spherical screen. Future work could investigate stimulus with different shapes on displays with other shapes.

We expect the findings could be extended to other display shapes such as planar or cylindrical displays, because the design of the task, the analysis of the projected size and the head/object movement are independent of the shape factor. We also expect the perceptual bias to be more pronounced

when the size of the stimuli is comparable to the size of the screen, because the screen can serve as a reference that makes the local changes of the on-screen imagery more noticeable. It is also interesting to consider how these findings could be transferred to HMDs. Unlike FTVR displays, users move together with HMD screens. Ideally, there is no relative movement between the viewpoint and screen when users wear HMDs tightly. Therefore we expect the result will be different from our study as the projected size will increase in both *HeadMove* and *ObjectMove*, making closer objects look larger.

Secondly, while we found that both *Stereo* and *WeakPersp* can alleviate the perceptual bias of perceived size, they were evaluated separately. It is unclear whether applying *WeakPersp* with *Stereo* will over correct the perceptual bias. In addition, *WeakPersp* renders enlarged on-screen imagery by warping the viewer-screen depth which caused visual artifacts such as distortions on the Fishbowl VR display. The enlarged imagery may also affect the perception of absolute scale in the virtual environment. Future studies would be necessary to investigate the influence of different projection matrices with the stereo cue regarding size perception and visual artifacts.

Lastly, our study assessed the effect of on-screen imagery on perceived size. As discussed in Section 6.4.1, we expect the influence of the on-screen imagery might not be limited to size perception. Similar to the perceptual bias observed with the real object [35], it may occur not only for the size of objects but also for other attributes such as position or orientation, which may help to explain some visual artifacts observed in the virtual environment such as the *floating effect*. More experiments are required to further understand the influence of on-screen imagery on spatial perception. Our investigation with the on-screen imagery is still at a preliminary stage, but initial findings are encouraging that the on-screen imagery plays a significant role in size perception, and the stereo cue as well as projection matrix can mitigate the perceptual bias caused by on-screen imagery.

6.5 Summary

FTVR displays render perspective-corrected imagery on a 2D screen. Because users perceive a 3D object by looking at pixels on the 2D screen, there exists a perceptual duality between the on-screen pixels and the 3D percept. In this chapter, we conducted two empirical studies to demonstrate the influence of the on-screen imagery, causing 83.3% under/over-estimation of perceived size. The addition of stereopsis significantly mitigated the per-

ceptual bias to 64.7%. When viewing without stereopsis, weak perspective projection alleviated the perceptual bias to 58.3% with a strong preference by 91.7% users compared to perspective projection. The results suggest that future designs of Fishbowl VR and volumetric displays should use stereopsis and weak perspective projection to mitigate perceptual bias and reconcile pixels with percept.

Chapter 7

Conclusions

This chapter reviews and summarizes the contributions of this dissertation. We discuss the applicability of our previous findings to a broader research domain, as well as the limitations of our work. Finally, we present future directions along with concluding comments.

7.1 Research Contributions

This dissertation provides four primary contributions to the research domain of 3D displays. These contributions are: creating automatic display calibration approach (Chapter 3), formulating visual error (Chapter 4), demonstrating the superiority of the spherical screen (Chapter 5) and the perceptual duality in FTVR displays (Chapter 6). Chapter 3 uses computer vision techniques to develop an automatic calibration approach to reconstruct the display surface. Chapter 4 uses computer graphic techniques to incorporate the viewpoint into the Fishbowl VR display and analyze the visual error from the perspective-corrected rendering. Chapter 5 and 6 are grounded on human factor studies to investigate the visual perception on the Fishbowl VR display. Publications from this dissertation are listed in Appendix A and the main contributions are summarized below.

Automatic calibration of a multiple-projector spherical display

i. **Created a novel automatic calibration approach of a multiple-projector spherical display.** We developed a novel automatic calibration method using a single camera for a multiple-projector spherical display. Modeling the projector as an inverse camera, we estimate the intrinsic and extrinsic projector parameters automatically using a set of projected images on the spherical screen. A calibrated camera is placed beneath to observe partially visible projected patterns. Using the correspondence between the observed pattern and the projected pattern, we reconstruct the shape of the spherical display and finally recover the 3D position of each projected pixel on the display.

ii. **Applied and evaluated the calibration approach to a Fishbowl VR prototype.** We applied the calibration approach in a prototype and achieved sub-millimeter calibration accuracy by estimating the on-surface error. The results were consistent with existing work. The calibrated display can support both view-dependent and view-independent applications.

Error analysis of a Fishbowl VR display

i. **Formulated visual error of Fishbowl VR displays.** We formulated the visual error of a Fishbowl VR display in terms of display calibration error and head-tracking error. Using this model, we analyzed the sensitivity of the viewer's visual error to errors in each part of the system. We also presented the design and implementation of two Fishbowl VR prototypes.

ii. **Established design guidelines for FTVR displays.** We applied the error analysis on our prototypes and provided guidelines to minimize visual error for FTVR displays. We found: head tracking error causes significantly more visual error than display calibration error; visual error becomes more sensitive to tracking error when the viewer moves closer to the sphere; and visual error is sensitive to the distance between the virtual object and its corresponding pixel on the surface. Taken together, these results provide practical guidelines for building a Fishbowl VR display and can be applied to other configurations of geometric displays.

Evaluation of spatial perception in a Fishbowl VR display

i. **Compared spatial perception on a Fishbowl VR display with a traditional flat FTVR display.** We conducted an experiment and found that the Fishbowl VR display provides better task performance with depth accuracy of 1cm which yielded significantly less error compared to the accuracy of 6.5 cm on the flat FTVR display in a depth-ranking task. In a size-matching task, the performance was also better with an accuracy of 2.3 mm on the spherical display as compared to 3.1 mm on the flat display.

ii. **Demonstrated superiority of the spherical display in providing better size-constancy.** We found that the perception of size-constancy is stronger on the Fishbowl VR display than the flat FTVR

display. This indicates that the natural affordances provided by the spherical form factor better preserved the size-constancy and improved size perception in 3D compared to a flat display.

Evaluation of perceptual duality of FTVR displays

i. **Demonstrated the influence of on-screen imagery on size perception.** We conducted two studies and found that the size of on-screen imagery significantly influenced object size perception, causing 83.3% under/overestimation of perceived size. This perceptual bias indicates there is a perceptual duality between the on-screen pixels and the 3D percept.

ii. **Demonstrated the influence of stereopsis on size perception.** We conducted a study and found adding stereopsis can mitigated under or overestimation of perceived size, which reduced to 64.7% when viewing with stereopsis. It is suggested that future designs of FTVR displays should include the stereo cue to alleviate the influence of on-screen imagery.

iii. **Compared size perception under different projection matrices.** We conducted a study and found using weak perspective projection significantly reduced the perceptual bias of size perception to 58.3% and was strongly preferred by 91.7% users compared to perspective projection. Projection matrices which produce a stable on-screen imagery, such as weak perspective projection, should be considered to mitigate the perceptual bias and make sure users perceive 3D scenes in a way as expected.

7.2 Indications to FTVR Designs

We investigated the visualization and spatial perception provided by Fishbowl VR displays. Based on the result, we now revisit the form factors of FTVR displays and discuss design indications to FTVR displays. As our studies are carried out on two Fishbowl VR prototypes, we also discuss the applicability of our findings to other FTVR displays with different shapes.

Stereopsis Stereopsis has been shown to be critically important for FTVR displays in various tasks [5, 6]. Consistent to the literature, we also found that stereopsis improved size perception in Chapter 6. While users' task

performance has been significantly improved with stereopsis, we did not find significant subjective preference on stereopsis. Relevant to our findings, recent work also found that while users' performance was significantly degraded without stereopsis, they did not have a strong preference for stereo rendering [36]. Therefore, non-stereo FTVR displays would be reasonable for use cases that simply provide subjective impression such as 3D showcase at the exhibition. However, these work were all carried on spherical displays. It is possible that the spherical display creates a strong "fishbowl" metaphor of virtual objects contained within the sphere. With the metaphor, it maintains the 3D illusion without stereopsis. Therefore, this finding may not be applied on traditional planar FTVR displays.

Head-coupling Head-coupling in FTVR displays provides motion parallax cue by rendering to viewer's perspective. It is suggested to include head-coupling in Fishbowl VR displays based on the results of Chapter 5 and 6. In Chapter 6, we found that head movement provided accurate size perception with low perceptual bias. Compared to the planar screen, users exhibited more head movements with the spherical screen. For a single viewer, it is probably unnecessary to cover the entire sphere with head-coupling to provide view-dependent 360° imagery, as we found that users were only using part of the spherical display space in Chapter 5. For multi-person viewing experience, supporting 360° visibility becomes important as each viewer gains own distinct perspective around the display [37].

On-screen Imagery When users visualize a virtual object in FTVR displays, they are actually looking at pixels on a 2D screen. A perceptual duality exists between the on-screen pixels and the 3D percept. The influence of on-screen imagery inherently exists due to the nature of FTVR displays, which we evaluated on a Fishbowl VR display in Chapter 6. Though our investigation is conducted using a spherical display, the observed effect could be expected independent of the display shape. When the stimuli size is comparable to the screen size, it is possible for the perceptual bias be more pronounced, because the screen can serve as a reference that makes the local changes of the on-screen imagery more noticeable. However, if a similar experiment is conducted in a completely dark environment with light only coming from the on-screen imagery, the result might be different because it becomes difficult to observe the local changes on the screen when the screen itself is invisible.

7.3 Discussion of Fishbowl VR Displays

We found that the Fishbowl VR display is promising to provide consistent perception between the physical and digital world in Chapter 5 compared to a traditional flat counterpart display. Therefore, we believe 3D applications which require high fidelity spatial perception, such as CAD and virtual surgery, can potentially benefit from the spherical screen shape over a traditional flat screen. Given the property of the Fishbowl VR display, it is more likely for this type of display to be applied for targeted domains rather than having a wide-spread adoption in all fields. In this section, we provide a general discussion of the Fishbowl VR display on its strength and limitations by comparing to other displays.

The spherical screen of the Fishbowl VR display has an enclosure nature, which presents a metaphor as if virtual objects are inside a glass globe or display case with a finite rendering volume. This is different from flat displays with a window metaphor and infinite rendering volume. While it is possible to render virtual objects outside the boundary of the spherical surface, it is inconsistent with the metaphor of having a glass globe and therefore potentially breaks the 3D illusion. Furthermore, the virtual objects are more sensitive to the tracking error when rendered outside the screen as discussed in Section 4.3.4 and Figure 4.7 compared to rendered inside at a equivalent distance to the screen. Therefore, we constrained the virtual scene inside the sphere in our user studies (Chapter 5 and 6) as well as demonstrated applications [47, 147, 153].

While the enclosure nature provides a strong metaphor, the finite rendering volume is a constraint when applying this type of display to applications. This limitation is not unique to the spherical surface. Other enclosed surfaces, such as cube and cylinder, are also subject to it. Due to this constraint, it is better for this type of display to be applied for targeted applications with exocentric tasks. During exocentric tasks, the position and orientation of objects are defined in an exocentric reference frame [17]. Object-centered applications, such as CAD, can potentially benefit from this type of display by presenting the model in the center of the display. Naturally, the model will be scaled based on the size of the display. A smaller display, such as the 12 inch prototype (Figure 4.2(left)), will be suitable for a single designer scenario, and a larger display, such as the 24 inch prototype (Figure 4.2(right)), will be suitable for a collaborative working scenario.

The targeted exocentric application domain is consistent with the non-immersive nature of FTVR displays, which provide an "outside-in" viewing experience, compared to immersive displays with an "inside-out" viewing ex-

perience, such as HMD and CAVE. Unlike immersive displays, which isolate the user by blocking out all physical surroundings, FTVR displays situate in the physical environment. It provides opportunities for users to interact with physical objects, such as note-taking, messaging or recording when users work in the virtual environment with existing physical tools. In a collaborative environment, FTVR displays allow direct communication between users. There are the major strength of FTVR displays compared to HMD. Therefore, despite of the lack of generalization for various applications, we believe Fishbowl VR displays can be applied for targeted applications with exocentric tasks and the requirement of interacting with the physical surrounding.

7.4 Limitations

Demographic of participants One of the limitations across all of our studies is the demographic of participants. Our participants generally came from a pool of engineering students. Hence, they might have higher than average technical skills and experiences. It is possible that they had more exposure to VR technologies and 3D displays than the general public. As such, it would be difficult to generalize our result to the common public. Yet, our conclusions from these studies stand for this type of users who could potentially use Fishbowl VR displays in the future. An issue related to the demographic is the sample size. Each experiment involved 12-17 participants, which is typical for human factor studies in the literature. Naturally, larger sample size would make the results more pronouncing. Yet, each experiment is a sub-study as part of an entire study which tried to answer the same research question but from different perspective. The triangulation from multiple metrics in sub-studies helps to build the internal validity from different perspectives.

Controlled laboratory experiments Another limitation is that all of our studies were short-term controlled laboratory experiments, which may not reflect the viewing experience under natural conditions when people use the system in a relaxing and comfortable way. We tried to address this issue by bringing the display to conferences and having attendees to try our experimental stimuli in demo sessions. However, the observations were subjective and potentially biased towards our display given the nature of live-demo in conferences.

Abstract task design An issue related to the study design is the design of tasks used in our studies, including a size matching task, depth judgement task and size judgement task. Different from real tasks in 3D applications, these tasks are abstract and artificial, which may not directly reflect users' performance in 3D applications. However, it is important to keep the task simple and abstract in controlled experiments to ensure the outcome is only associated with the independent variables. Though these perceptual abstract tasks differentiate from real tasks, to some degree, they imitate general 3D tasks in applications as a combination of depth, size and orientation perception of objects. For example, the size matching task in Chapter 5 reflects user's ability to perceive and adjust the scale of 3D objects which is common in computer aided design systems; the spatial impression task in Chapter 6 imitates walk-around visualization experience when 3D displays are used as an attention-drawing showcase or a casual entertainment device.

Experimental stimuli A more specific issue, relevant to Chapter 5 and 6, is that we used spherical stimuli across different visual tasks. We chose the spherical shape of stimuli in our study due to its isotropic property, which has been commonly used in the literature of perceptual studies [73, 89, 101]. As our display is also spherical, it remains to be a question that whether stimuli with different shapes such as a cube would reproduce similar results. The same type of study will need to be carried out again for stimuli with other shapes on the Fishbowl VR display. It is also worth noting the discrepancy between the visual quality of experimental stimuli and the *visual fidelity* mentioned in this dissertation. There has been tremendous effort in computer graphics to develop photo-realistic rendering from the perspective of graphic content. Our goal of improving the *visual fidelity* takes a different perspective via studying the technical and perceptual factors of 3D displays. Therefore the experimental stimuli do not necessarily require the photo-realistic high quality of graphic rendering. However, future experiments could consider to include the variation of visual stimuli used in the study, such as using different textures or shading models, and investigate the potential interaction between the visual quality of stimuli and the display factors such as the screen shape in Chapter 5 and the movement type in Chapter 6.

7.5 Directions

Short term directions for the research components of this dissertation are discussed at the end of each chapter. Here we discuss broader directions for supporting perceptual consistency as well as the future potential of Fishbowl VR displays.

Evaluating spatial perception beyond controlled laboratory studies. As discussed in Section 7.4, our findings are based on a small group of users in a series of short-term controlled laboratory studies, which is common for perception studies in the domain of HCI. This type of study design provides strong internal validity and helps us to explore the complexity of spatial perception with specific insights under controlled conditions. One challenge when applying the results to real-world use cases is that the experimental tasks may not be equivalent to real-world tasks that could potentially benefit from improved spatial perception. Therefore, evaluating spatial perception in natural viewing conditions, such as in a field study, would provide data on the effective spatial perception supported by Fishbowl VR displays. In Chapter 6, we designed a walk-around visualization task which simulates the viewing experience when users browse the 3D data by walking around, and inspect points of interest by moving closer to the display. As head movements can be complicated with a combination of all possible directions, designing tasks which allows free movement would provide practical insights for realistic scenarios such as visual analytic, training and gaming. In addition to the movement, using real visual stimuli, such as scanned volume datesets [84] generated from ultrasound or computed tomography, would make the results more pronounced in use cases such as scientific visualization. These specific experimental design variations are examples of what we see as a promising future direction for understanding and improving spatial perception in 3D displays.

Comparing with other 3D displays. In Chapter 5, we compared the performance of a Fishbowl VR display to a traditional flat FTVR display and found the spherical display provided better spatial perception than the flat display. It is an open question as how the results can be applied to other 3D displays such as HMD and CAVE. While there exists studies which compared spatial perception from CAVE, FTVR and HMD [31, 41, 78, 106], they all used flat FTVR displays which supports limited field of view. Comparing the performance of Fishbowl VR displays with other 3D displays would

give us a broader understanding of the unique properties of this type of display. Furthermore, as the user performance is also task-related, comparing with other 3D displays helps to understand the visual fidelity of different displays and find appropriate applications the could benefit from Fishbowl VR displays.

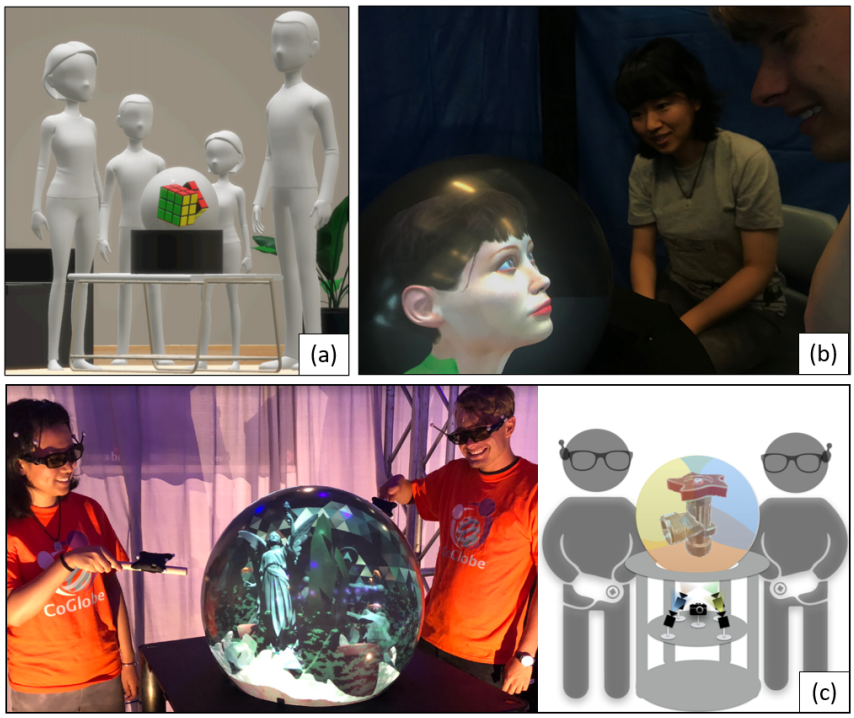


Figure 7.1: Examples of potential 3D applications using the Fishbowl VR display: (a) smart home assistant coupled with a spherical screen situated in the living room, (b) users making eye contact with a virtual agent [47] using a 12 inch prototype, and (c) collaborative tasks for two users to work together in the virtual environment using a 24 inch prototype [36].

Identifying appropriate 3D applications. Our efforts show potential for Fishbowl VR displays to provide consistent spatial perception in a way similar to the real world. It is unclear which applications can directly benefit from it. Here we discuss several possible areas which can potentially benefit from our findings and the two Fishbowl VR prototypes we developed (Figure 4.2). Naturally, smaller Fishbowl VR display, such as the 12 inch prototype, would benefit tasks and applications with personal use

and collaborative tasks would favor larger displays like the 24 inch prototype. Home assistant devices, such as Google Home [57] and Amazon Echo [52], have seen rapid adoption in the home over the past few years. So far, they have been limited to audio devices. As illustrated in Figure 7.1(a), the Fishbowl VR display extends the home assistant concept to use a volumetric 3D display as a focal point of the living room to create an embodied assistant with 360° visibility. Other augmented reality living room systems have been proposed, such as headset AR [61] and projection onto living room surfaces [116]. However, these systems lack the tangible nature of having a volumetric display that is part of the real living room. The spherical display is tangible and could show a virtual object situated in a real living environment. Furthermore, several perceptual studies have found spherical displays provided better eye contact [47] (shown in Figure 7.1(b)) and gaze perception [99] when interacting with virtual agents and avatars. This makes Fishbowl VR as appropriate displays applied in home assistant devices visible from all angles, such as using the 12 inch prototype.

Computer Aided Design (CAD) is another area that could benefit from consistent spatial perception. Realistic visual perception is crucial as users need to accurately perceive the size, shape and position of the 3D designs. Misperception of the distance and size of virtual models would decrease the performance of tasks ranging from 3D scaling to assembly in CAD. The ability to preserve size-constancy as shown in Chapter 5 allows users to effectively position, scale and design prototypes in the virtual environment. It is not just about making correct size judgments between two virtual entities, but also preserving consistent size perception between the real and virtual environment such that the perceived scale of a virtual model is consistent when it gets fabricated in the real world. In comparison to immersive 3D displays such as HMD, Fishbowl VR displays provide more consistent perception as it allows users to see virtual contents situated in the real world. In particular, in the collaborative environment when multiple users work together, a larger Fishbowl VR display, such as the 24 inch prototype, allows people to share their experience and visualize the same virtual entity situated in the real world. Through co-location, users can explore a virtual environment together while at the same time interacting with each other in the real world as shown in Figure 7.1(c). We suggest future studies to identify challenges and develop techniques to provide consistent perception for multiple co-located users simultaneously. We hope that our work on spatial perceptions shows potentials and help to pave the way for future applications in the area like CAD and home assistant devices.

7.6 Concluding Remarks

To conclude, this dissertation has presented a Fishbowl VR display with new techniques and perceptual studies to improve *visual fidelity* towards bridging the perceptual gap between the virtual and real world. We first formulated the visual error (Chapter 4) and created an automatic calibration approach to generate a seamless imagery with multiple projectors (Chapter 3), allowing us to subsequently investigate perceptual issues. Once the technical foundation is established, we demonstrated the superiority of the spherical screens as well as the perceptual duality via a series of human factor studies (Chapter 5 and 6), which play an important role in improving *visual fidelity* and enhancing the 3D experience. Providing realistic visualization in the virtual environment is difficult and inherently has considerable technical and perceptual challenges. While it was not possible to exhaust all possible aspects, we focus on the core technical and perceptual issues, allowing us to lay the groundwork for reconciling pixels with percept. Our work can thus serve as a baseline for future experiments, theoretical models, interaction techniques, and 3D applications to build on.

Bibliography

[1] Johnny Accot and Shumin Zhai. Beyond fitts' law: models for trajectory-based hci tasks. In *Proceedings of the ACM SIGCHI Conference on Human factors in computing systems*, pages 295–302. ACM, 1997. → page 14.

[2] Atif Ahmed, Rehan Hafiz, Muhammad Murtaza Khan, Yongju Cho, and Jihun Cha. Geometric correction for uneven quadric projection surfaces using recursive subdivision of bézier patches. *ETRI Journal*, 35(6):1115–1125, 2013. → page 28.

[3] JM Andrade and MG Estévez-Pérez. Statistical comparison of the slopes of two regression lines: a tutorial. *Analytica chimica acta*, 838:1–12, 2014. → page 76.

[4] Claudia Armbrüster, Marc Wolter, Torsten Kuhlen, Will Spijkers, and Bruno Fimm. Depth perception in virtual reality: distance estimations in peri-and extrapersonal space. *Cyberpsychology & Behavior*, 11(1):9–15, 2008. → page 22.

[5] Roland Arsenault and Colin Ware. The importance of stereo and eye-coupled perspective for eye-hand coordination in fish tank vr. *Presence: Teleoperators and Virtual Environments*, 13(5):549–559, 2004. → pages 21 and 115.

[6] Kevin W Arthur, Kellogg S Booth, and Colin Ware. Evaluating 3d task performance for fish tank virtual worlds. *ACM Transactions on Information Systems (TOIS)*, 11(3):239–265, 1993. → pages 13, 15, 21, 22, and 115.

[7] Ronald Azuma and Gary Bishop. Improving static and dynamic registration in an optical see-through hmd. In *Proceedings of the 21st annual conference on Computer graphics and interactive techniques*, pages 197–204. ACM, 1994. → pages 15 and 25.

[8] Martin Bauer. *Tracking Errors in Augmented Reality*. PhD thesis, Technische Universität München, 2007. → page 25.

[9] Hrvoje Benko, Ricardo Jota, and Andrew Wilson. Miragetable: freehand interaction on a projected augmented reality tabletop. In *Proceedings of the SIGCHI conference on human factors in computing systems*, pages 199–208. ACM, 2012. → pages 14, 23, and 66.

[10] Hrvoje Benko, Andrew D Wilson, and Ravin Balakrishnan. Sphere: multi-touch interactions on a spherical display. In *Proceedings of the 21st annual ACM symposium on User interface software and technology*, pages 77–86. ACM, 2008. → pages 16, 29, 63, 64, and 83.

[11] Hrvoje Benko, Andrew D Wilson, and Federico Zannier. Dyadic projected spatial augmented reality. In *Proceedings of the 27th annual ACM symposium on User interface software and technology*, pages 645–655. ACM, 2014. → pages 4, 11, 12, 14, 15, 20, and 23.

[12] Francois Berard and Thibault Louis. The object inside: Assessing 3d examination with a spherical handheld perspective-corrected display. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*, pages 4396–4404. ACM, 2017. → pages 17 and 23.

[13] Oliver Bimber and Ramesh Raskar. *Spatial augmented reality: merging real and virtual worlds*. AK Peters/CRC Press, 2005. → page 14.

[14] Barry G Blundell and Adam J Schwarz. The classification of volumetric display systems: characteristics and predictability of the image space. *IEEE Transactions on Visualization and Computer Graphics*, 8(1):66–75, 2002. → page 16.

[15] John Bolton, Kibum Kim, and Roel Vertegaal. Snowglobe: a spherical fish-tank vr display. In *CHI'11 Extended Abstracts on Human Factors in Computing Systems*, pages 1159–1164. ACM, 2011. → page 16.

[16] John Bolton, Kibum Kim, and Roel Vertegaal. A comparison of competitive and cooperative task performance using spherical and flat displays. In *Proceedings of the ACM 2012 conference on Computer Supported Cooperative Work*, pages 529–538. ACM, 2012. → pages 16 and 64.

[17] Doug A Bowman, Sabine Coquillart, Bernd Froehlich, Michitaka Hirose, Yoshifumi Kitamura, Kiyoshi Kiyokawa, and Wolfgang Stuerzlinger. 3d user interfaces: New directions and perspectives. *IEEE computer graphics and applications*, 28(6):20–36, 2008. → pages 10, 17, and 117.

[18] Mark F Bradshaw, Andrew D Parton, and Richard A Eagle. The interaction of binocular disparity and motion parallax in determining perceived depth and perceived size. *Perception*, 27(11):1317–1331, 1998. → page 93.

[19] Mark F Bradshaw, Andrew D Parton, and Andrew Glennerster. The task-dependent use of binocular disparity and motion parallax information. *Vision research*, 40(27):3725–3734, 2000. → page 4.

[20] Gerd Bruder, Ferran Argelaguet, Anne-Héléne Olivier, and Anatole Lécuyer. Cave size matters: Effects of screen distance and parallax on distance estimation in large immersive display setups. *Presence: Teleoperators and Virtual Environments*, 25(1):1–16, 2016. → pages 22 and 79.

[21] Han Chen, Rahul Sukthankar, Grant Wallace, and Kai Li. Scalable alignment of large-format multi-projector displays using camera homography trees. In *Proceedings of the conference on Visualization'02*, pages 339–346. IEEE Computer Society, 2002. → pages 24 and 28.

[22] Han Chen, Rahul Sukthankar, Grant Wallace, and Kai Li. Scalable alignment of large-format multi-projector displays using camera homography trees. In *Proceedings of the conference on Visualization'02*, pages 339–346. IEEE Computer Society, 2002. → page 56.

[23] Yuqun Chen, Douglas W Clark, Adam Finkelstein, Timothy C Housel, and Kai Li. Automatic alignment of high-resolution multi-projector display using an un-calibrated camera. In *Proceedings of the conference on Visualization'00*, pages 125–130. IEEE Computer Society Press, 2000. → pages 24 and 28.

[24] Lung-Pan Cheng, Eyal Ofek, Christian Holz, Hrvoje Benko, and Andrew D Wilson. Sparse haptic proxy: Touch feedback in virtual environments using a general passive prop. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*, pages 3718–3728. ACM, 2017. → pages 11 and 12.

[25] Lung-Pan Cheng, Thijs Roumen, Hannes Rantzsch, Sven Köhler, Patrick Schmidt, Robert Kovacs, Johannes Jasper, Jonas Kemper, and Patrick Baudisch. Turkdeck: Physical virtual reality based on people. In *Proceedings of the 28th Annual ACM Symposium on User Interface Software & Technology*, pages 417–426. ACM, 2015. → pages 11 and 12.

[26] Nusrat Choudhury, Nicholas Gélinas-Phaneuf, Sébastien Delorme, and Rolando Del Maestro. Fundamentals of neurosurgery: virtual reality tasks for training and evaluation of technical skills. *World neurosurgery*, 80(5):e9–e19, 2013. → page 2.

[27] Zeynep Cipiloglu, Abdullah Bulbul, and Tolga Capin. A framework for enhancing depth perception in computer graphics. In *Proceedings of the 7th Symposium on Applied Perception in Graphics and Visualization*, pages 141–148. ACM, 2010. → page 3.

[28] Sarah H Creem-Regehr, Jeanine K Stefanucci, and William B Thompson. Perceiving absolute scale in virtual environments: How theory and application have mutually informed the role of body-based perception. In *Psychology of Learning and Motivation*, volume 62, pages 195–224. Elsevier, 2015. → pages 22 and 23.

[29] Carolina Cruz-Neira, Daniel J Sandin, and Thomas A DeFanti. Surround-screen projection-based virtual reality: the design and implementation of the cave. In *Proceedings of the 20th annual conference on Computer graphics and interactive techniques*, pages 135–142. ACM, 1993. → pages 12, 14, 25, 55, and 86.

[30] Michael Deering. High resolution virtual reality. *ACM SIGGRAPH Computer Graphics*, 26(2):195–202, 1992. → page 1.

[31] Cagatay Demiralp, Cullen D Jackson, David B Karelitz, Song Zhang, and David H Laidlaw. Cave and fishtank virtual-reality displays: A qualitative and quantitative comparison. *IEEE transactions on visualization and computer graphics*, 12(3):323–330, 2006. → pages 4, 13, and 120.

[32] JP Djajadiningrat, Gerda JF Smets, and CJ Overbeeke. Cubby: a multiscreen movement parallax display for direct manual manipulation. *Displays*, 17(3-4):191–197, 1997. → page 14.

[33] C BROWN Duane. Close-range camera calibration. *Photogramm. Eng*, 37(8):855–866, 1971. → pages 33, 145, and 146.

[34] Robert G Eggleston, William P Janson, and Kenneth A Aldrich. Virtual reality system effects on size-distance judgements in a virtual environment. In *Virtual Reality Annual International Symposium, 1996., Proceedings of the IEEE 1996*, pages 139–146. IEEE, 1996. → page 23.

[35] Kevin W Elner and Helen Wright. Phenomenal regression to the real object in physical and virtual worlds. *Virtual Reality*, 19(1):21–31, 2015. → pages 20, 23, 91, 105, and 111.

[36] Dylan Fafard, Ian Stavness, Martin Dechant, Regan Mandryk, Qian Zhou, and Sidney Fels. Ftvr in vr: Evaluation of 3d perception with a simulated volumetric fish-tank virtual reality display. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, pages 1–12, 2019. → pages 85, 86, 91, 96, 105, 116, and 121.

[37] Dylan Fafard, Qian Zhou, Chris Chamberlain, Georg Hagemann, Sidney Fels, and Ian Stavness. Design and implementation of a multi-person fish-tank virtual reality display. In *Proceedings of the 24th ACM Symposium on Virtual Reality Software and Technology*, VRST '18. ACM, 2018. → pages v, vi, 48, 50, 94, and 116.

[38] Dylan Brodie Fafard. A virtual testbed for fish-tank virtual reality: Improving calibration with a virtual-in-virtual display. Master's thesis, University of Saskatchewan, 2019. → page 52.

[39] Gabriel Falcao, Natalia Hurtos, and Joan Massich. Plane-based calibration of a projector-camera system. *VIBOT master*, 9(1):1–12, 2008. → pages 32 and 145.

[40] G. E. Favalora. Volumetric 3d displays and application infrastructure. *Computer*, 38(8):37–44, 2005. → pages 3, 11, 12, and 16.

[41] Andrew Forsberg, Michael Katzourin, Kristi Wharton, Mel Slater, et al. A comparative study of desktop, fishtank, and cave systems for the exploration of volume rendered confocal data sets. *IEEE Transactions on Visualization and Computer Graphics*, 14(3):551–563, 2008. → pages 4 and 120.

[42] Davide Gadia, Gianfranco Garipoli, Cristian Bonanomi, Luigi Albani, and Alessandro Rizzi. Assessing stereo blindness and stereo acuity on digital displays. *Displays*, 35(4):206–212, 2014. → page 93.

[43] Michael N Geuss, Jeanine K Stefanucci, Sarah H Creem-Regehr, and William B Thompson. Effect of viewing plane on perceived distances in real and virtual environments. *Journal of Experimental Psychology: Human Perception and Performance*, 38(5):1242, 2012. → page 22.

[44] James J Gibson. The information available in pictures. *Leonardo*, 4(1):27–35, 1971. → pages 20 and 86.

[45] Tovi Grossman and Ravin Balakrishnan. An evaluation of depth perception on volumetric displays. In *Proceedings of the working conference on Advanced visual interfaces*, pages 193–200. ACM, 2006. → pages 16, 22, and 66.

[46] Ralph Norman Haber. How we perceive depth from flat pictures: The inherent dual reality in pictorial art enables us to perceive a scene as three dimensional at the same time we see that the painting or photograph is actually flat. *American Scientist*, 68(4):370–380, 1980. → pages 20 and 86.

[47] Georg Hagemann, Qian Zhou, Ian Stavness, and Sidney Fels. Investigating spherical fish tank virtual reality displays for establishing realistic eye-contact. In *2019 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*, pages 950–951. IEEE, 2019. → pages 117, 121, and 122.

[48] Richard Hartley and Andrew Zisserman. *Multiple view geometry in computer vision*. Cambridge university press, 2003. → pages 32, 36, 37, 56, and 145.

[49] Michael Harville, Bruce Culbertson, Irwin Sobel, Dan Gelb, Andrew Fitzhugh, and Donald Tanguay. Practical methods for geometric and photometric correction of tiled projector. In *Computer Vision and Pattern Recognition Workshop, 2006. CVPRW'06. Conference on*, pages 5–5. IEEE, 2006. → pages 24 and 28.

[50] Anuruddha Hettiarachchi and Daniel Wigdor. Annexing reality: Enabling opportunistic use of everyday objects as tangible proxies in augmented reality. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, pages 1957–1967. ACM, 2016. → page 11.

[51] Richard L Holloway. Registration error analysis for augmented reality. *Presence: Teleoperators and Virtual Environments*, 6(4):413–432, 1997. → pages 25, 47, 60, and 105.

[52] Amazon Inc. Amazon echo, 2019. `https://en.wikipedia.org/wiki/Amazon_Echo`. → page 122.

[53] ASUS Inc. Asus p2b portable led projector, 2020. `https://www.asus.com/ca-en/Commercial-Projectors/P2B/`. → pages 40, 49, and 65.

[54] Facebook Inc. Oculus, 2020. `https://www.oculus.com`. → pages 10, 12, and 15.

[55] FLIR Systems Inc. Flea3 usb3 — flir systems, 2020. `https://www.flir.com/products/flea3-usb3?model=FL3-U3-13Y3M-C`. → page 41.

[56] Fujifilm Inc. Fujinon 3.8-13mm f1.4 — fujinon lens, 2020. `https://www.fujifilm.com/products/optical_devices/`. → page 41.

[57] Google Inc. Google home, 2019. `https://en.wikipedia.org/wiki/Google_Home`. → page 122.

[58] HTC Inc. Vive, 2020. `https://www.vive.com`. → page 15.

[59] Matlab Inc. Symbolic math toolbox from matlab, 2019. `https://www.mathworks.com/help/symbolic/generate-c-or-fortran-code.html`. → page 148.

[60] Microsoft Inc. Kinect, 2020. `https://developer.microsoft.com/en-us/windows/kinect`. → pages 25 and 59.

[61] Microsoft Inc. Microsoft HoloLens — mixed reality technology for business, 2020. `https://www.microsoft.com/en-us/hololens`. → pages 10, 12, 15, and 122.

[62] NaturalPoint Inc. Optitrack, 2020. `https://optitrack.com`. → pages 25, 50, 60, and 94.

[63] OPTOMA Inc. Optoma mobile led projector ml750, 2020. `https://www.optoma.com/us/product/ml750/`. → page 50.

[64] Polhemus Inc. Polhemus motion tracking systems, 2020. `https://polhemus.com/motion-tracking/overview/`. → pages 40 and 50.

[65] Pufferfish Inc. Pufferfish interactive spherical displays, 2020. `https://pufferfishdisplays.com`. → page 16.

[66] VICON Inc. Vicon, 2019. `https://www.vicon.com`. → page 25.

[67] XPAND Inc. Xpand 3d glasses lite rf, 2020. `http://xpandvision.com/products/xpand-3d-glasses-lite-ir-rf/`. → page 50.

[68] Brett Jones, Rajinder Sodhi, Michael Murdock, Ravish Mehra, Hrvoje Benko, Andrew Wilson, Eyal Ofek, Blair MacIntyre, Nikunj Raghuvanshi, and Lior Shapira. Roomalive: magical experiences enabled by scalable, adaptive projector-camera units. In *Proceedings of the 27th annual ACM symposium on User interface software and technology*, pages 637–644. ACM, 2014. → page 14.

[69] Brett R Jones, Hrvoje Benko, Eyal Ofek, and Andrew D Wilson. Illumiroom: peripheral projected illusions for interactive experiences. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 869–878. ACM, 2013. → page 14.

[70] J Adam Jones, J Edward Swan II, Gurjot Singh, Eric Kolstad, and Stephen R Ellis. The effects of virtual reality, augmented reality, and motion parallax on egocentric depth perception. In *Proceedings of the 5th symposium on Applied perception in graphics and visualization*, pages 9–14. ACM, 2008. → page 22.

[71] Naoki Kawakami, Masahiko INAMI, Taro MAEDA, and Susumu TACHI. Proposal for the object-oriented display: The design and implementation of the media3. In *Proceedings of ICAT*, volume 97, pages 57–62, 1997. → page 14.

[72] Daniel F Keefe, Daniel Acevedo Feliz, Tomer Moscovich, David H Laidlaw, and Joseph J LaViola Jr. Cavepainting: a fully immersive 3d artistic medium and interactive experience. In *Proceedings of the 2001 symposium on Interactive 3D graphics*, pages 85–93. Citeseer, 2001. → pages 11 and 12.

[73] Jonathan W Kelly, Lucia A Cherep, and Zachary D Siegel. Perceived space in the htc vive. *ACM Transactions on Applied Perception (TAP)*, 15(1):2, 2017. → pages 64, 91, 104, 110, and 119.

[74] Jonathan W Kelly, Lisa S Donaldson, Lori A Sjolund, and Jacob B Freiberg. More than just perception–action recalibration: Walking

through a virtual environment causes rescaling of perceived space. *Attention, Perception, & Psychophysics*, 75(7):1473–1485, 2013. → pages 22, 23, and 82.

[75] Jonathan W Kelly, Jack M Loomis, and Andrew C Beall. Judgments of exocentric direction in large-scale space. *Perception*, 33(4):443–454, 2004. → page 22.

[76] Robert V Kenyon, Daniel Sandin, Randall C Smith, Richard Pawlicki, and Thomas Defanti. Size-constancy in the cave. *Presence: Teleoperators and Virtual Environments*, 16(2):172–187, 2007. → pages 64, 66, 67, 70, 75, 81, 82, and 87.

[77] Kibum Kim, John Bolton, Audrey Girouard, Jeremy Cooperstock, and Roel Vertegaal. Telehuman: effects of 3d perspective on gaze and pose estimation with a life-size cylindrical telepresence pod. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 2531–2540. ACM, 2012. → page 63.

[78] Kwanguk Kim, M Zachary Rosenthal, David Zielinski, and Rachel Brady. Comparison of desktop, head mounted display, and six wall fully immersive systems using a stressful task. In *2012 IEEE Virtual Reality Workshops (VRW)*, pages 143–144. IEEE, 2012. → page 120.

[79] Volodymyr V Kindratenko. A survey of electromagnetic position tracker calibration techniques. *Virtual Reality*, 5(3):169–182, 2000. → page 26.

[80] Georg Klein. *Visual tracking for augmented reality*. PhD thesis, University of Cambridge City of Cambridge, United Kingdom, 2006. → page 25.

[81] Sirisilp Kongsilp and Matthew N Dailey. Motion parallax from head movement enhances stereoscopic displays by improving presence and decreasing visual fatigue. *Displays*, 49:72–79, 2017. → pages 13 and 85.

[82] Robert Kooima. Generalized perspective projection. *J. Sch. Electron. Eng. Comput. Sci*, 2009. → page 54.

[83] Gregory Kramida. Resolving the vergence-accommodation conflict in head-mounted displays. *IEEE transactions on visualization and computer graphics*, 22(7):1912–1931, 2015. → page 15.

[84] B. Laha, D. A. Bowman, and J. J. Socha. Effects of vr system fidelity on analyzing isosurface visualization of volume datasets. *IEEE Transactions on Visualization and Computer Graphics*, 20(4):513–522, 2014. → pages 18, 19, and 120.

[85] Billy Lam, Yichen Tang, Ian Stavness, and Sidney Fels. A 3d cubic puzzle in pcubee. In *3D User Interfaces (3DUI), 2011 IEEE Symposium on*, pages 135–136. IEEE, 2011. → page 14.

[86] Billy Shiu Fai Lam. Evaluation of a tangible outward-facing geometric display. Master's thesis, UNIVERSITY OF BRITISH COLUMBIA (Vancouver, 2011. → page 4.

[87] Joseph J LaViola, Andrew S Forsberg, John Huffman, and Andrew Bragdon. Poster: Effects of head tracking and stereo on non-isomorphic 3d rotation. In *2008 IEEE Symposium on 3D User Interfaces*, pages 155–156. IEEE, 2008. → page 13.

[88] Ivan KY Li, Edward M Peek, Burkhard C Wünsche, and Christof Lutteroth. Enhancing 3d applications using stereoscopic 3d and motion parallax. In *Proceedings of the Thirteenth Australasian User Interface Conference-Volume 126*, pages 59–68. Australian Computer Society, Inc., 2012. → page 13.

[89] Sally A Linkenauger, Markus Leyrer, Heinrich H Bülthoff, and Betty J Mohler. Welcome to wonderland: The influence of the size and shape of a virtual hand on the perceived size and shape of virtual objects. *PloS one*, 8(7):e68594, 2013. → page 119.

[90] Kok-Lim Low, Greg Welch, Anselmo Lastra, and Henry Fuchs. Life-sized projector-based dioramas. In *Proceedings of the ACM symposium on Virtual reality software and technology*, pages 93–101. ACM, 2001. → page 12.

[91] Xun Luo, Robert Kenyon, Derek Kamper, Daniel Sandin, and Thomas DeFanti. The effects of scene complexity, stereovision, and motion parallax on size constancy in a virtual environment. In *Virtual Reality Conference, 2007. VR'07. IEEE*, pages 59–66. IEEE, 2007. → pages 4, 23, 70, 93, and 104.

[92] Blair MacIntyre, Enylton Machado Coelho, and Simon J Julier. Estimating and adapting to registration errors in augmented reality sys-

tems. In *Virtual Reality, 2002. Proceedings. IEEE*, pages 73–80. IEEE, 2002. → page 25.

[93] Ryan P McMahan, Doug A Bowman, David J Zielinski, and Rachael B Brady. Evaluating display fidelity and interaction fidelity in a virtual reality game. *IEEE transactions on visualization and computer graphics*, 18(4):626–633, 2012. → pages 18 and 19.

[94] Paul Milgram and Fumio Kishino. A taxonomy of mixed reality visual displays. *IEICE TRANSACTIONS on Information and Systems*, 77(12):1321–1329, 1994. → pages xiii, 11, 12, 15, 17, and 18.

[95] Mark R Mine, Jeroen Van Baar, Anselm Grundhofer, David Rose, and Bei Yang. Projection-based augmented reality in disney theme parks. *Computer*, 45(7):32–40, 2012. → page 14.

[96] Betty J Mohler, Sarah H Creem-Regehr, and William B Thompson. The influence of feedback on egocentric distance judgments in real and virtual environments. In *Proceedings of the 3rd symposium on Applied perception in graphics and visualization*, pages 9–14. ACM, 2006. → page 22.

[97] Daniel Moreno and Gabriel Taubin. Simple, accurate, and robust projector-camera calibration. In *2012 Second International Conference on 3D Imaging, Modeling, Processing, Visualization & Transmission*, pages 464–471. IEEE, 2012. → page 37.

[98] opencv. opencv, 2019. `https://opencv.org/`. → pages 41 and 145.

[99] Ye Pan and Anthony Steed. Preserving gaze direction in teleconferencing using a camera array and a spherical display. In *3DTV-Conference: The True Vision-Capture, Transmission and Display of 3D Video (3DTV-CON), 2012*, pages 1–4. IEEE, 2012. → pages 16 and 122.

[100] Anne Parent. A virtual environment task analysis workbook for the creation and evaluation of virtual art exhibits. *National Research Council of Canada-Reports-ERB*, 1998. → page 101.

[101] Etienne Peillard, Thomas Thebaud, Jean-Marie Normand, Ferran Argelaguet, Guillaume Moreau, and Anatole Lécuyer. Virtual objects look farther on the sides: The anisotropy of distance perception in virtual reality. In *IEEE VR Conference 2019*, 2019. → page 119.

[102] Fastrak Polhemus. 3space fastrak user's manual. *F. Polhemus Inc., Colchester, VT*, 1993. → pages 25, 50, and 60.

[103] Kevin Ponto, Michael Gleicher, Robert G Radwin, and Hyun Joon Shin. Perceptual calibration for immersive display environments. *IEEE transactions on visualization and computer graphics*, 19(4):691–700, 2013. → pages 23, 80, 81, and 105.

[104] David H Laidlaw Prabhat, Thomas F Banchoff, and Cullen D Jackson. Comparative evaluation of desktop and cave environments for learning hypercube rotations. Technical report, Brown University, 2005. → page 13.

[105] ALGLIB Project. Alglib, 2019. `https://www.alglib.net/`. → pages 41 and 147.

[106] Wen Qi, Russell M Taylor II, Christopher G Healey, and Jean-Bernard Martens. A comparison of immersive hmd, fish tank vr and fish tank with haptics displays for volume visualization. In *Proceedings of the 3rd Symposium on Applied Perception in Graphics and Visualization*, pages 51–58. ACM, 2006. → pages 4, 11, 13, 22, and 120.

[107] Qian Zhou. Geometric fish tank vr display calibration library, 2019. `https://github.com/GeometricFishTankVR/SphericalDisplayCalibration`. → page 148.

[108] Eric D Ragan, Doug A Bowman, Regis Kopper, Cheryl Stinson, Siroberto Scerbo, and Ryan P McMahan. Effects of field of view and visual complexity on virtual reality training effectiveness for a visual scanning task. *IEEE transactions on visualization and computer graphics*, 21(7):794–807, 2015. → pages 18 and 19.

[109] Eric D Ragan, Regis Kopper, Philip Schuchardt, and Doug A Bowman. Studying the effects of stereo, head tracking, and field of regard on a small-scale spatial judgment task. *IEEE transactions on visualization and computer graphics*, 19(5):886–896, 2012. → page 18.

[110] Andrew Raij, Gennette Gill, Aditi Majumder, Herman Towles, and Henry Fuchs. Pixelflex2: A comprehensive, automatic, casually-aligned multi-projector display. In *IEEE International Workshop on Projector-Camera Systems*, pages 203–211. Nice, France, 2003. → pages 24 and 28.

[111] Andrew Raij and Marc Pollefeys. Auto-calibration of multi-projector display walls. In *Pattern Recognition, 2004. ICPR 2004. Proceedings of the 17th International Conference on*, volume 1, pages 14–17. IEEE, 2004. → page 37.

[112] Ramesh Raskar. Immersive planar display using roughly aligned projectors. In *Virtual Reality, 2000. Proceedings. IEEE*, pages 109–115. IEEE, 2000. → pages 24 and 28.

[113] Ramesh Raskar, Michael S Brown, Ruigang Yang, Wei-Chao Chen, Greg Welch, Herman Towles, Brent Scales, and Henry Fuchs. Multi-projector displays using camera-based registration. In *Visualization'99. Proceedings*, pages 161–522. IEEE, 1999. → pages 24, 28, 31, 35, 40, 41, and 53.

[114] Ramesh Raskar, Jeroen van Baar, Paul Beardsley, Thomas Willwacher, Srinivas Rao, and Clifton Forlines. Ilamps: Geometrically aware and self-configuring projectors. In *ACM SIGGRAPH 2006 Courses*, SIGGRAPH '06, page 7–es, New York, NY, USA, 2006. Association for Computing Machinery. → pages 12 and 14.

[115] Ramesh Raskar, Jeroen van Baar, Srinivas Rao, and Thomas Willwacher. Multi-projector imagery on curved surfaces. *Mitsubishi Electric Research Labs*, pages 1–8, 2003. → pages 24, 28, 38, 40, and 43.

[116] Ramesh Raskar, Greg Welch, Matt Cutts, Adam Lake, Lev Stesin, and Henry Fuchs. The office of the future: A unified approach to image-based modeling and spatially immersive displays. In *Proceedings of the 25th annual conference on Computer graphics and interactive techniques*, pages 179–188. ACM, 1998. → pages 14 and 122.

[117] Ramesh Raskar, Greg Welch, Kok-Lim Low, and Deepak Bandyopadhyay. Shader lamps: Animating real objects with image-based illumination. In *Rendering Techniques 2001*, pages 89–102. Springer, 2001. → page 14.

[118] Rebekka S Renner, Boris M Velichkovsky, and Jens R Helmert. The perception of egocentric distances in virtual environments-a review. *ACM Computing Surveys (CSUR)*, 46(2):23, 2013. → page 22.

[119] Brian Rogers and Maureen Graham. Motion parallax as an independent cue for depth perception. *Perception*, 8(2):125–134, 1979. → page 4.

[120] Behzad Sajadi and Aditi Majumder. Scalable multi-view registration for multi-projector displays on vertically extruded surfaces. In *Computer Graphics Forum*, volume 29, pages 1063–1072. Wiley Online Library, 2010. → pages 24, 38, and 43.

[121] Behzad Sajadi and Aditi Majumder. Automatic registration of multiprojector domes using a single uncalibrated camera. In *Computer Graphics Forum*, volume 30, pages 1161–1170. Wiley Online Library, 2011. → pages 14, 43, and 45.

[122] Behzad Sajadi and Aditi Majumder. Autocalibration of multiprojector cave-like immersive environments. *Visualization and Computer Graphics, IEEE Transactions on*, 18(3):381–393, 2012. → pages 14, 24, 38, and 43.

[123] Zachary D Siegel, Jonathan W Kelly, and Lucia A Cherep. Rescaling of perceived space transfers across virtual environments. *Journal of Experimental Psychology: Human Perception and Performance*, 43(10):1805, 2017. → pages 67, 70, and 91.

[124] Adalberto L Simeone, Eduardo Velloso, and Hans Gellersen. Substitutional reality: Using the physical environment to design virtual reality experiences. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*, pages 3307–3316. ACM, 2015. → page 12.

[125] Duncan Smith and Sameer Singh. Approaches to multisensor data fusion in target tracking: A survey. *IEEE transactions on knowledge and data engineering*, 18(12):1696–1710, 2006. → page 25.

[126] Ian Stavness, Billy Lam, and Sidney Fels. pcubee: a perspective-corrected handheld cubic display. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 1381–1390. ACM, 2010. → pages 4, 11, 14, 23, and 105.

[127] Ian Stavness, Florian Vogt, and Sidney Fels. Cubee: a cubic 3d display for physics-based interaction. In *ACM SIGGRAPH 2006 Sketches*, page 165. ACM, 2006. → pages 14 and 63.

[128] Jeanine K Stefanucci, Sarah H Creem-Regehr, William B Thompson, David A Lessard, and Michael N Geuss. Evaluating the accuracy of size perception on screen-based displays: Displayed objects appear smaller than real objects. *Journal of Experimental Psychology: Applied*, 21(3):215, 2015. → pages 1, 19, 23, 77, 96, 104, and 105.

[129] Brett Stevens, Jennifer Jerrams-Smith, David Heathcote, and David Callear. Putting the virtual into reality: Assessing object-presence with projection-augmented models. *Presence: Teleoperators & Virtual Environments*, 11(1):79–92, 2002. → page 2.

[130] G Strang. *Introduction to Applied Mathematics*. Wellesley-Cambridge, 1986. → pages 33, 57, and 145.

[131] Ivan E Sutherland. A head-mounted three dimensional display. In *Proceedings of the December 9-11, 1968, fall joint computer conference, part I*, pages 757–764. ACM, 1968. → page 15.

[132] Thiago Teixeira, Gershon Dublon, and Andreas Savvides. A survey of human-sensing: Methods for detecting presence, count, location, track, and identity. *ACM Computing Surveys*, 5(1):59–69, 2010. → page 25.

[133] F Teubl, Celso S Kurashima, MC Cabral, Roseli D Lopes, Junia C Anacleto, Marcelo K Zuffo, and Sidney Fels. Spheree: An interactive perspective-corrected spherical 3d display. In *3DTV-Conference: The True Vision-Capture, Transmission and Display of 3D Video (3DTV-CON), 2014*, pages 1–4. IEEE, 2014. → pages 16, 29, and 64.

[134] Fernando Teubl, Celso Kurashima, Marcio Cabral, and Marcelo Zuffo. Fastfusion: A scalable multi-projector system. In *Virtual and Augmented Reality (SVR), 2012 14th Symposium on*, pages 26–35. IEEE, 2012. → pages 14, 24, and 28.

[135] Unity. Unity technologies inc., 2020. `https://unity.com/`. → pages 53, 65, and 94.

[136] D Vishwanath. Information in surface and depth perception: Reconciling pictures and reality. *Perception beyond inference*, pages 201–240, 2011. → page 20.

[137] Andrejs Vorozcovs, Wolfgang Stürzlinger, Andrew Hogue, and Robert S Allison. The hedgehog: a novel optical tracking method

for spatially immersive displays. *Presence*, 15(1):108–121, 2006. → page 26.

[138] Andrew Wagemakers, Dylan Fafard, and Ian Stavness. Interactive visual calibration of volumetric head-tracked 3d displays. In *2017 SIGCHI Conference on Human Factors in Computing Systems (CHI '17)*. ACM, 2017, to appear. → pages 51, 93, and 94.

[139] Andrew John Wagemakers, Dylan Brodie Fafard, and Ian Stavness. Interactive visual calibration of volumetric head-tracked 3d displays. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*, pages 3943–3953. ACM, 2017. → pages 14 and 46.

[140] Mark Wagner. Sensory and cognitive explanations for a century of size constancy research. *Visual experience: Sensation, cognition, and constancy*, pages 1–35, 2012. → page 64.

[141] L. C. Wanger, D. P. Greenberg, and J. A. Ferwerda. Perceiving spatial relationships in computer-generated images. *IEEE Computer Graphics and Applications*, 12(03):44–51, 54–58, may 1992. → page 21.

[142] Colin Ware. *Information visualization: perception for design*. Elsevier, 2012. → pages 3, 4, 15, 20, 21, 64, 77, 85, 86, 91, 105, and 106.

[143] Colin Ware, Kevin Arthur, and Kellogg S Booth. Fish tank virtual reality. In *Proceedings of the INTERACT'93 and CHI'93 conference on Human factors in computing systems*, pages 37–42. ACM, 1993. → pages 12 and 13.

[144] Colin Ware and Glenn Franck. Evaluating stereo and motion cues for visualizing information nets in three dimensions. *ACM Transactions on Graphics (TOG)*, 15(2):121–140, 1996. → pages 13, 21, and 22.

[145] Christian Weichel, Manfred Lau, David Kim, Nicolas Villar, and Hans W Gellersen. Mixfab: a mixed-reality environment for personal fabrication. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 3855–3864. ACM, 2014. → page 2.

[146] Wikipedia contributors. Spinning dancer — Wikipedia, the free encyclopedia, 2019. `https://en.wikipedia.org/w/index.php?title=Spinning_Dancer&oldid=916380798`. → page 21.

[147] Fan Wu, Qian Zhou, Kyoungwon Seo, Toshiro Kashiwaqi, and Sidney Fels. I got your point: An investigation of pointing cues in a spherical fish tank virtual reality display. In *2019 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*, pages 1237–1238. IEEE, 2019. → page 117.

[148] Xu Xu and Raymond W McGorry. The validity of the first and second generation microsoft kinect for identifying joint center locations during static postures. *Applied ergonomics*, 49:47–54, 2015. → pages 59 and 60.

[149] Alper Yilmaz, Omar Javed, and Mubarak Shah. Object tracking: A survey. *Acm computing surveys (CSUR)*, 38(4):13, 2006. → page 25.

[150] Suya You, Ulrich Neumann, and Ronald Azuma. Orientation tracking for outdoor augmented reality registration. *Computer Graphics and Applications, IEEE*, 19(6):36–42, 1999. → page 25.

[151] Zhengyou Zhang. A flexible new technique for camera calibration. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 22(11):1330–1334, 2000. → pages 32 and 145.

[152] Qian Zhou, Georg Hagemann, Dylan Fafard, Ian Stavness, and Sidney Fels. An evaluation of depth and size perception on a spherical fish tank virtual reality display. *IEEE Transactions on Visualization and Computer Graphics*, 25(5):2040–2049, 2019. → pages vi, 91, and 105.

[153] Qian Zhou, Georg Hagemann, Sidney Fels, Dylan Fafard, Andrew Wagemakers, Chris Chamberlain, and Ian Stavness. Coglobe: a co-located multi-person ftvr experience. In *ACM SIGGRAPH 2018 Emerging Technologies*, page 5. ACM, 2018. → pages vi, vii, 97, and 117.

[154] Qian Zhou, Gregor Miller, Kai Wu, Daniela Correa, and Sidney Fels. Automatic calibration of a multiple-projector spherical fish tank vr display. In *Applications of Computer Vision (WACV), 2017 IEEE Winter Conference on*, pages 1072–1081. IEEE, 2017. → pages v and 94.

[155] Qian Zhou, Gregor Miller, Kai Wu, Ian Stavness, and Sidney Fels. Analysis and practical minimization of registration error in a spherical fish tank virtual reality system. In *Asian Conference on Computer Vision*, pages 519–534. Springer, 2016. → pages v, 28, 29, 38, 41, and 43.

[156] Qian Zhou, Fan Wu, Sidney Fels, and Ian Stavness. Closer object looks smaller: Investigating the duality of size perception in a spherical fish tank vr display. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, pages 1–9. ACM, 2020. → page vi.

[157] Qian Zhou, Kai Wu, Gregor Miller, Ian Stavness, and Sidney Fels. 3dps: An auto-calibrated three-dimensional perspective-corrected spherical display. In *Virtual Reality (VR), 2017 IEEE*, pages 455–456. IEEE, 2017. → pages vi, vii, and 63.

[158] Yi Zhou, Shuangjiu Xiao, Ning Tang, Zhiyong Wei, and Xu Chen. Pmomo: projection mapping on movable 3d object. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, pages 781–790. ACM, 2016. → pages 11, 12, and 14.

Appendix A

List of Publications

A.1 Journal Publication

Qian Zhou, Georg Hagemann, Dylan Fafard, Ian Stavness, and Sidney Fels. An evaluation of depth and size perception on a spherical fish tank virtual reality display. *IEEE Transactions on Visualization and Computer Graphics*, 25(5): 2040–2049, 2019.

A.2 Conference Publication

Qian Zhou, Fan Wu, Sidney Fels, and Ian Stavness. Closer object looks smaller: Investigating the duality of size perception in a spherical fish tank vr display. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, pages 1–9. ACM, 2020.

Qian Zhou, Fan Wu, Ian Stavness, and Sidney Fels. Match the cube: Investigation of the head-coupled input with a spherical fish tank virtual reality display. In *2019 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*, pages 1281–1282. IEEE, 2019.

Qian Zhou, Georg Hagemann, Sidney Fels, Dylan Fafard, Andrew Wagemakers, Chris Chamberlain, and Ian Stavness. Coglobe: a co-located multi-person ftvr experience. In *ACM SIGGRAPH 2018 Emerging Technologies*, page 5. ACM, 2018.

Dylan Fafard, **Qian Zhou**, Chris Chamberlain, Georg Hagemann, Sidney Fels, and Ian Stavness. Design and implementation of a multi-person fish tank virtual reality display. In *Proceedings of the 24th ACM Symposium on Virtual Reality Software and Technology (VRST)*, page 5. ACM, 2018.

Dylan Fafard, Andrew Wagemakers, Ian Stavness, **Qian Zhou**, Gregor Miller, and Sidney S Fels. Calibration methods for effective fish

tank VR in multi-screen displays. In *Proceedings of the 2017 CHI Conference Extended Abstracts on Human Factors in Computing Systems*, pages 373–376. ACM, 2017.

Qian Zhou, Gregor Miller, Kai Wu, Daniela Correa, and Sidney Fels. Automatic calibration of a multiple-projector spherical fish tank VR display. In *2017 IEEE Winter Conference on Applications of Computer Vision (WACV)*, pages 1072–1081. IEEE, 2017.

Qian Zhou, Kai Wu, Gregor Miller, Ian Stavness, and Sidney Fels. 3dps: An auto-calibrated three-dimensional perspective-corrected spherical display. In *Virtual Reality (VR)*, pages 455–456. IEEE, 2017.

Qian Zhou, Gregor Miller, Kai Wu, Ian Stavness, and Sidney Fels. Analysis and practical minimization of registration error in a spherical fish tank virtual reality system. In *Asian Conference on Computer Vision (ACCV)*, pages 519–534. Springer, 2016.

A.3 Research Talk

March 2018 - Spherical Fish Tank VR

- Emerging Media Lab, Vancouver, BC, Canada

August 2018 - CoGlobe: A Co-Located Multi-Person FTVR Experience

- Siggraph 2018, Vancouver, BC, Canada

A.4 Additional Publication

Qian Zhou, Sarah Sykes, Sidney Fels, and Kenrick Kin. Gripmarks: Using hand grips to transform in-hand objects into mixed reality input. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, pages 1–11. ACM, 2020.

Fan Wu, **Qian Zhou**, Kyoungwon Seo, Toshiro Kashiwaqi, and Sidney Fels. I got your point: An investigation of pointing cues in a spherical fish tank virtual reality display. In *2019 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*, pages 1237–1238. IEEE, 2019.

Georg Hagemann, **Qian Zhou**, Ian Stavness, and Sidney Fels. Investigating spherical fish tank virtual reality displays for establishing realistic eye-contact. In *2019 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*, pages 950–951. IEEE, 2019.

Toshiro Kashiwagi, Kaoru Sumi, Sidney Fels, **Qian Zhou**, and Fan Wu. Crystal palace: Merging virtual objects and physical handheld tools. In *2019 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*, pages 1411–1412. IEEE, 2019.

Dylan Fafard, Ian Stavness, Martin Dechant, Regan Mandryk, **Qian Zhou**, and Sidney Fels. FTVR in VR: Evaluation of 3D perception with a simulated volumetric fish-tank virtual reality display. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, page 533. ACM, 2019.

Georg Hagemann, **Qian Zhou**, Ian Stavness, Oky Dicky Ardiansyah Prima, and Sidney S Fels. Here's looking at you: A spherical FTVR display for realistic eye-contact. In *Proceedings of the 2018 ACM International Conference on Interactive Surfaces and Spaces (ISS)*, pages 357–362. ACM, 2018.

Appendix B

Multi-Projector Display Calibration Optimization

This chapter describes the implementation and functions for the multi-projector spherical display calibration in Chapter 3. The problem of the display calibration can be formulated as the estimation of a set of parameters $\vec{p}$:

$$\vec{p} = (\vec{p}_c,\ \vec{p}_s,\ \vec{p}_{p_1},\ ...,\ \vec{p}_{p_N}) \tag{B.1}$$

Assuming we have one camera and N projectors. Camera parameters $\vec{p}_c$ have 9 degree of freedom (DOF): 4 for the focal length and the principle point; 5 for lens distortion [33]. Each projector has parameters $\vec{p}_{p_i}$ with 10 DOF: 4 for the focal length and the principle point; 3 for rotation and 3 for translation. Sphere parameters $\vec{p}_s$ have 4 DOF: 3 for the center position and 1 for radius.

As shown in Figure 3.3, we first obtain the initial guess of camera parameters $\vec{p}_c$ and projector parameters $\vec{p}_{p_i}$. Camera parameters (intrinsic) are estimated based on [151] implemented with OpenCV [98]. Projector parameters (intrinsic) are estimated based on [39] implemented with OpenCV. Projector parameters (extrinsic) are estimated based on the canonical solution of a two view geometry described in [48] (page 257-260). Sphere parameters $\vec{p}_s$ are estimated based on a weighted linear least square solution [130].

As the initial guess of all parameters has been obtained, we describe the nonlinear optimization with the error function and the Jacobian matrix. For a pixel j located at $\vec{x}_{p_{ij}}$ in the image plane of projector P_i, a ray is back-projected and intersects with the sphere at the 3D point $\vec{X}_{ij}$. The back-projection and ray-sphere intersection can be expressed as a function f based on variables $\vec{p}_{p_i}$ and $\vec{p}_s$:

$$\vec{X}_{ij} = f(\vec{x}_{p_{ij}}; \vec{p}_{p_i}, \vec{p}_s) \tag{B.2}$$

We now describe the formulation of $\vec{X}_{ij}$. Let KK_{p_i} represent the intrinsic matrix (3x3), R_i and T_i represent the rotation (3x3) and translation matrix (3x1) of projector P_i. For a pixel j with the 2D coordinate of $\vec{x}_{p_{ij}}$, the back-projection of a ray starting at the pixel j can be formulated as:

$$\begin{aligned} Ray(\lambda) &= \begin{pmatrix} x_{ij} \\ y_{ij} \\ z_{ij} \end{pmatrix} = \vec{C_i} + \vec{V_{ij}}\lambda \\ \vec{C_i} &= -R_i^T T_i \\ \vec{V_{ij}} &= (KK_{p_i} R_i)^{-1} \vec{x}_{p_{ij}} \end{aligned} \tag{B.3}$$

Substituting equation B.3 into the sphere equation with (a, b, c) as the center of the sphere and r as the radius:

$$(x_{ij} - a)^2 + (y_{ij} - b)^2 + (z_{ij} - c)^2 = r^2 \tag{B.4}$$

We can solve for $\vec{X}_{ij} = (x_{ij}, y_{ij}, z_{ij})^T$. Let the 3D point $\vec{X}_{ij}$ captured by the camera at pixel $\vec{x}_{c_{ij}}$ on the image plane. $\vec{x}_{c_{ij}}$ can be expressed as a function g based on $\vec{X}_{ij}$ and $\vec{p}_c$:

$$\vec{x}_{c_{ij}} = g(\vec{X}_{ij}; \vec{p}_c) \tag{B.5}$$

The formulation of $\vec{x}_{c_{ij}}$ is based on the lens distortion and projection in camera. We use a five parameter model for lens distortion [33]:

$$\begin{aligned} \begin{pmatrix} x'_{ij} \\ y'_{ij} \end{pmatrix} &= \begin{pmatrix} x_{ij}/z_{ij} \\ y_{ij}/z_{ij} \end{pmatrix} \\ r_c^2 &= {x'_{ij}}^2 + {y'_{ij}}^2 \\ x''_{ij} &= x'_{ij}(1 + k_1 r_c^2 + k_2 r_c^4 + k_3 r_c^6) + 2p_1 x'_{ij} y'_{ij} + p_2(r_c^2 + 2{x'_{ij}}^2) \\ y''_{ij} &= y'_{ij}(1 + k_1 r_c^2 + k_2 r_c^4 + k_3 r_c^6) + 2p_2 x'_{ij} y'_{ij} + p_1(r_c^2 + 2{y'_{ij}}^2) \end{aligned} \tag{B.6}$$

where (x'_{ij}, y'_{ij}) is the projected 2D point from $\vec{X}_{ij}$ before lens distortion, and (x''_{ij}, y''_{ij}) is the projected point after lens distortion. k_1, k_2, k_3 are the radial distortion factors and p_1, p_2 are the tangential distortion factors. Finally, $\vec{x}_{c_{ij}}$ can be computed using (x''_{ij}, y''_{ij}), where KK_c is the intrinsic matrix of the camera:

$$\vec{x}_{c_{ij}} = KK_c \begin{pmatrix} x''_{ij} \\ y''_{ij} \\ 1 \end{pmatrix} \tag{B.7}$$

The error function is formulated as the re-projection error in the camera, which is the mean squared error across all pixels and projectors between the observed pixel $\vec{x}_{c_{ij}}$ and the estimated pixel $\hat{x}_{c_{ij}}$ computed using above equations.

$$E = \sum_i \sum_j (\vec{x}_{c_{ij}} - \hat{x}_{c_{ij}})^2, \tag{B.8}$$

The above error function has been implemented with a numeric analysis library Alglib [105] using a Levenberg-Marquardt algorithm. To facilitate the solver, we also compute the analytic Jacobian matrix. For illustrative purpose, assuming we have 4 projectors with M pixels in each projector. The Jacobian matrix can be computed as:

$$J = \frac{\partial F}{\partial \vec{p}} = \left[\begin{array}{cccc|cc}
A_{11} & 0 & 0 & 0 & B_{11} & C_{11} \\
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\
A_{1M} & 0 & 0 & 0 & B_{1M} & C_{1M} \\
\hline
0 & A_{21} & 0 & 0 & B_{21} & C_{21} \\
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\
0 & A_{2M} & 0 & 0 & B_{2M} & C_{2M} \\
\hline
0 & 0 & A_{31} & 0 & B_{31} & C_{31} \\
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\
0 & 0 & A_{3M} & 0 & B_{3M} & C_{3M} \\
\hline
0 & 0 & 0 & A_{41} & B_{41} & C_{41} \\
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\
0 & 0 & 0 & A_{4M} & B_{4M} & C_{4M}
\end{array}\right]$$

(projector 1 pixels; projector 2 pixels; projector 3 pixels; projector 4 pixels)

($\underbrace{\text{projectors } 1-4}$; $\underbrace{\text{camera}}$; $\underbrace{\text{sphere}}$)

$$\begin{aligned}
A_{ij} &= \frac{\partial \vec{x}_{c_{ij}}}{\partial \vec{X}_{ij}} \frac{\partial \vec{X}_{ij}}{\partial \vec{p}_{p_i}} \\
B_{ij} &= \frac{\partial \vec{x}_{c_{ij}}}{\partial \vec{p}_c} \\
C_{ij} &= \frac{\partial \vec{x}_{c_{ij}}}{\partial \vec{X}_{ij}} \frac{\partial \vec{X}_{ij}}{\partial \vec{p}_s}
\end{aligned} \tag{B.9}$$

where A_{ij} is the partial derivative of the camera 2D point $\vec{x}_{c_{ij}}$ to projector i parameters $\vec{p}_{p_i}$. B_{ij} is the partial derivative of $\vec{x}_{c_{ij}}$ to camera parameters $\vec{p}_c$. C_{ij} is the partial derivative of $\vec{x}_{c_{ij}}$ to sphere parameters $\vec{p}_s$.

The analytic expressions of equation B.9 can be derived using equations B.3, B.4, B.6 and B.7. In practice the analytic expression of the Jacobian matrix is auto-generated using the Symbolic Math Toolbox in MATLAB [59]. The generated analytic expressions are in C format and imported in the optimization code with Alglib. The source code of the entire calibration are open source and can be found on [107].

Appendix C

User Study Questionnaires

C.1 User Study on Size and Depth Perception

Subject ID: Condition Order: Date:

Electrical and Computer Engineering

3D Perspective-Corrected Spherical Display
Experiment Questionnaire Form

Using a checkmark (✓) please rate the statements below about the different display conditions:

SD (-2)	D (-1)	N (0)	A (+1)	SA (+2)
Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree

Flat Screen		SD	D	N	A	SA
Depth Task	It was easy to judge the depth					
	It was intuitive to judge the depth					
	I felt I performed the depth task well					
Size Task	It was easy to judge the size and compare it to the real ball					
	It was intuitive to adjust the size of the virtual ball					
	I felt I performed the size task well					

Spherical Screen		SD	D	N	A	SA
Depth Task	It was easy to judge the depth					
	It was intuitive to judge the depth					
	I felt I performed the depth task well					
Size Task	It was easy to judge the size and compare it to the real ball					
	It was intuitive to adjust the size of the virtual ball					
	I felt I performed the size task well					

(after all tasks)

Flat Screen	SD	D	N	A	SA
I was able to perceive the virtual pool ball like a real pool ball on this screen					
I enjoyed doing tasks on this screen					

Spherical Screen	SD	D	N	A	SA
I was able to perceive the virtual pool ball like a real pool ball on this screen					
I enjoyed doing tasks on this screen					

Subject ID: Condition Order: Date:

1. Please specify your gender:
   - ☐ Male ☐ Female
2. Please specify your age range:
   - ☐ 18-25 ☐ 26-35 ☐ 36-45
   - ☐ 46-55 ☐ 56-65 ☐ Above 65
3. Are you a user of 3D graphic interface (e.g. playing 3D Games or using CAD, Blender, Unity3D)?
   - ☐ Yes
     i. Please mention names of software: .....
     ii. What is the frequency of usage?
     1. Daily 2. Weekly 3. Occasionally
   - ☐ No
4. Have you experienced with Virtual Reality before?
   - ☐ Yes
     i. Please mention type of device: .....
     ii. How many times of usage?
     1. 1-5 times 2. 5-10 times 3. Above 10 times
   - ☐ No

5. For the depth task, you feel you performed better on:
☐ Flat Screen ☐ Spherical Screen ☐ No difference on performance
Please specify the reason:

6. For the depth task, you prefer to:
☐ Flat Screen ☐ Spherical Screen ☐ No preference
Please specify the reason:

7. For the size task, you feel you performed better on:
☐ Flat Screen ☐ Spherical Screen ☐ No difference on performance
Please specify the reason:

8. For the size task, you prefer to:
☐ Flat Screen ☐ Spherical Screen ☐ No preference
Please specify the reason:

C.2 User Study 1 on Perceptual Duality

Subject ID: Condition Order: Date:

Electrical and Computer Engineering

3D Perspective-Corrected Spherical Display

Questionnaire Form

Using a checkmark (✓) please rate the statements below about the different display conditions:

SD (-2)	D (-1)	N (0)	A (+1)	SA (+2)
Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree

	Head-Move	SD	D	N	A	SA
NonStereo	I was able to perceive the virtual pool ball like a real ball					
	I was confident of my answers when completing the task.					

	Head-Move	SD	D	N	A	SA
Stereo	I was able to perceive the virtual pool ball like a real ball					
	I was confident of my answers when completing the task.					

	Object-Move	SD	D	N	A	SA
NonStereo	I was able to perceive the virtual pool ball like a real ball					
	I was confident of my answers when completing the task.					

	Object-Move	SD	D	N	A	SA
Stereo	I was able to perceive the virtual pool ball like a real ball					
	I was confident of my answers when completing the task.					

Subject ID: Condition Order: Date:

Electrical and Computer Engineering

3D Perspective-Corrected Spherical Display

Questionnaire Form

1. Please specify your gender:
   - ☐ Male ☐ Female
2. Please specify your age range:
   - ☐ 18-25 ☐ 26-35 ☐ 36-45
   - ☐ 46-55 ☐ 56-65 ☐ Above 65
3. Have you experienced with Virtual Reality before?
   - ☐ Yes
     - i. Please mention type of device: .....
     - ii. How many times of usage?
       1. 1-5 times 2. 5-10 times 3. Above 10 times
   - ☐ No

Please feel free to leave any comments:

C.3 User Study 2 on Perceptual Duality

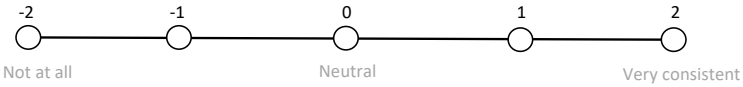
Subject ID: Condition Name: Date:

Electrical and Computer Engineering

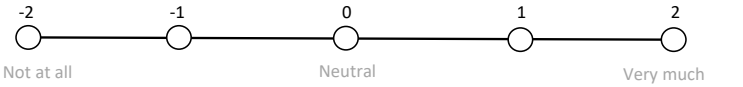
CRYSTAL: Spherical Fish Tank Virtual Reality Display

Experiment Questionnaire Form

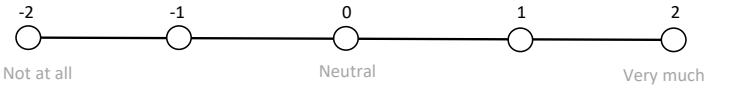
1. How much did your overall experience with virtual objects seem *consistent* with your real-world experience?



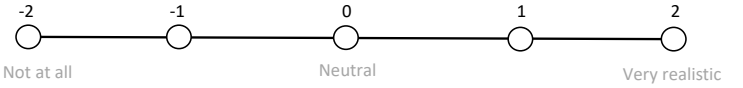
2. To what extent did you feel you could *reach and grasp* an object in the virtual environment?



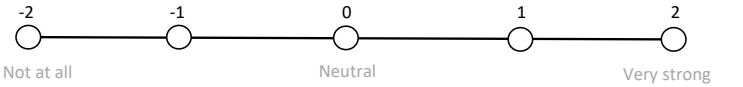
3. To what extent did virtual objects appear *geometrically correct* during your movement, did they seem to have the right *size and distance* in relation to your position?



4. Overall, how *realistic* did the virtual object appear?



5. Overall, how strong did you feel the object *exist* in the real world?



a place of mind
THE UNIVERSITY OF BRITISH COLUMBIA

Subject ID: Condition Order: Date:

Electrical and Computer Engineering

CRYSTAL: Spherical Fish Tank Virtual Reality Display

Demographic Questionnaire Form

1. Please specify your gender:
   - ☐ Male ☐ Female
2. Please specify your age range:
   - ☐ 18-25 ☐ 26-35 ☐ 36-45
   - ☐ 46-55 ☐ 56-65 ☐ Above 65
3. Have you experienced Virtual Reality before?
   - ☐ Yes
     - i. Please mention type of device: ……
     - ii. How many times of usage?
       1. 1-5 times 2. 5-10 times 3. Above 10 times
   - ☐ No

4. Which viewing condition that you would prefer:

☐ Perspective1 ☐ Perspective2

Please specify the reason:

a place of mind
THE UNIVERSITY OF BRITISH COLUMBIA

Appendix D

User Study Result

D.1 User Study on Size and Depth Perception

Within Subjects Effects

	Sum of Squares	df	Mean Square	F	p	η^2_p
Display	0.00202	1	0.00202	11.000	0.005	0.423
Residual	0.00275	15	1.83e-4			
Distance	1.87e-5	2	9.37e-6	0.287	0.753	0.019
Residual	9.81e-4	30	3.27e-5			
Display ✻ Distance	2.71e-5	2	1.35e-5	0.448	0.643	0.029
Residual	9.06e-4	30	3.02e-5			

Note. Type 3 Sums of Squares

Figure D.1: Repeated measures two-way ANOVA on the absolute depth error with factors of Display and Distance in the depth ranking task in Chapter 5.

Within Subjects Effects

	Sum of Squares	df	Mean Square	F	p	partial η^2
Display	0.00957	1	0.00957	5.132	0.039	0.255
Residual	0.02797	15	0.00186			
Distance	0.00333	2	0.00167	2.093	0.141	0.122
Residual	0.02389	30	7.96e-4			
InitialSize	0.00119	1	0.00119	0.546	0.472	0.035
Residual	0.03265	15	0.00218			
Display ✻ Distance	4.57e-4	2	2.28e-4	0.348	0.709	0.023
Residual	0.01969	30	6.56e-4			
Display ✻ InitialSize	1.32e-4	1	1.32e-4	0.133	0.720	0.009
Residual	0.01492	15	9.95e-4			
Distance ✻ InitialSize	0.00250	2	0.00125	1.430	0.255	0.087
Residual	0.02619	30	8.73e-4			
Display ✻ Distance ✻ InitialSize	0.00119	2	5.97e-4	1.851	0.175	0.110
Residual	0.00968	30	3.23e-4			

Note. Type 3 Sums of Squares

Figure D.2: Repeated measures three-way ANOVA on the absolute size error with factors of Display, Distance and InitialSize in the size matching task in Chapter 5.

Within Subjects Effects

	Sum of Squares	df	Mean Square	F	p	partial η^2
Display	0.02342	1	0.02342	4.88	0.043	0.246
Residual	0.07196	15	0.00480			
Distance	0.01462	2	0.00731	5.13	0.012	0.255
Residual	0.04276	30	0.00143			
InitialSize	0.07849	1	0.07849	36.64	< .001	0.710
Residual	0.03213	15	0.00214			
Display ✻ Distance	0.00641	2	0.00320	5.13	0.012	0.255
Residual	0.01875	30	6.25e-4			
Display ✻ InitialSize	0.01634	1	0.01634	28.71	< .001	0.657
Residual	0.00854	15	5.69e-4			
Distance ✻ InitialSize	0.01604	2	0.00802	12.97	< .001	0.464
Residual	0.01855	30	6.18e-4			
Display ✻ Distance ✻ InitialSize	0.00133	2	6.65e-4	1.47	0.245	0.089
Residual	0.01355	30	4.52e-4			

Note. Type 3 Sums of Squares

Post Hoc Comparisons - Display ✻ Distance

Comparison					Mean Difference	SE	df	t	$p_{bonferroni}$
Display	Distance	-	Display	Distance					
Flat	Far	-	Flat	Mid	0.02347	0.008	52.1	2.9318	0.045
Flat	Far	-	Flat	Near	0.03388	0.008	52.1	4.2325	0.001
Flat	Far	-	Fishbowl	Far	-0.00576	0.01122	23.1	-0.5133	1
Flat	Mid	-	Flat	Near	0.01041	0.008	52.1	1.3007	1
Flat	Mid	-	Fishbowl	Mid	-0.02969	0.01122	23.1	-2.6455	0.126
Flat	Near	-	Fishbowl	Near	-0.03081	0.01122	23.1	-2.7453	0.108
Fishbowl	Far	-	Fishbowl	Mid	-4.66e−4	0.008	52.1	-0.0582	1
Fishbowl	Far	-	Fishbowl	Near	0.00883	0.008	52.1	1.1027	1
Fishbowl	Mid	-	Fishbowl	Near	0.00929	0.008	52.1	1.1609	1

Post Hoc Comparisons - Display ✻ InitialSize

Comparison					Mean Difference	SE	df	t	$p_{bonferroni}$
Display	InitialSize	-	Display	InitialSize					
Flat	Small	-	Flat	Large	-0.05889	0.00752	22.4	-7.835	< .001
Flat	Small	-	Fishbowl	Small	-0.04054	0.01057	18.5	-3.834	0.004
Flat	Large	-	Fishbowl	Large	-0.00364	0.01057	18.5	-0.344	1
Fishbowl	Small	-	Fishbowl	Large	-0.02199	0.00752	22.4	-2.926	0.032

Figure D.3: Repeated measures three-way ANOVA on the size ratio with factors of Display, Distance and InitialSize, followed by post-hoc analysis t-test with Bonferroni correction on the interaction between Display x Distance and Display x InitialSize in the size matching task in Chapter 5.

Paired Samples T-Test

			statistic	df	p	Mean difference	SE difference	Cohen's d
Flat	Fishbowl	Student's t	-1.02	15.0	0.322	-0.0110	0.0107	-0.256

Figure D.4: Results of pairwise t-test on the normalized head movement in the depth ranking task in Chapter 5.

Paired Samples T-Test

			statistic	df	p	Mean difference	SE difference	Cohen's d
Flat	Fishbowl	Student's t	-2.76	15.0	0.015	-0.0295	0.0107	-0.689

Figure D.5: Results of pairwise t-test on the normalized head movement in the size matching task in Chapter 5.

Paired Samples T-Test

			statistic	p	Cohen's d
Fishbowl-Depth-easy	Flat-Depth-easy	Wilcoxon W	78.0 [a]	0.002	1.134
Fishbowl-Depth-intuitive	Flat-Depth-intuitive	Wilcoxon W	62.0 [b]	0.009	0.834
Fishbowl-Depth-confident	Flat-Depth-confident	Wilcoxon W	45.0 [d]	0.007	0.854
Fishbowl-Size-easy	Flat-Size-easy	Wilcoxon W	35.0 [d]	0.146	0.413
Fishbowl-Size-intuitive	Flat-Size-intuitive	Wilcoxon W	14.5 [e]	1.000	0.000
Fishbowl-Size-confident	Flat-Size-confident	Wilcoxon W	15.0 [f]	0.374	0.250
Fishbowl-Overall-real	Flat-Overall-real	Wilcoxon W	72.5 [a]	0.006	0.968
Fishbowl-Overall-enjoy	Flat-Overall-enjoy	Wilcoxon W	15.0 [g]	0.048	0.606

[a] 4 pair(s) of values were tied

[b] 5 pair(s) of values were tied

[d] 7 pair(s) of values were tied

[e] 9 pair(s) of values were tied

[f] 10 pair(s) of values were tied

[g] 11 pair(s) of values were tied

Figure D.6: Results of Wilcoxon Signed Rank Test on the Likert-scale post-experiment questions in the depth and size task in Chapter 5. A complete copy of the questionnaire can be found in Appendix C.1.

D.2 User Study 1 on Perceptual Duality

Paired Samples T-Test

			statistic	p	Cohen's d
NonStereo-Object	NonStereo-Head	Wilcoxon W	151.0	< .001	1.850
Stereo-Object	Stereo-Head	Wilcoxon W	119.0 [a]	0.009	0.875
NonStereo-Object	Stereo-Object	Wilcoxon W	94.0 [b]	0.010	0.761
NonStereo-Head	Stereo-Head	Wilcoxon W	20.5 [d]	0.027	-0.664

[a] 1 pair(s) of values were tied

[b] 3 pair(s) of values were tied

[d] 2 pair(s) of values were tied

Figure D.7: Results of pairwise Wilcoxon Signed Rank Test on the bias error in the size judgement task with p-values before Bonferroni correction in Chapter 6.

Paired Samples T-Test

			statistic	p	Mean difference	SE difference	Cohen's d
Stereo-Head	NonStereo-Head	Wilcoxon W	50.5 [a]	0.015	1.000	0.226	0.695
Stereo-Object	NonStereo-Object	Wilcoxon W	45.0 [a]	0.059	1.000	0.193	0.518
Stereo-Head	Stereo-Object	Wilcoxon W	51.0 [b]	0.097	1.000	0.259	0.441
NonStereo-Head	NonStereo-Object	Wilcoxon W	38.5 [a]	0.227	1.000	0.182	0.313

[a] 7 pair(s) of values were tied

[b] 6 pair(s) of values were tied

Figure D.8: Results of pairwise Wilcoxon Signed Rank Test on the confidence rating with p-values before Bonferroni correction in Chapter 6. Results are not significant after Bonferroni correction. A complete copy of the questionnaire can be found in Appendix C.2.

Paired Samples T-Test

			statistic	p	Mean difference	SE difference	Cohen's d
Stereo-Head	NonStereo-Head	Wilcoxon W	25.5 [a]	0.305	1.000	0.254	0.2810
Stereo-Object	NonStereo-Object	Wilcoxon W	42.0 [b]	0.022	1.500	0.358	0.6369
Stereo-Head	Stereo-Object	Wilcoxon W	19.5 [b]	0.761	-4.66e−5	0.277	-0.0514
NonStereo-Head	NonStereo-Object	Wilcoxon W	52.0 [d]	0.090	1.000	0.344	0.4152

[a] 9 pair(s) of values were tied

[b] 8 pair(s) of values were tied

[d] 6 pair(s) of values were tied

Figure D.9: Results of pairwise Wilcoxon Signed Rank Test on the realism rating with p-values before Bonferroni correction in Chapter 6. Results are not significant after Bonferroni correction. A complete copy of the questionnaire can be found in Appendix C.2.

D.3 User Study 2 on Perceptual Duality

Within Subjects Effects

	Sum of Squares	df	Mean Square	F	p	η^2_p
Projection	0.0978	1	0.0978	4.14	0.067	0.273
Residual	0.2598	11	0.0236			
Movement	2.2245	1	2.2245	36.64	< .001	0.769
Residual	0.6678	11	0.0607			
Projection ✻ Movement	0.4537	1	0.4537	10.10	0.009	0.479
Residual	0.4942	11	0.0449			

Note. Type 3 Sums of Squares

Post Hoc Comparisons - Projection ✻ Movement

Comparison					Mean Difference	SE	df	t	$p_{bonferroni}$
Projection	Movement	-	Projection	Movement					
Perspective	Object	-	Perspective	Head	0.625	0.0938	21.5	6.66	< .001
Perspective	Object	-	WeakPerspective	Object	0.285	0.0756	20.1	3.77	0.005
Perspective	Head	-	WeakPerspective	Head	-0.104	0.0756	20.1	-1.38	0.732
WeakPerspective	Object	-	WeakPerspective	Head	0.236	0.0938	21.5	2.52	0.08

Figure D.10: Results of repeated measures two-way ANOVA on the bias error with factors of Projection and Movement, followed by post-hoc pairwise t-test with Bonferroni correction on the interaction between Projection x Movement.

Paired Samples T-Test

			statistic	p	Cohen's d
WeakPersp-consistent	Persp-consistent	Wilcoxon W	28.0 [a]	0.019	1.018
WeakPersp-reachable	Persp-reachable	Wilcoxon W	55.5 [b]	0.047	0.720
WeakPersp-correctGeometry	Persp-correctGeometry	Wilcoxon W	50.0 [d]	0.024	0.894
WeakPersp-realistic	Persp-realistic	Wilcoxon W	32.5 [e]	0.040	0.751
WeakPersp-presence	Persp-presence	Wilcoxon W	37.0 [f]	0.090	0.568

[a] 5 pair(s) of values were tied
[b] 1 pair(s) of values were tied
[d] 2 pair(s) of values were tied
[e] 4 pair(s) of values were tied
[f] 3 pair(s) of values were tied

Figure D.11: Results of Wilcoxon Signed Rank Test on the Likert-scale post-experiment questions in the subjective impression task in Chapter 6. A complete copy of the questionnaire can be found in Appendix C.3.