

Touching the 3rd Dimension

Interaction with Stereoscopic Data On and Above Interactive Surfaces

Dissertation zur Erlangung des Grades
Doktor der Ingenieurwissenschaften (Dr.-Ing.)
der Naturwissenschaftlich-Technischen Fakultät I
der Universität des Saarlandes

vorgelegt von
Florian Daiber, Dipl. Geoinf.
Saarbrücken, den 11. März 2015



UNIVERSITÄT
DES
SAARLANDES

Dean:

Prof. Dr. Markus Bläser

Head of Committee:

Prof. Dr. Andreas Zeller

Reviewers:

Prof. Dr. Antonio Krüger

Prof. Dr. Frank Steinicke

Committee member:

Dr. Sven Gehring

Day of Defense:

May 4th, 2015

Abstract

In recent years, interaction with three-dimensional (3D) data has become more and more popular. However, current 3D user interfaces (3DUIs), as for example provided by virtual reality (VR) systems, are very often expert systems with complex user interfaces and high instrumentation. While stereoscopic displays allow users to perceive 3D content in an intuitive and natural way, interaction with stereoscopic data is still a challenging task, especially with objects that are displayed with different parallaxes. To overcome this interaction problem, multi-touch or depth sensing as commodity tracking technologies can be used. This thesis therefore investigates the challenges that occur when the flat world of surface computing meets the spatially complex 3D space.

The thesis contributes a number of interactive research prototypes, interaction techniques and insights from the corresponding studies. 3D interaction techniques will be designed and studied in canonical 3DUI tasks, which leads to the conclusion that significant differences arise when interacting with objects that are stereoscopically displayed at different parallaxes. The results give implications for the design of usable 3D interaction techniques that might enable VR in the living room or at public places. To sum up, this work contributes to a better understanding of stereoscopic 3D applications for end users and can be seen as a step towards a ubiquitous distribution of interactive 3D technologies and applications.

Zusammenfassung

Die Interaktion mit dreidimensionalen (3D) Daten hat in den letzten Jahren an Bedeutung gewonnen. Allerdings handelt es sich bei heutigen 3D Benutzerschnittstellen meist um Expertensysteme mit komplexen Benutzeroberflächen und -instrumentierung. Während stereoskopische Displays die intuitive Wahrnehmung von 3D erlauben, ist die Interaktion mit diesen Displays immer noch wenig benutzerfreundlich, insbesondere wenn Objekte mit verschiedenen Parallaxen angezeigt werden. In der vorliegenden Arbeit wird deshalb der Bedarf erforscht, der sich aus den Herausforderungen ergibt, die auftreten, wenn die flache Welt der interaktiven Oberflächen auf den komplexen, virtuellen 3D Raum trifft.

Der Beitrag dieser Arbeit liegt in der Entwicklung von interaktiven Forschungsprototypen, neuen Interaktionstechniken sowie Erkenntnissen aus Studien. Es werden 3D Interaktionstechniken entwickelt und in universellen 3D Interaktionsaufgaben evaluiert. Die Ergebnisse zeigen, dass signifikante Unterschiede bei der Interaktion mit Objekten bestehen, die mit unterschiedlichen Parallaxen angezeigt werden. Des Weiteren ergeben sich Implikationen für das Design von Interaktionstechniken, die einen intuitiven Zugang zu Virtueller Realität (VR) im Wohnzimmer oder an öffentlichen Orten ermöglichen. Die Arbeit bildet damit einen wichtigen Beitrag zum Verständnis von stereoskopischem 3D für den Endanwender und kann als Schritt zur Verbreitung von interaktiven 3D Technologien und Anwendungen betrachtet werden.

Acknowledgments

Over the last years I met a lot of great people that supported me and my work in various ways. In the following, I want to thank some people in particular and I apologize to the ones I forgot. First of all, I want to thank Antonio Kruüger for the opportunity to work on this thesis and for all the helpful support and guidance. Thank you for always pushing me to the limit! I thank Frank Steinicke for providing me with helpful advice, and for reviewing my thesis.

I had the pleasure to work in inspiring and creative environments. For this, I want to thank my colleagues at DFKI (in random order) Denise Paradowski, Frederic Kerber, Sven Gehring, Christian Lander, Marco Speicher, Matthias Böhmer, Johannes Schöning, Ralf Jung, Markus Löchtefeld, Pascal Lessel, and Gerrit Kahl. I also thank Patrick Maué, Oliver Paczkowski, Jörg Müller, Alexander Walkowski, Christoph Stasch, Anusuriya Devaraju, Mareike Kritzler, and the whole Institute for Geoinformatics for being supportive colleagues during my time at University of Münster. I had a great time working with all of you!

I am thankful to all people whom I had the pleasure to collaborate with over the last few years. I am especially thankful for the input, discussion and joint work with Bruno di Araujo, Tsvika Kuflik, Antti Oulasvirta, Michael Rohs, Gabor Sörös, Wolfgang Stürzlinger, Dimitar Valkov, Tomer Weller, and Ulrich von Zadow.

I am also grateful to the students who shared their distinct and unique abilities to add important bits and pieces to the result of this dissertation. In particular, I thank Eric Falk, Felix Kosmalla, Lianchao Li, Marco Speicher, Michael Mauderer, and Oliver Schönleben. Also many thanks to Margaret De Lap for thoroughly proofreading my thesis.

Finally, I want to thank my family for their support over the last years. Last but not least, I want to thank Ramona and Nicolas without whom this thesis would have not been possible. Thank you for your patience and support!

Contents

I. Introduction and Background	15
1. Introduction	17
1.1. Context and Motivation	17
1.2. Background and Problem	18
1.3. Thesis Statement	19
1.4. Methods	20
1.5. Results and Contributions	21
1.6. Thesis Overview and Conventions Used	26
2. Perceptual and Technical Foundations of Interactive Stereoscopic Surfaces	27
2.1. Stereoscopic Vision and Depth Perception	27
2.1.1. Depth Cues	28
2.1.2. Depth Perception on Stereoscopic Multi-Touch Displays	31
2.1.3. Depth Perception on Handheld Stereoscopic Displays for AR	33
2.2. 3D Technology	36
2.2.1. Output Technology	36
2.2.2. Input Technology	42
2.3. 3D User Interfaces and 3D Interaction	49
2.3.1. Canonical 3D Tasks	49
2.3.2. 3D Interaction	56
2.4. Conclusion	63
3. Classification of Multi-Modal 3D Interaction on Interactive Surfaces	65
3.1. Multi-Touch Interaction with Spatial 3D Data	65
3.1.1. Multi-Touch Interactions	66
3.1.2. Study	67
3.1.3. Multi-Touch Framework	68
3.2. Extended Framework for Whole Body Interaction	69
3.2.1. Multi-Touch and Foot Interaction	70
3.2.2. Gaze-based Interaction	70
3.2.3. Extended Framework for Foot and Gaze-based Interaction	71
3.3. Taxonomy for Multi-Modal 3D Interaction On and Above Interactive Surfaces	71
3.3.1. Example 1: 3D Interaction on Mobile Devices	73
3.3.2. Example 2: Gestural and Mobile 3D Interaction Above the Interac- tive Surface	74

3.4. Conclusion	75
II. Canonical 3D Tasks	77
4. Indirect 3D Selection	79
4.1. Multi-touch 3D Selection Techniques	79
4.1.1. Ballon/Fishnet Selection	80
4.1.2. Corkscrew Selection	83
4.2. Experiment	83
4.2.1. Participants	84
4.2.2. Conditions	84
4.2.3. Task	84
4.2.4. Design	84
4.2.5. Procedure	85
4.2.6. Apparatus	85
4.2.7. Independent and Dependent Variables	85
4.2.8. Hypotheses	86
4.3. Results	86
4.3.1. Balloon/Fishnet Selection	86
4.3.2. Corkscrew Selection	87
4.3.3. Balloon/Fishnet vs. Corkscrew Selection	87
4.3.4. Post-Study Questionnaire	88
4.4. Discussion	88
4.5. Conclusion	90
5. 3D Manipulation	91
5.1. Bimanual Gestural and Mobile 3D Manipulation	91
5.2. Experiment	92
5.2.1. Participants	93
5.2.2. Conditions	93
5.2.3. Task	93
5.2.4. Design	94
5.2.5. Procedure	94
5.2.6. Apparatus	95
5.2.7. Independent and Dependent Variables	95
5.2.8. Hypotheses	95
5.3. Results	96
5.3.1. Task Completion Time	96
5.3.2. Translation Task Precision	96
5.3.3. Rotation Task Precision	98
5.3.4. NASA TLX	98
5.3.5. Observations During the Study	99

5.4.	Discussion	100
5.4.1.	Interaction Technique	100
5.4.2.	Monoscopic vs. Stereoscopic Display	101
5.4.3.	Docking Tasks	101
5.5.	Conclusion	101
6.	3D Travel	103
6.1.	Mobile and Gestural 3D Travel	104
6.1.1.	Bimanual Grabbing	104
6.1.2.	Whole-Body Tilt and Grab	105
6.1.3.	Mobile Multi-Touch	107
6.1.4.	Mobile Tilt and Touch	107
6.2.	Experiment	109
6.2.1.	Participants	109
6.2.2.	Conditions	109
6.2.3.	Task	110
6.2.4.	Design	110
6.2.5.	Procedure	111
6.2.6.	Apparatus	111
6.2.7.	Independent and Dependent Variables	112
6.2.8.	Hypotheses	112
6.2.9.	Improvement to Existing Methodology	112
6.3.	Results	112
6.3.1.	Task completion time	113
6.3.2.	Error rate	114
6.3.3.	NASA TLX	114
6.4.	Discussion	116
6.4.1.	Experimental results	116
6.4.2.	Advantages	116
6.4.3.	Limitations	117
6.5.	Conclusion	117
III.	Beyond the Multi-touch Surface	119
7.	Interaction Context for Multi-touch 3D Interaction	121
7.1.	Interaction Context	121
7.2.	Modalities	122
7.2.1.	Whole-Body Postures	122
7.2.2.	Hand Postures and Grasp	123
7.2.3.	Eye Gaze	123
7.3.	Conclusion	123

8. Reach to Grasp Interaction	125
8.1. Grasp Pre-Study	125
8.1.1. Experiment	126
8.1.2. Results	128
8.2. Grasp Corpus Study	130
8.2.1. Experiment	130
8.2.2. Analysis	133
8.2.3. Results	135
8.3. Discussion	136
8.4. Design Considerations	138
8.4.1. Optional Menus	138
8.4.2. Level of Detail Interaction	138
8.4.3. Adapting the 3D User Interface	139
8.4.4. Improving Object Recognition	139
8.5. Conclusion	140
 IV. Handheld 3D Interaction	 143
9. Interactive Handheld Stereoscopic Devices	145
9.1. Interaction with Handheld Stereoscopic Devices	145
9.1.1. Selection	146
9.1.2. Manipulation	146
9.1.3. Travel	147
9.2. Flight Control 3D	149
9.3. User Study	150
9.3.1. Participants	150
9.3.2. Task	150
9.3.3. Procedure	151
9.3.4. Apparatus	151
9.3.5. Independent and Dependent Variables	151
9.4. Results	151
9.4.1. Game Usability	151
9.4.2. Interaction techniques	153
9.5. Discussion	155
9.5.1. 3D Game Experiences	155
9.6. Conclusion	157
 10. Handheld Stereoscopic Augmented Reality	 159
10.1. Handheld Stereoscopic Augmented Reality	159
10.1.1. Interaction	161
10.1.2. Proof-of-Concept Application	161

10.2. Experiment	162
10.2.1. Participants	163
10.2.2. Task	163
10.2.3. Procedure	164
10.2.4. Apparatus	165
10.2.5. Independent and Dependent Variables	166
10.3. Results	167
10.3.1. Object Size	167
10.3.2. Autostereoscopy and Parallax	167
10.4. Discussion	168
10.5. Conclusion	170
11. Conclusion	173
11.1. Thesis Summary	173
11.2. Contributions	173
11.3. Future Work	174
11.4. Concluding Remarks	176

Part I.

Introduction and Background

1. Introduction

1.1. Context and Motivation

In recent years, interaction with three-dimensional (3D) data has become more and more popular. Current 3DUIs, as for example provided by VR systems, consist of stereoscopic projection and tracked input devices. But these are often expert systems with complex user interfaces and high instrumentation. Nonetheless, stereoscopic displays allow users to perceive 3D data in an intuitive and natural way. On stereoscopic displays objects might be displayed with different parallax paradigms resulting in different stereoscopic effects. Objects may appear behind (positive parallax), on top of (zero parallax), or in front (negative parallax) the screen. Interaction with objects that are displayed with different parallaxes is still a challenging task, even in virtual environments (VEs) [182].

Novel input technologies such as multi-touch or depth sensing devices have received considerable attention in recent years, especially for 2D user interfaces (UIs). These technologies make it possible to control applications with several simultaneously performed touch or mid-air gestures. These devices have paved the way for the next generation of UIs that go beyond those that rely on the windows, icons, menus and pointer (WIMP) metaphor [150] by allowing a more natural interaction [207]. But they also have a great potential for exploring complex content in an easy and natural manner, e.g. in the case of 3D data. However, while multi-touch has shown its usefulness for 2D interaction by providing more natural and intuitive techniques such as 2D translation, scaling and rotation, it has rarely been considered whether and how these concepts can be extended to 3DUIs. To overcome the difficulties that occur during interaction with stereoscopic data, multi-touch or depth sensing as affordable tracking technologies can be used. Such technologies allow a rich set of interactions without high instrumentation. In combination with autostereoscopic displays, any instrumentation of the user can be avoided entirely while providing an advanced user experience. However, the benefits and limitations of using such devices in combination with stereoscopic displays have not yet been examined in depth and are not well understood [182].

Thus, there is a need to investigate the challenges of how users interact with stereoscopically displayed 3D data, in particular when the interaction is restricted to a two-dimensional multi-touch surface. Thus, ergonomic and perceptual issues during interaction on and above stereoscopic displays need to be investigated in depth. This thesis therefore investigates interaction with stereoscopic data on and above interactive surfaces by studying universal 3D tasks and extensions of the interaction space that specifically address this class of devices. The next section introduces the background with a small literature review

1. Introduction

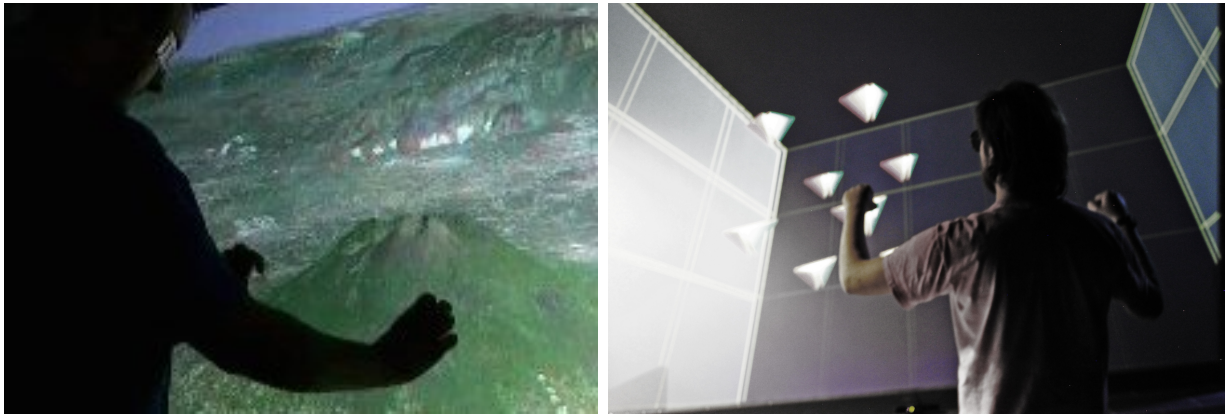


Figure 1.1.: Interaction with stereoscopically displayed geo-spatial data on a multi-touch surface with anaglyph display (left). Mobile and gestural 3D interaction with a large stereoscopic display (right).

(a comprehensive discussion of related work is provided in Chapter 2) and the challenges that occur when the flat world of surface computing meets the spatially complex 3D space.

1.2. Background and Problem

This research is grounded in a variety of fields within computer science, most prominently human-computer interaction (HCI) and 3DUI. In recent years there has been an increasing interest in 3D related technology, e.g. 3D movies, augmented reality applications and gaming. Although 3D interaction has a long research tradition, recent interactive 3D technology lacks natural interaction because it often requires high user instrumentation and special input devices. Novel input devices such as interactive tabletops, smartphones and depth sensors have the potential to close this gap. Most of these devices are affordable and already in use in our everyday life, as smartphones, depth sensors and 3D television. Moreover, current 3DUIs, as for example provided by VR systems, consist of stereoscopic projection and tracked input devices. But these are often expert systems with complex user interfaces and high instrumentation. For instance, travel in virtual environments is a universal interaction task and has been an intensive research topic. However, it is still a challenging task even in VR-based environments. In order to address these challenges in VR research, but also to introduce interactive 3D applications to the living room, affordable input devices can be used (cf. [169, 124]).

Major UI paradigms for graphical user interfaces, e.g. the WIMP [150] paradigm are becoming obsolete and are increasingly being replaced by interfaces that follow the old vision of ubiquitous computing [204] or the recent paradigm of the Natural user interface (NUI) [207]. This trend also influences 3D user interfaces [21] and some research has been done in the field of 3DUI design and touch-based interaction with stereoscopic 3D displays. While traditional 3DUIs are often restricted to heavily instrumented virtual environments, Weiser's [204] vision of ubiquitous computing goes beyond this by integrating

the technology in the environment. The user then interacts with everyday objects, often without even being totally aware that an explicit interaction between a human and a computer has taken place. Virtual and augmented reality play an important role in Weiser's vision and therefore the ubiquitous computing perspective also needs to be considered in the scope of this thesis research.

Multi-touch surfaces and interaction has been an active research field in the last decades. In the last years, re-inspired by Han's multi-touch display that relies on the principle of frustrated total internal reflection [84], much work has been carried out in general but also on the definition of frameworks and taxonomies for gesture-based multi-touch input. Wu et al. [213] defined the principle of gesture registration, relaxation and reuse. Wobbrock et al. [211] investigated user-defined gestures and developed a taxonomy of gestures for surface computing.

In 3DUI research, only a few researchers have addressed the problem of 3D interaction on a 2D multi-touch surface so far. In a seminal approach Grossman and Wigdor [80] proposed a taxonomy of 3D on the tabletop. Schöning et al. [174] considered general challenges of multi-touch interaction with stereoscopically rendered projections. First multi-touch 3D interaction techniques (e.g. [35, 36, 85, 161]) have also been proposed. However, most of these interaction techniques have in common that the interaction and visualization is limited to almost zero parallax (i.e. the plane of the interactive surface). This restricts the interaction space more or less to the 2D surface. Hilliges et al. [88] addressed this restriction and proposed interactions above the tabletop. In particular, multi-touch interaction with stereoscopic data leads to the question of where the user is actually touching an object in a stereoscopic projection (cf. [24, 196]). These perceptual issues of touching 3D stereoscopic data have rarely been considered so far and will therefore be covered in this thesis as well.

Natural interaction via touch and gesture offers the potential to bridge the gap towards natural and immersive 3DUI for ubiquitous computing. As stated above, little work on multi-touch 3D interaction exists, especially for stereoscopic multi-touch surfaces. Thus, there is a strong need to better understand the user interacting in such environments. Thus we aim to investigate interactive stereoscopic surfaces of different sizes that go beyond the restriction of the 2D surface. But we also covered perceptual issues that occur during interaction with stereoscopic data. A thesis statement and hypotheses that address this problem are proposed in the following section.

1.3. Thesis Statement

This research addressed the overall question of how users interact with stereoscopically displayed 3D data on the next generation of interactive stereoscopic commodity devices. The interaction can either be restricted to a two-dimensional multi-touch surface or take place in the space above the stereoscopic display. More specifically, this work investigates novel ways to interact with interactive surfaces and can be described by the following thesis statement:

1. Introduction

Usable and natural interaction techniques and UI concepts can be designed for interactive surfaces for interaction with stereoscopic data. In particular, extensions of the interaction space that specifically address these interactive surfaces will lead to effective and usable 3D interaction techniques and 3DUIs, even for commodity 3D devices.

While specific research questions and hypotheses are formulated in each part of the thesis, the following general research questions are investigated in the scope of this thesis:

1. Can usable interactions with stereoscopically displayed 3D data be designed for universal 3DUI tasks?
2. Are these interactions feasible for the 3D manipulation of and navigation in stereoscopic data with commodity devices?
3. How can multi-touch be used to select stereoscopic 3D objects that are displayed with differing parallax?
4. How can mobile and gestural interactions be used to remotely manipulate and navigate stereoscopic data on large screens?
5. How can the phase before the actual (touch) interaction occurs be used to adapt the 3DUI for an improved touch experience (i.e. resolve the spatial discrepancy of touch position and object position)?
6. Is it possible to design usable 3D interactions with stereoscopic mobile devices that allow travel in the scene?
7. Can we extend see-through augmented reality (AR) by mobile devices that provide stereoscopic augmentations, and what are effective (monoscopic and stereoscopic) depth cues for handheld stereoscopic AR?

The above thesis statement and research questions address the ergonomics, perception, and usability of interaction on and above stereoscopic displays and lead to the following goals and methods.

1.4. Methods

The thesis aims to gain profound insights on natural interaction with stereoscopically displayed 3D data. The research goals that encompass this field range from perceptual issues of stereoscopic displays to usability of the gestural interaction for stereoscopic 3DUIs. To achieve this, UIs for stereoscopic displays of different size and interaction modalities are investigated. Affordable commodity devices and devices that require less or no instrumentation are of particular interest.

This thesis evaluates the field of multi-touch and gestural 3D interaction on and above interactive surfaces and explores the design space of interaction with stereoscopic data. The methodology can be divided into several steps. First we perform an exhaustive literature review. Furthermore, a classification of multi-modal 3D interaction on interactive surfaces is proposed. Based on this, we develop and evaluate interactive research probes. Usability research methods (also including psychophysical experiments) are used for the experiments and evaluations. Knowledge is built in a user-centered approach through a number of experiments and studies. This approach results in contributions that are described in the following section.

1.5. Results and Contributions

This work contributes insights on how the next generation of multi-touch enabled stereoscopic displays can be used to provide a natural and immersive user experience. Research probes will be explored that go beyond classical approaches in HCI, AR and VR. An extensible framework and taxonomy for multi-touch interaction with spatial data as well as a concept for *interaction context* will be proposed. 3D interaction techniques will be designed and evaluated. Studying these interaction techniques in canonical 3DUI tasks leads to the conclusion that significant differences exist that arise when displaying objects stereoscopically at differing parallax.

This research mainly focuses on novel and affordable input devices without high user-instrumentation, namely multi-touch surfaces, depth cameras and mobile devices. Almost all of these devices target the mass market, and thus the potential target group is naive to VR and 3DUI. The results have implications for the design of intuitive 3D interaction techniques that might enable VR in the living room or at public places. This results in the following contributions:

- We provide a taxonomy for multi-modal 3D interaction on interactive surfaces, which serves as the basis for the categorization of our later parts of our work (Chapter 3).
- We investigate touch and gestural 3D interaction with stereoscopic data for the canonical 3D tasks (selection, manipulation and travel). Indirect multi-touch 3D selection of objects projected at differing parallax will be designed and studied. The results show that parallax has an effect on selection performance (Chapter 4). Touch and gestural 3D manipulation techniques will be explored and studied with a focus on monoscopic vs. stereoscopic displays. (Chapter 5). 3D travel techniques will be investigated and evaluated with an extensible search task (Chapter 6).
- The concept of *interaction context* will be proposed (Chapter 7) and further explored for the *Reach to Grasp* interaction (Chapter 8). Our studies show that a recognition of the grasp posture during the *Reach to Grasp* phase is feasible within a certain amount of time before the user actually reaches the surface.

1. Introduction

- The extension of VR and AR to stereoscopic handheld devices will be investigated and its general applicability and usability will be shown. We identify issues especially perceptual problems, that need to be carefully addressed and provide guidelines when designing 3D interactions for handheld VR and AR. We are able to show that our sensor-based 3D interaction concepts for handheld stereoscopic devices have proved to work. This implies that sensor-based mobile 3D interaction, when carefully designed, provides an intuitive and joyful means of interaction (Chapter 9). By investigating psychophysical aspects of handheld stereoscopic AR, we identify issues, especially perceptual problems, that need to be carefully addressed and provide guidelines when designing 3D interactions for handheld stereoscopic AR (Chapter 10).

The following publications resulted directly from this dissertation work. The outline of the thesis is reflected by these contributions and is presented afterward.

- | | | |
|-------|---|-----------|
| [42] | F. Daiber. Interaction with stereoscopic data on and above multi-touch surfaces. In <i>Proceedings of the ACM International Conference on Interactive Tabletops and Surfaces</i> , ITS '11, pages 2:1–2:1, New York, NY, USA, 2011. ACM | Chapter 1 |
| [43] | F. Daiber. 3d interaction on and above the surface. In <i>Dagstuhl-Seminar Report, 12151. Schloß Dagstuhl</i> , Schloß Dagstuhl, Germany, 2012 | Chapter 1 |
| [44] | F. Daiber, B. R. De Araujo, F. Steinicke, and W. Stuerzlinger. Interactive Surfaces for Interaction with Stereoscopic 3D (ISIS3D): Tutorial and Workshop at ITS 2013. In <i>Proceedings of the 2013 ACM international conference on Interactive tabletops and surfaces</i> , ITS '13, pages 483–486, New York, NY, USA, 2013. ACM | Chapter 1 |
| [181] | F. Steinicke, H. Benko, F. Daiber, D. Keefe, and J.-B. de la Rivière. Touching the 3rd Dimension (T3D). In <i>CHI '11 Extended Abstracts on Human Factors in Computing Systems</i> , CHI EA '11, pages 161–164, New York, NY, USA, 2011. ACM | Chapter 1 |
| [50] | F. Daiber, A. Krüger, J. Schöning, and J. Müller. Context-sensitive display environments. In A. Krüger and T. Kuflik, editors, <i>Ubiquitous Display Environments</i> , Cognitive Technologies, pages 31–51. Springer Berlin Heidelberg, 2012 | Chapter 3 |
| [52] | F. Daiber, J. Schöning, and A. Krüger. Whole body interaction with geospatial data. In A. B. B. F. . M. Christie, editor, <i>Smart Graphics. Spain</i> , volume 5531/2009, pages 81–92. Springer, 2009 | Chapter 3 |
| [53] | F. Daiber, J. Schöning, and A. Krüger. Towards a framework for whole body interaction with geospatial data. In D. England, editor, <i>Whole Body Interaction</i> , Human-Computer Interaction Series, pages 197–207. Springer London, 2011 | Chapter 3 |

- [171] J. Schöning, F. Daiber, M. Rohs, and A. Krüger. Using hands and feet to navigate and manipulate spatial data. In *CHI '09: CHI '09 extended abstracts on Human factors in computing systems*, New York, NY, USA, 2009. ACM Chapter 3
- [170] J. Schöning, F. Daiber, and A. Krüger. Advanced navigation techniques for spatial information using whole body motion. 2009 Chapter 3
- [45] F. Daiber, E. Falk, and A. Krüger. Balloon Selection revisited - Multi-touch Selection Techniques for Stereoscopic Data. In *Proceedings of the International Conference on Advanced Visual Interfaces*, AVI '12, pages 441–444, New York, NY, USA, 2012. ACM Chapter 4
- [49] F. Daiber, A. Krekhov, M. Speicher, J. Krüger, and A. Krüger. A framework for prototyping and evaluation of sensor-based mobile interaction with stereoscopic 3d. In *Proceedings of ACM ITS Workshop on Interactive Surfaces for Interaction with Stereoscopic 3D (ISIS3D)*, pages 13–16, 2013 Chapter 5
- [54] F. Daiber, M. Speicher, S. Gehring, M. Löchtefeld, and A. Krüger. Interacting with 3d content on stereoscopic displays. In *Proceedings of The International Symposium on Pervasive Displays*, PerDis '14, pages 32:32–32:37, New York, NY, USA, 2014. ACM Chapter 6
- [55] F. Daiber, D. Valkov, F. Steinicke, A. Krüger, and K. H. Hinrichs. Towards Object Prediction based on Hand Postures for Reach to Grasp Interaction. In *CHI 2012 Workshop on Touching the 3rd Dimension of CHI: Touching and Designing 3D User Interfaces*, pages 99–106, 2012 Chapter 8
- [197] D. Valkov, F. Steinicke, G. Bruder, K. H. Hinrichs, J. Schöning, F. Daiber, and A. Krüger. Touching floating objects in projection-based virtual reality environments. In *Joint Virtual Reality Conference*. Eurographics, 2010 Chapter 8
- [51] F. Daiber, L. Li, and A. Krüger. Designing gestures for mobile 3d gaming. In *Proceedings of the 11th International Conference on Mobile and Ubiquitous Multimedia*, MUM '12, New York, NY, USA, 2012. ACM Chapter 9
- [109] F. Kerber, P. Lessel, M. Mauderer, F. Daiber, A. Oulasvirta, and A. Krüger. Is autostereoscopy useful for handheld ar? In *Proceedings of the 12th International Conference on Mobile and Ubiquitous Multimedia*, MUM '13, pages 4:1–4:4, New York, NY, USA, 2013. ACM Chapter 10

1. Introduction

The following publications informed this work as well. Although they are not directly linked to the thesis, the research in these publications inspired this dissertation in many ways (e.g. by exploring other 3D input devices or modalities):

- [46] F. Daiber, S. Gehring, M. Löchtefeld, and A. Krüger. Touchposing - multi-modal interaction with geospatial data. In *Proceedings of the 11th International Conference on Mobile and Ubiquitous Multimedia*, MUM '12, New York, NY, USA, 2012. ACM
- [47] F. Daiber, F. Kosmalla, and A. Krüger. Boulder: Using augmented reality to support collaborative boulder training. In *CHI '13 Extended Abstracts on Human Factors in Computing Systems*, CHI EA '13, pages 949–954, New York, NY, USA, 2013. ACM
- [48] F. Daiber, F. Kosmalla, M. Löchtefeld, S. Gehring, and A. Krüger. Hand-held augmented reality for collaborative boulder training. In *Proceedings of ACM CHI Workshop: HCI and Sports*, 2014
- [75] S. Gehring, M. Löchtefeld, F. Daiber, M. Böhmer, and A. Krüger. Using intelligent natural user interfaces to support sales conversations. In *Proceedings of the 2012 ACM international conference on Intelligent User Interfaces*, IUI '12, pages 97–100, New York, NY, USA, 2012. ACM
- [74] S. Gehring, M. Löchtefeld, F. Daiber, M. Böhmer, and A. Krüger. Using intelligent natural user interfaces to support sales conversations. In *Proceedings of the 2012 ACM international conference on Intelligent User Interfaces*, IUI '12, pages 97–100, New York, NY, USA, 2012. ACM
- [96] S. Hoppe, F. Daiber, and M. Löchtefeld. Eype - using eye-traces for eye-typing. In *CHI 2013 Workshop on Grand Challenges in Text Entry*, 2013
- [108] F. Kerber, P. Lessel, F. Daiber, and A. Krüger. Shift 'n' touch: Combining wii balance board and cubtile. In *Proceedings of the 7th Nordic Conference on Human-Computer Interaction: Making Sense Through Design*, NordiCHI '12, pages 789–790, New York, NY, USA, 2012. ACM
- [113] F. Kosmalla, F. Daiber, and A. Krüger. Climbsense: Automatic climbing route recognition using wrist-worn inertia measurement units. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*, CHI '15, pages 2033–2042, New York, NY, USA, 2015. ACM
- [131] M. Löchtefeld, S. Gehring, J. Schöning, F. Daiber, and A. Krüger. Tracking pointing gestures to support sales conversations. In *Adjunct Proceedings of the 28th International Conference on Human Factors in Computing Systems. Workshop on Performative Interaction in Public Spaces*. ACM, 2011
- [141] M. Mauderer, F. Daiber, and A. Krüger. Combining Touch and Gaze for Distant Selection in a Tabletop Setting. In *CHI 2013 Workshop on Gaze Interaction in the Post-WIMP World*, 2013

- [178] G. Sörös, F. Daiber, and T. Weller. Cyclo: A personal bike coach through the glass. In *SIGGRAPH Asia 2013 Symposium on Mobile Graphics and Interactive Applications*, SA '13, pages 99:1–99:4, New York, NY, USA, 2013. ACM
- [197] D. Valkov, F. Steinicke, G. Bruder, K. H. Hinrichs, J. Schöning, F. Daiber, and A. Krüger. Touching floating objects in projection-based virtual reality environments. In *Joint Virtual Reality Conference*. Eurographics, 2010
- [198] U. von Zadow, F. Daiber, J. Schöning, and A. Krüger. Globaldata: Multi-user interaction with geographic information systems on interactive surfaces. In *ACM International Conference on Interactive Tabletops and Surfaces*, ITS '10, pages 318–318, New York, NY, USA, 2010. ACM

The following Bachelor's and Master's Theses that I supervised also directly or indirectly informed this dissertation:

- [5] M. Barz. Computational modeling and prediction of gaze estimation error for head-mounted eye trackers. Master's thesis, Saarland University, Saarbrücken, Germany, Department of Computer Science, 2015
- [31] A. Chernov. A method for 3d reconstruction of a foot with kinect. Bachelor's thesis, Saarland University, Saarbrücken, Germany, Department of Computer Science, 2014
- [67] E. Falk. Multi-touch selection techniques for stereoscopic 3d content. Master's thesis, Saarland University, Saarbrücken, Germany, Department of Computer Science, 2011
- [70] P. Flotho. Persisten user identification with the kinect. Bachelor's thesis, Saarland University, Saarbrücken, Germany, Department of Computer Science, 2013
- [72] A. Freund. Mobicube: A novel approach to 3d menus on mobile devices - a comparative study on 2d vs. 3d mobile menus. Bachelor's thesis, Saarland University, Saarbrücken, Germany, Department of Computer Science, 2015
- [107] F. Kerber. Openindoormap - smartphone-based capture of uninstrumented indoor environments. Master's thesis, Saarland University, Saarbrücken, Germany, Department of Computer Science, 2012
- [111] F. Kosmalla. Boulder: Design and evaluation of a mobile augmented reality system for collaborative boulder training. Bachelor's thesis, Saarland University, Saarbrücken, Germany, Department of Computer Science, 2013
- [112] F. Kosmalla. Climbsense - automatic climbing route recognition using wrist-worn inertia measurement units. Master's thesis, Saarland University, Saarbrücken, Germany, Department of Computer Science, 2014

1. Introduction

- [126] L. Li. Interaction with stereoscopic data displayed on mobile devices. Master's thesis, Saarland University, Saarbrücken, Germany, Department of Computer Science, 2011
- [140] M. Mauderer. Combining touch and gaze for distant selection. Master's thesis, Saarland University, Saarbrücken, Germany, Department of Computer Science, 2012
- [179] M. Speicher. Exploring 3d interaction techniques for stereoscopic content using consumer tracking devices. Master's thesis, Saarland University, Saarbrücken, Germany, Department of Computer Science, 2014

1.6. Thesis Overview and Conventions Used

The thesis is structured as follows. In Chapter 2, the background and related work in the variety of fields that are relevant for this work is discussed in depth. Chapter 3 presents an extensible framework and taxonomy for multi-touch interaction with spatial stereoscopic data. Part II discusses 3D interaction techniques for basic 3DUI tasks, namely 3D selection (Chapter 4), 3D manipulation (Chapter 5) and 3D navigation (Chapter 6). Part III discusses interaction that goes beyond traditional approaches of surface computing. In Chapter 7 the concept of *interaction context* is proposed, which goes beyond touch by also incorporating the space above the device in the touch interaction lifecycle. In Chapter 8 the *Reach to Grasp* task will be investigated in more depth as an instance of *interaction context*. Part IV discusses interaction with handheld stereoscopic devices in more detail by investigating methods for mobile stereoscopic interaction (Chapter 9) and studying perceptual issues in stereoscopic handheld AR (Chapter 10). Finally, the thesis is summarized with a conclusion and a discussion of future work in Chapter 11.

To improve the readability of the thesis the following writing conventions are used:

- A part is denoted by a roman numeral (I), a single number X denotes a chapter, two numbers separated by a period (X.X) are called sections and all other parts are called subsections.
- Neutral persons will be referred to using the female pronoun only (she instead of he or she).
- All links are provided as footnotes. All links were last accessed March 1st, 2015. Links to companies, products, websites, etc. are listed only at first mention. Trademark symbols are not used.
- Although the thesis represents the work of a single PhD candidate, it is written using plural pronouns. This is done to acknowledge the contribution of others without whom this work would not have been possible.

2. Perceptual and Technical Foundations of Interactive Stereoscopic Surfaces

This chapter aims to cover the wide range of fields of this thesis. In particular, an overview of the background and related work is given, focusing mainly on HCI and 3DUI. The chapter is structured as follows. (1) Depth perception and stereoscopic vision are a prerequisite for the interaction with stereoscopic data. Thus, a brief introduction to depth cues that allow humans to visually perceive depth is given. Then, perceptual issues of devices that are investigated in this thesis, namely touch interactive stereoscopic displays and handheld stereoscopic displays, are discussed in detail. (2) An overview of 3D technologies is given. First, the generation of stereoscopic output is introduced, followed by a brief overview of visual 3D display technologies and the most relevant technologies that are best suited for interaction with stereoscopic data on and above interactive surfaces. Second, 3D input techniques are explored, starting with an overview of 3D input devices and followed by a discussion of commodity input hardware that is of particular interest in the scope of this thesis. (3) The main part of this chapter covers the discussion of related work of 3DUIs and 3D interaction. After an introduction to 3DUI and its canonical 3D tasks, related 3D interaction techniques are discussed, in particular touch and gestures, hand postures and grasp, gaze, and handheld interaction. Finally, a conclusion sets this thesis in the context of the concepts and related work discussed in the chapter.

2.1. Stereoscopic Vision and Depth Perception

A variety of depth cues allow humans to visually perceive depth. In this section human depth perception is briefly introduced. For stereoscopic 3D displays the binocular disparity and stereopsis depth cues are of main interest and will thus be addressed in more depth, followed by a review of research on depth perception on stereoscopic displays. Touch-enabled stereoscopic displays may cause additional perceptual issues that need to be carefully taken into account when designing interactions for this class of stereoscopic displays.

2. Perceptual and Technical Foundations of Interactive Stereoscopic Surfaces

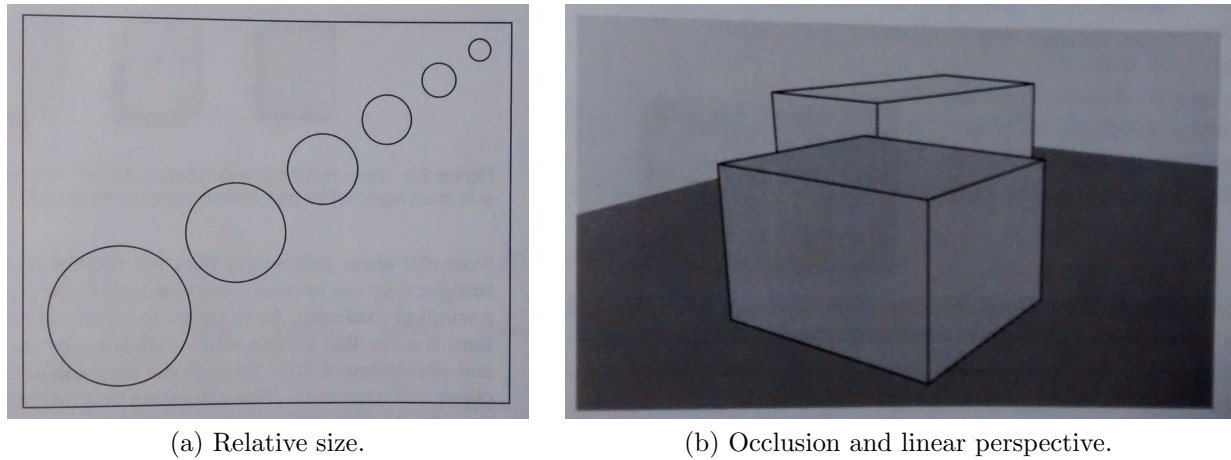


Figure 2.1.: Monoscopic depth cues [21].

2.1.1. Depth Cues

Depth cues are differentiated into monocular and stereoscopic depth cues. Monocular cues are static cues that can be perceived with one eye. Stereoscopic cues are dynamic and require either interactive changes of the scene's viewport or physical changes of both eyes.

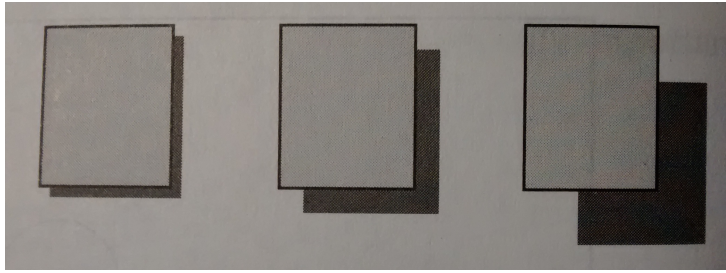
Monocular Depth Cues

Monocular, static cues can be perceived by looking with one-eye (monocularly) at features like perspective, relative size, occlusion, shadow, lighting and gradients that can be perceived even in a single (static) image [21]. The monoscopic depth cues are briefly introduced in the following.

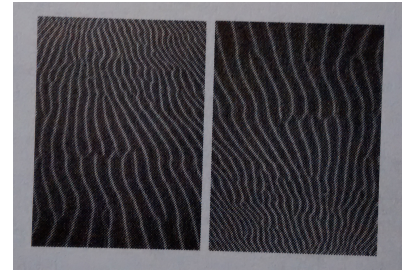
Relative size of objects (e.g. two persons of similar size who stand at different distances from the viewer) is a strong depth cue known as the relative size cue (see Figure 2.1a). Additionally, the height of objects relative to the horizon can influence the impression of depth.

Occlusion occurs when one object is partially occluded by another (see Figure 2.1b). This effect guides the viewer in determining which object is closer and which one is further away.

Linear and aerial perspective represents the effect of how the appearance of objects changes with increasing distance from the viewer. Linear perspective is the cue that occurs when parallel lines appear to converge at a far distance (a common example is train rails that appear to converge in distance). Aerial perspective is affected by atmospheric



(c) Shadows.



(d) Texture gradients.

Figure 2.0.: Monoscopic depth cues [21].

illumination of objects, i.e. the color and saturation of objects changes when positioned at different depths.

Shadows imply a light source that is positioned with respect to an object (see Figure 2.1c). Depth is perceived based on the viewer's assumptions about the spatial relation to this light source. This effect can also lead to depth illusions. Lighting itself affects depth perception because objects that are illuminated more brightly appear closer to the viewer.

Texture gradients also provide (relative) depth cues (see Figure 2.1d). These gradients emerge from the surface structure that is represented by textures.

In summary, monocular, static cues can be perceived by looking with one-eye (monocularly) and can even be perceived in a single (static) image. These cues provide a strong perception of depth without stereoscopic technology. But since they also affect natural depth perception, one has to keep in mind that they also exist in stereoscopic virtual scenes. Thus they need to be carefully considered as an experimental condition, or excluded when investigating stereoscopic cues in experiments. The stereoscopic depth cues that can be used to virtually generate 3D scenes are discussed next.

Stereoscopic Depth Cues

Stereoscopic depth cues can be perceived by interactive changes of the scene's viewport (motion parallax) or by physical changes of both eyes (vergence and binocular disparity) when looking at different objects, which results in different depth impressions [21]. The stereoscopic depth cues are depicted in Figure 2.1 and briefly discussed below.

Motion parallax creates the impression of depth when the scene is moving relative to the viewer, i.e. 1) when objects are moving relative to a stationary viewer, 2) when the viewer is moving relative to stationary objects or 3) both viewer and objects are moving accordingly. Motion parallax is an intuitive depth cue that has been introduced in recent web and (mobile) game design. The effect of moving different objects with different speed

2. Perceptual and Technical Foundations of Interactive Stereoscopic Surfaces

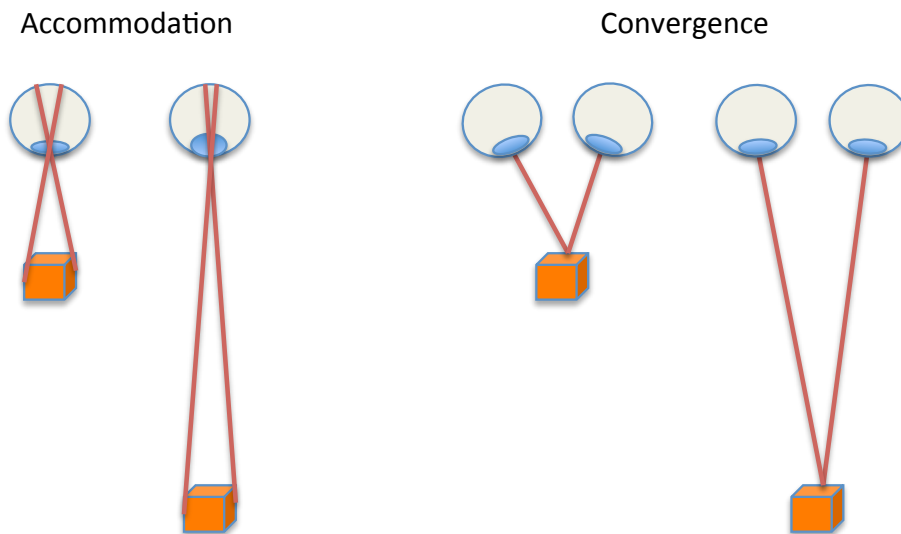


Figure 2.1.: Stereoscopic depth cues: accommodation and convergence.

(e.g. scrolling a website or navigating in a game) generates an impression of depth in an otherwise 2D interface.

Vergence (accommodation and convergence) relies on muscular tension of the eyes referred to as the oculomotor cues. Accommodation is the physical deformation of the eye lens in order to focus on the distance of an object. The eye muscles are stretched to flatten the eye lens to focus on nearby objects and are relaxed to focus on distant objects. Convergence is the rotation of the eyeballs in order to fuse the images from each eye correctly. The eyes converge when viewing nearby objects and diverge when viewing far-away objects. These depth cues are crucial for stereoscopic display technology due to the accommodation-convergence conflict (see Figure 2.2). This conflict results from different positions that accommodation and convergence focus when viewing stereoscopically projected objects, by converging to the projected object but accommodating to the projection screen.

Binocular disparity and stereopsis is produced by the fact that (most) humans view their environment with two eyes. This results in two slightly different images for the left and the right eye, which is referred to as the binocular parallax. The fusion of these two images (through accommodation and convergence) into one single stereoscopic image is called stereopsis. This strong stereoscopic depth cue tends to be much more pronounced the closer the object is to the viewer [21]. However, few studies exist that address this problem and describe these relations for small stereoscopic displays. This motivated us to conduct our perception study on small stereoscopic displays for handheld AR (see Chapter 10).

To sum up, all these depth cues (monocular and stereoscopic) enable humans to perceive depth. Stereoscopic cues are mainly used to produce the impression of depth in immersive

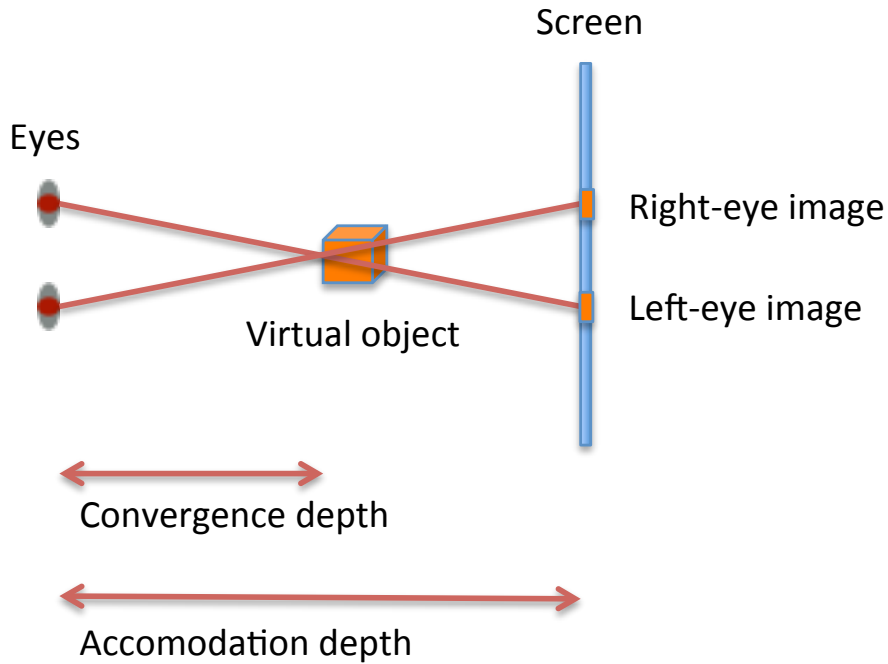


Figure 2.2.: Accommodation-convergence conflict.

VR setups. In the following section the depth perception of touch-interactive stereoscopic displays is discussed. These depth cues are also used to determine distances. While all these cues can be used to assess distances in real environments, some are not available in virtual and mixed realities [60]. Thus, depth perception on handheld stereoscopic displays for AR is discussed afterwards.

2.1.2. Depth Perception on Stereoscopic Multi-Touch Displays

In this section we discuss the perceptual issues of stereoscopic displays that support touch input (3D workbenches and autostereoscopic mobile devices). First, the perceptual challenges of interaction with stereoscopically displayed objects that are manipulated by touching a 2D surface are discussed. Second, depth interpretation in handheld AR is reviewed.

As mentioned above, stereoscopic effects on screens are achieved by showing each eye of an observer a different image. As can be seen in Figure 2.3 the arrangement of these two images forces the eyes to converge accordingly which results in different parallaxes. On stereoscopic displays objects can be displayed with different parallax resulting in different stereoscopic effects. Objects may appear behind (positive parallax), on top (zero parallax), or in front (negative parallax) of the screen. The effect of objects floating in front of the screen is reached while the depth cues the brain obtains are ambiguous. The eye's convergence presumes that two different images are seen, but the eyes need to focus on the screen instead of the objects in front. This leads to an accommodation contradictory to the convergence, i.e. the already mentioned accommodation-convergence conflict. But

2. Perceptual and Technical Foundations of Interactive Stereoscopic Surfaces

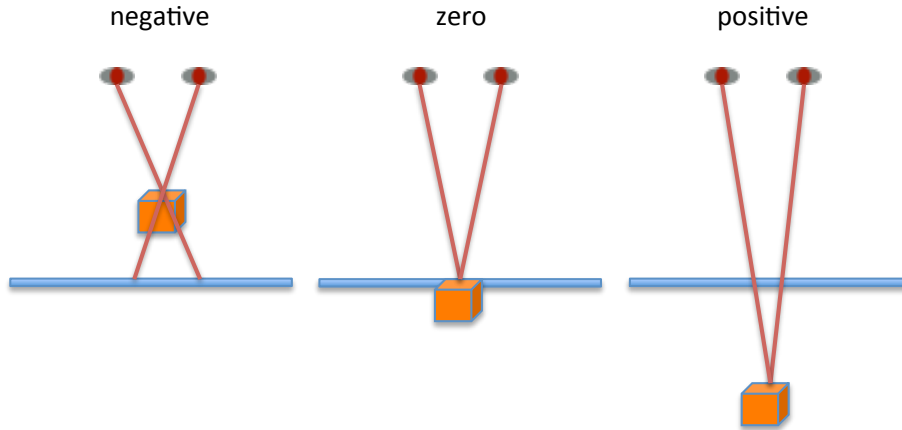


Figure 2.3.: Parallax spaces: On stereoscopic displays objects may appear in front of (negative parallax), on top of (zero parallax), or behind (positive parallax) the screen.

when it comes to touch interaction with stereoscopic displays this effect can get even more critical, and especially for negative parallax, additional degrees of ambiguity become relevant. Objects appearing in front of the screen are clearly behind the user's hand when touching the screen. A result of the inconsistency between accommodation and convergence is that if the user's hand is focused on, the stereoscopic effect gets lost, whereas focusing on the scene objects may be inappropriate for tasks requiring the user's attention.

In a study on touching floating objects on stereoscopic display walls, Valkov et al. [197] addressed the perceptual challenges that occur when users interact with stereoscopically displayed objects and the input is constrained to a 2D (touch) surface (see Figure 2.4). The study reveals that objects displayed with positive parallax cannot be accessed by direct touch interaction, since the screen surface limits the user's reach. While one can use indirect selection and manipulation techniques for such objects, it is difficult to apply these techniques to objects in front of the screen [190]. In a follow-up study, Valkov et al. [196] found that the user is limited to touch interaction on the area behind the object, since without additional instrumentation, touch feedback is only provided at the surface. Therefore the user has to reach through the visual object to reach the touch surface with her finger. If the user reaches into an object while focusing on her finger, the stereoscopic effect for the object will be disturbed, since the user's eyes are not accommodated and converged on the projection screen's surface. Thus the left and right stereoscopic images of the object's projection would appear blurred and could not be merged anymore. However, focusing on the virtual object leads to a disturbance of the stereoscopic perception of the user's finger, since her eyes are converged to the object's 3D position. In both cases touching an object may become ambiguous [196].

However, as suggested by Valkov et al. [197], users are less sensitive to discrepancies between visual penetration and touch feedback when they try to touch stereoscopic objects which are displayed close to the surface. In particular, they found that users are less



Figure 2.4.: The problem of touching stereoscopically objects [197].

sensitive to discrepancies between visual and tactile feedback if objects are displayed with negative parallax. In the monoscopic case, the mapping between an on-surface touch point and the intended object point in the virtual scene is straightforward, but with stereoscopic projection this mapping introduces problems [190]. Since there are different projections for each eye, the question arises: where do users touch the surface when they try to “touch” a stereoscopic object? In principle, the user may touch anywhere on the surface to select a stereoscopically displayed object. However, perceptual experiments reveal that users actually touch an intermediate point that is located between both projections with a significant offset to the user’s dominant eye [196].

This illustrates how perceptual aspects limit humans during the pre-touch phase. However, these implications can drive novel interface concepts. We built our concept of interaction context based on these insights (see Chapter 7). One goal is to adapt the 3D scene to allow a passive haptic sensation. For instance, when interacting with stereoscopically displayed virtual objects, one could shift a virtual object before the user actually touches the surface in such a way that the object appears exactly on the surface at the moment of touch. The dominance of vision over tactile feedback might then evoke tactual illusions (i.e. induce false haptic sensations that reflect visual properties of the object such as texture). Furthermore, taking into account this first phase before the user actually touches the surface, other adaptations of the user interface can be applied. In contrast to large stereoscopic displays, depth perception on small, mobile stereoscopic displays is even more critical, especially in AR, which is discussed in detail in the next section.

2.1.3. Depth Perception on Handheld Stereoscopic Displays for AR

Depth perception is even more critical when it comes to small, mobile stereoscopic displays. In the following AR is briefly introduced and discussed, with a special focus on depth interpretation in AR.



(a) The first AR system by Sutherland [186].



(b) The first mobile AR system by Feiner et al. [68].

Figure 2.5.: Early examples of AR systems.

History of Handheld AR

The first AR system was presented by Sutherland in 1968 [186] (see Figure 2.5a). The system consists of a tracked (in 6-degrees of freedom (DOF)) optical see-through head-mounted display that projects simple wireframe graphics. With the release of the first laptops and the global positioning system (GPS), the first mobile AR systems appeared. One of the first systems was the Touring Machine by Feiner et al. [68] (see Figure 2.5b). They presented a head-tracked, see-through head mounted display connected to a laptop in a backpack. In addition, a handheld display was used as an input device. In their approach tracking was realized by a magnetometer in the handheld display and GPS. Rekimoto [163] introduced one of the first marker-based AR systems. Tracking was realized with computer vision and in contrast to the previous systems, virtual objects could be accurately registered in the real world. This visual tracking approach has attracted a large body of AR research (see Wagner and Schmalstieg [199] for an early example of handheld AR). Today smartphones are powerful enough to process computer vision for AR and thus many commercial mobile AR applications exist.

As a first approach towards a definition for AR, Milgram and Kishino [146] proposed the reality-virtuality continuum. In their understanding, UIs can be aligned within this continuum according to their proportion of reality and virtuality (see Figure 2.6). The real environment consists exclusively of real objects. At the other end of the continuum, in the virtual environment everything is computer generated. A mixed reality (MR) environment is located between the real environment and the virtual environment and thus consists of both real world and virtual objects.

According to a second definition by Azuma [2] an AR application needs to fulfill at least the following three properties: It combines real and virtual images, such that both can be

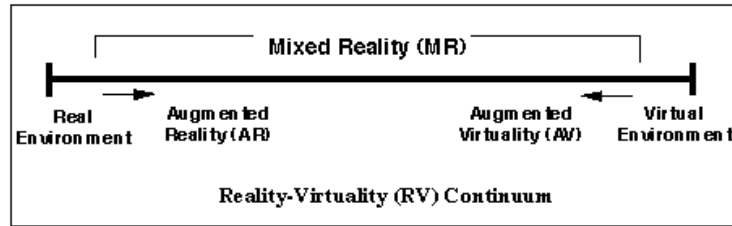


Figure 2.6.: The reality-virtuality continuum by Milgram and Kishino [147].

seen at the same time (1); it is interactive, and virtual content can be manipulated in real time (2); and it is registered in 3D, which leads to the impression that virtual objects are fixed in real space (3).

Depth perception in AR

Depth perception in MR and AR (mainly using head-mounted displays (HMDs)) has been investigated in indoor and outdoor environments. It has been shown, that people often underestimate depth in indoor environments, while they overestimate depth in outdoor environments [60, 110, 130, 144, 187].

Handheld see-through AR relies on the tool glass and magic lenses metaphor [10]. While this concept has been widely adopted, relatively little literature on depth perception on magic lens displays exists (e.g. [60]). In handheld AR, depth interpretation is a common problem and creating a perceptually correct augmentation is still a challenge [116]. Flat augmentations in the magic lens display are not suited very well to guide the user's depth perception (see Figure 2.7). Not much research has investigated the use of autostereoscopic mobile devices for AR. Nevertheless, some relevant work on depth perception on mobile devices exists.

Dey et al. [61] investigated depth perception on handheld devices with different screen sizes (i.e. iPad and iPhone). None of these devices had autostereoscopic capabilities. The results showed that there is no significant effect of screen resolution on depth perception, but there is an effect on distance estimation.

Huhtala et al. [97] investigated whether autostereoscopy could help users in a selection task where relevant parts were highlighted. In an experiment the participants performed a simple thumbnail visual search task, where color shading, horizontal disparity, and a combination of the two were used to highlight the relevant parts. Thus, two of the four conditions involve autostereoscopic cues, but the results did not show that using stereoscopy alone improved the performance. However, the combination with a second visual cue performed better.

The work of Mikkola et al. [145] considered the importance of different depth cues on a mobile autostereoscopic display. Participants were presented with several virtual balls that had been placed at different depths on a virtual background. For different depth cues the participants had to decide which of the balls was at the same depth as a reference object.



Figure 2.7.: The problem of depth perception in AR.

The results showed that the stereoscopic depth cues outperform the monocular ones in accuracy and speed of depth estimation.

To conclude, little work has been done in the field of depth interpretation in handheld AR. We believe stereoscopic AR can improve the creation of perceptually correct augmentations in handheld AR. Thus, we investigate stereoscopic handheld AR starting with a study on depth perception in an autostereoscopic handheld AR setup (see Chapter 10).

In the next section 3D output and input technologies are discussed. The basics of depth perception introduced above need to be kept in mind when designing 3D technologies. 3D output devices that are not carefully designed to provide correct stereoscopic images might break the depth perception, which could even lead to cybersickness [123]. 3D input can also influence depth perception; for example, touch-based input can disturb depth perception when the touching hand enters the field of view.

2.2. 3D Technology

This section provides a brief overview of related 3D technologies. Both 3D output and input technologies that are most relevant will be introduced. This includes a thorough discussion of commodity hardware that is of particular interest in the scope of this thesis, since it has the potential to enable 3DUI for everyday use.

2.2.1. Output Technology

There has been a long research tradition in output devices for 3D virtual environments and augmented reality, ranging from highly immersive cave setups with fully instrumented users



Figure 2.8.: Stereo glasses: active shutter glasses (left), passive polarized glasses (middle), passive spectral (anaglyph) glasses (right).

wearing gloves and goggles, to light weight (auto-)stereoscopic display technologies. After a short introduction to the generation of stereoscopic imagery, an overview of 3D display technologies is given. Special emphasis is placed on autostereoscopy and other technologies that allow 3D visual display with low instrumentation. This class of stereoscopic displays is best suited for combination with multi-touch and gestural input technologies because it allows natural interaction without the need for heavy user instrumentation. Nevertheless, standard monitors, HMD and volumetric displays are also briefly introduced to complete the whole picture of stereoscopic display technologies.

Stereoscopic Output

In general, stereoscopic output that relies on binocular disparity is produced by the generation of different images for each of the two eyes (the binocular disparity and the stereopsis effect is discussed in more depth in Section 2.1). The most common way to generate stereoscopic images uses stereo glasses that can be either active or passive (see Figure 2.8).

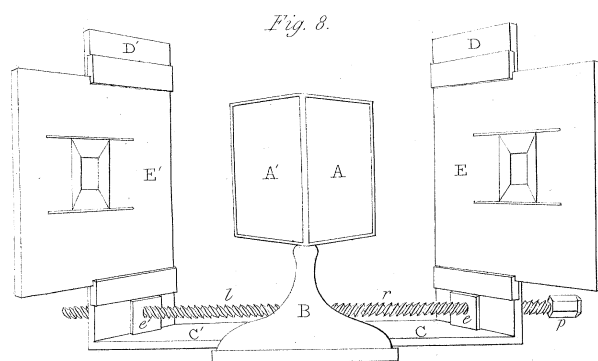
Active stereo glasses, also known as shutter glasses, use temporal multiplexing to project images to each eye. By opening and closing the shutters for each eye at a rate that is synchronized with the display's refresh rate, only one eye at a time sees the display in an alternating order. The human visual system is able to fuse a stereo pair within a time lag of up to 50 ms [155] and thus the different stereoscopic images that are projected to the corresponding eyes generate a binocular depth cue.

In passive stereo the separation of the images for each of the two eyes is realized by passive filters or a separation of the two displays. Passive stereo can be generated either by polarization, spectral or location multiplexing.

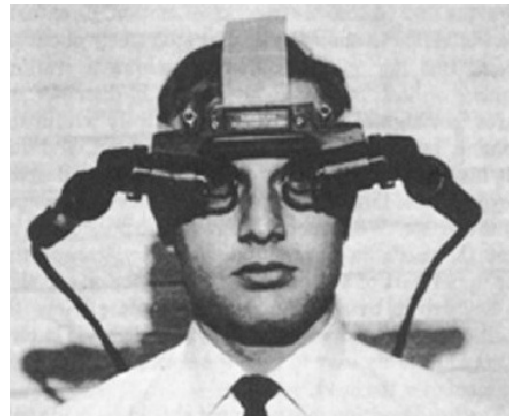
Polarization multiplexing separates the two overlaid stereoscopic images by oppositely polarized filters. Vertical and horizontal polarized filters or clockwise and counter-clockwise circular polarized filters can be used in combination with corresponding filter glasses.

Spectral multiplexing uses different colors to display the overlaid stereoscopic images. The corresponding glasses (anaglyph glasses) use color filters that allow only one color to pass through the filter (e.g. cyan/red, red/blue, red/green). Anaglyph stereo can be produced with any color display but this approach obviously has some color limitations.

2. Perceptual and Technical Foundations of Interactive Stereoscopic Surfaces



(a) The Wheatstone stereoscope [206], the first stereoscopic display.



(b) The first HMD [186].

Figure 2.9.: Location multiplexing examples.

For example, when using cyan/red glasses, false colors can be perceived. Nevertheless, this technology is well suited for research, because it can be used for rapid prototyping, and device classes that do not yet provide stereoscopic output can be investigated using anaglyph stereo.

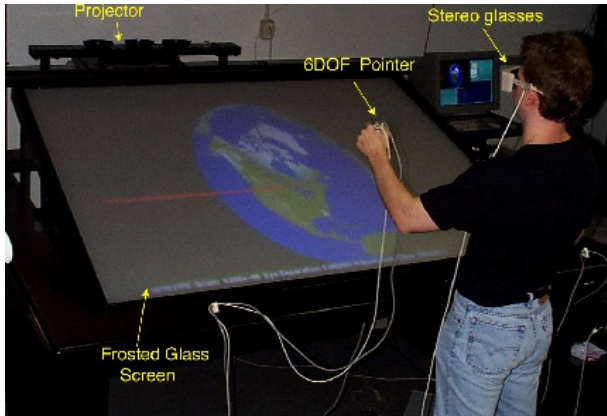
Location multiplexing projects the two images to two different locations where they can only be perceived by one of the eyes. The *Wheatstone stereoscope* [206], the first stereoscopic display, used two mirrors that each provided a different image to the viewer (see Figure 2.9a). Another example of this approach is HMD technology (see Figure 2.9b).

An approach to generate stereoscopic output without the need for stereo glasses is to use autostereoscopic displays. In the following section an overview of display technologies for 3DUI is given, including a discussion of autostereoscopic displays.

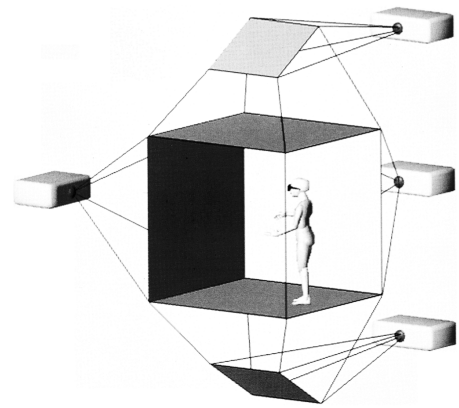
Overview of Visual Displays for 3DUI

Visual displays for 3DUI are introduced that range from standard monitors to autostereoscopic displays. The display classes are then set in the context of a simple classification of 3D displays in order to provide an overview and initial insights on the applicability of these technologies for light weight displays that require less instrumentation and in addition are best suited for touch and gestural interactive 3D surfaces.

(1) Standard monitors are widely used for 3DUI for both monoscopic as well as stereoscopic displays. Many commercial applications for scientific visualization, computer aided design (CAD) and 3D modeling in general are designed for this device class. These devices are partially suited for multi-touch 3DUI when enhanced with additional tracking hard-



(a) The Perceptive Workbench [125].



(b) The CAVE [40].

Figure 2.10.: Projected 3D Displays.

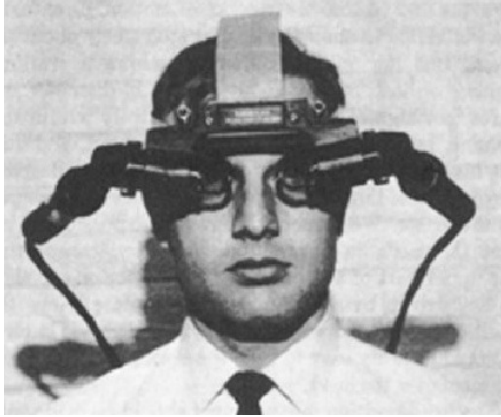
ware. For this reason, we have used such setups in our studies within the scope of this thesis as well.

(2) Projected 3D displays allow large-scale setups ranging from workbenches up to surround screen setups (see Figure 2.10). Workbenches are mid-size displays that often provide high-resolution 3D displays (see Figure 2.10a). Tabletop and wall-screen workbench setups also exist, as to combinations of the two. Some of these workbenches already provide touch sensitive displays. Surround-screen displays consist of multiple projected wall-sized displays (more than three) that surround the user and allow her to move within a certain range (see Figure 2.10b). Surround-screen setups provide a good immersion for the user but they are very expensive and hard to maintain. Thus, we ignored surround screen displays in our studies and focused instead on workbench setups. We have used different setups from tabletops to wall-sized multi-touch enabled stereoscopic projections.

(3) Head-mounted displays are another important class of 3D displays. In an HMD the image is directly projected on one (mono) or two (stereo) displays in front of the eyes (see Figure 2.11). Thus an HMD allows a complete physical immersion by providing a 360° field of regard (FOR). However they often provide a small and unnatural field of view (FOV) ranging from 30° to 60° which can lead to perception and performance problems. HMDs have some other drawbacks as well. Most of them are bulky and thus cause ergonomic problems like head strain. They are also very expensive and have only been used by a few research groups and in very specialized application domains. Most recently a low-cost HMD solution has been proposed with the Oculus Rift¹ that targets the 3D gaming market (see Figure 2.11b). This device has the potential to reinspire HMD research and

¹<https://www.oculusvr.com>

2. Perceptual and Technical Foundations of Interactive Stereoscopic Surfaces



(a) The first HMD [186].



(b) The Oculus Rift.

Figure 2.11.: HMD examples.

open HMD 3DUIs to a broad audience. Since these commodity devices did not exist until very recently, we excluded this class of stereoscopic displays from our studies.

(4) Autostereoscopic displays provide 3D images without the need to wear additional glasses. Lenticular displays, volumetric displays and holographic displays fall into this category. Lenticular displays use parallax barriers [101, 100] or cylindrical lenses [128] to separate different parts of the displays for one eye from the pixels for the other eye. The main drawback of this approach is that the user has to remain in a fixed position in relation to the device. This makes it hardly usable for mobile displays that are frequently moved. A volumetric display is based on a transparent physical volume in which image components are displayed [11]. Most of these displays build images using 3D pixels (voxels) by using swept- or static-volume techniques [3]. Holographic displays use a different technique to produce 3D imagery. The holographic approaches basically record and reproduce the properties of light waves that are emitted from a real 3D scene [133]. Both volumetric and holographic displays produce 3D imagery that can be perceived from multiple positions by multiple users without glasses. However, volumetric and holographic display techniques are still immature and have many technical issues (e.g. display only small-size volumes and cannot provide many monoscopic depth cues) [21]. Due to these technological challenges we ignored volumetric and holographic displays in our studies. Instead we explore displays that use the parallax barrier technique by investigating the perception and interaction design of mobile autostereoscopic devices that use this technique.

A classification for 3D displays was proposed by Blundell [12]. This simple classification consists of four general classes of 3D displays: Monocular, Stereoscopic, Autostereoscopic Class I, and Autostereoscopic Class II (see Table 2.1). While standard monitors fall in the Monocular class, these displays can easily be extended to Monocular tracked 3D displays by adding head tracking. In combination with 3D glasses these monitors can even be used

Monocular	Stereoscopic	Autostereoscopic (Class I)	Autostereoscopic (Class II)
The conventional flat screen display	<div> <div>Chromatically coded (anaglyph)</div> <div>Non-coded</div> <div>Temporally coded</div> <div>Spatially coded</div> </div>	Immersive/augmented virtual reality	<div>Volumetric</div> <div>Varifocal</div> <div>Holographic</div>
Support for only pictorial depth cues	Support for pictorial depth cues and binocular parallax	Support for pictorial depth cues, binocular and motion parallax	Support for pictorial depth cues, binocular and motion parallax and oculomotor cues
Direct viewing	Direct or indirect viewing (via glasses)	Direct or indirect viewing (via glasses or other headgear)	Direct viewing
No head tracking	No head tracking	May or may not require head tracking	No head tracking required

Table 2.1.: Blundell’s classification for 3D displays [12].

for the display of stereoscopic data (Stereoscopic class). Blundell distinguishes between two categories of autostereoscopic displays (Autostereoscopic Class I and Autostereoscopic Class II). They both provide glasses-free stereoscopic displays but differ in varying degrees of (natural) stereoscopic perception. Class I displays use a parallax barrier to split the display into multiple views that can be perceived from one or more positions. In contrast to Class I displays, Class II displays support accommodation and convergence and thus avoid the accommodation-convergence conflict (see Section 2.1 for more details on perceptual aspects of 3D displays).

Blundell’s classification gives a fairly good indication for choosing an appropriate stereoscopic display technology. However, since this classification mainly covers output, additional considerations need to be taken into account when designing interactions for stereoscopic displays. Since few market-ready technologies exist, the combination of a stereoscopic display with an appropriate input device is a crucial task. Stereoscopic, Stereoscopic tracked, Autostereoscopic Class I displays were mainly used in this work because they were assumed to be best suited for the combination with commodity input technologies such as multi-touch. Potential multi-touch stereoscopic display technologies are explored in more detail in the following section.

Visual Displays for Interactive Surfaces

The display technologies introduced above range from fully immersive setups that require highly instrumented users to light weight (auto-)stereoscopic display technologies. We aim for light weight technologies in order to enable 3D touch technology for intuitive and natural interaction without high instrumentation.

Only a few prototypes (e.g. [34, 56, 185, 197]) and almost no commercial solutions at all exist in the field of projected multi-touch 3D displays. One outstanding exception is the Ilight 3D Touch, a multi-user multi-touch tabletop with shutter-based 3D [57].

2. Perceptual and Technical Foundations of Interactive Stereoscopic Surfaces

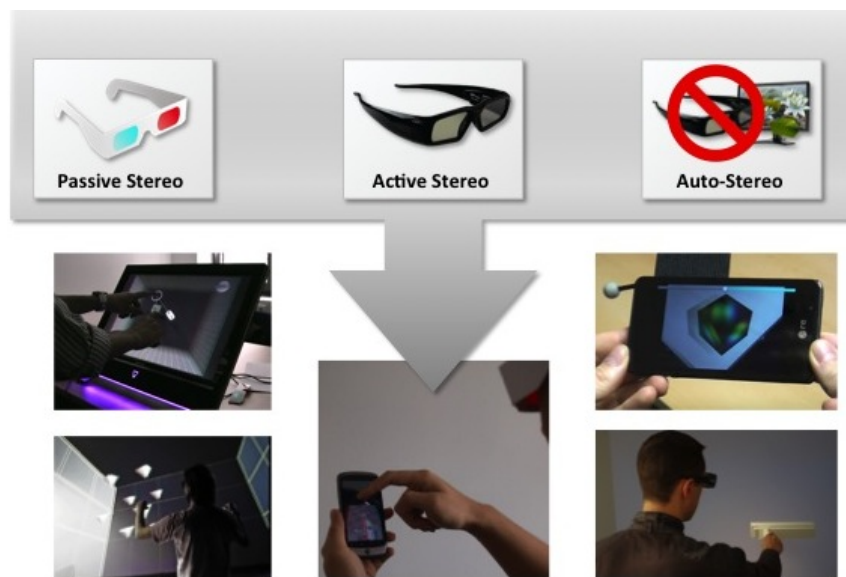


Figure 2.12.: 3D technology that was used in this research: passive and active stereo glasses as well as autostereoscopic displays were used in combination with multi-touch and gestural tracking technologies.

Beyond that, some special setups exist that separate the touch surface from the projection in such a way that the actual interaction takes place behind the projection screen, either by multi-touch [82] or mid-air [89] gestures.

In the mobile domain, some commercial products exist that are equipped with multi-touch autostereoscopic displays. The HTC Evo 3D and the LG Optimus 3D Max are examples of autostereoscopic smartphones that use the parallax barrier technique².

However, for the mobile domain few research exists regarding the stereoscopic perception on mobile devices. On small mobile screens touch accuracy is a critical issue also referred to as the fat finger problem. Colley et al. investigated this issue on autostereoscopic mobile devices [38]. They found that the touch target size tends to be bigger for stereoscopic than monoscopic mobile touch displays.

In this dissertation, interactive prototypes are developed and evaluated on a couple of different 3D displays technologies. We use passive and active stereo glasses as well as autostereoscopic displays in combination with multi-touch and gestural tracking technologies (see Figure 2.12).

2.2.2. Input Technology

Research on 3D input devices has a long tradition in VR. While early systems are characterized by heavily-instrumented users (e.g. wands, gloves, markers, etc.) and environments (e.g. caves, tracking systems etc.) recent developments require less instrumentation of

²<https://www.htcdev.com/devcenter/opensense-sdk/legacy-apis/stereoscopic-3d>

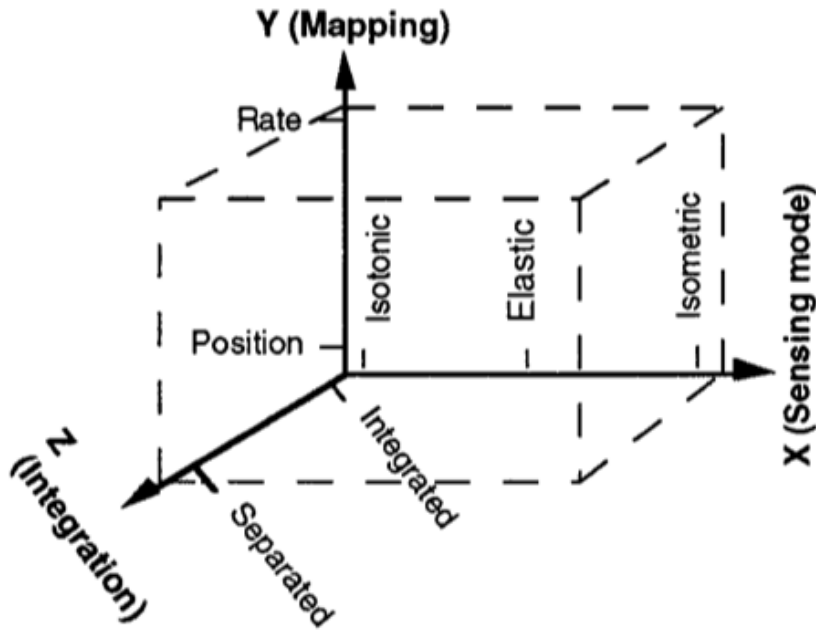


Figure 2.13.: 6-DOF Input Taxonomy by Zhai and Milgram [216]

users, or even no instrumentation at all. In the following an overview of 3D input technologies is given. Recent commodity hardware (multi-touch tabletops, mobile devices, depth cameras, etc.) is nowadays powerful and flexible enough and well suited as input devices for 3DUI. Thus, a thorough discussion of the multi-touch, mobile and other commodity input technology that is mainly addressed within the scope of this thesis, completes this section.

The 3D interaction predominantly requires input in six DOF. 3D tasks often require three DOF for translation and three DOF for rotation (e.g. to move and rotate an object in order to manipulate it, or move through the scene and look around during navigation). Zhai et al. [216] proposed a taxonomy of 6-DOF input, which consists of three dimensions of 6-DOF input devices (see Figure 2.13). The Mapping Relationship determines whether the input is rate-controlled or position-controlled. While position-controlled input depends on where the user directly maps to (e.g. tracking), rate-controlled input means that the user's input is related to the applied force (e.g. the speed of the cursor movement). The Sensing Mode is determined by the feedback of the input device ranging from isotonic to elastic to isometric. Isometric input devices provide resistance when moved, and center themselves back to their home position when released. Integration describes the degree to which all six DOF are controlled together. Integrated devices are controlled with a single 6-DOF device, or one or more of the six DOF are controlled separately by different devices (separated).

2. Perceptual and Technical Foundations of Interactive Stereoscopic Surfaces

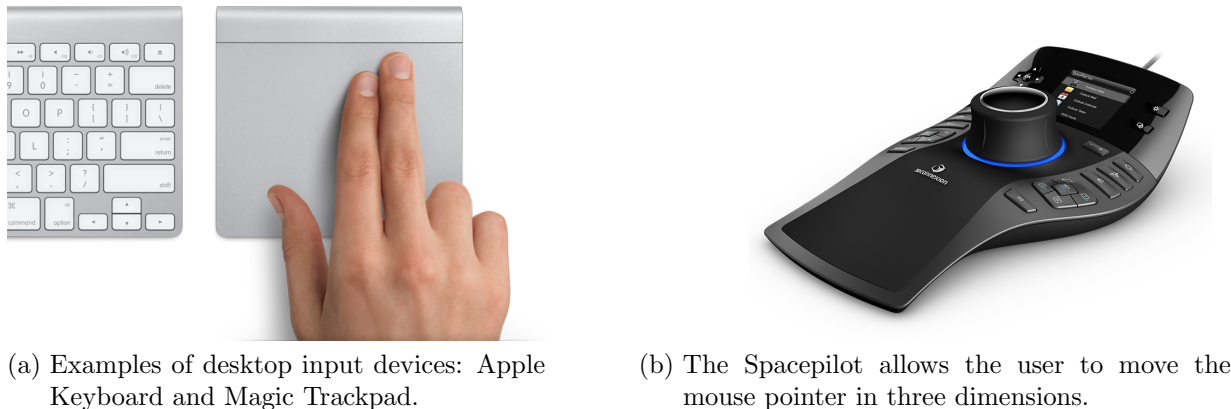


Figure 2.14.: Desktop input devices.

Overview of 3D Input Devices

Bowman et al. [21] divide input devices for 3DUI into the five following categories: (1) desktop input devices, (2) tracking devices, (3) 3D mice, (4) special-purpose input devices and (5) direct human input. Those devices can be described as follows and provide the basis for the overview of 3D input devices.

(1) Desktop input devices cover a wide range of devices, predominantly 2D input devices such as mouse, trackball, joystick, and keyboard (e.g. the Apple Keyboard and Magic Trackpad³ in Figure 2.14a). Most of these devices only provide 2-DOF input. Nevertheless, some 6-DOF desktop input devices exist as well. The Spacepilot⁴, for example, is a mouse that extends the degrees of freedom by adding a push and pull mechanism and thus makes it possible to move the mouse pointer in three dimensions (see Figure 2.14b). Tilting and spinning the cap further enables rotation and scaling for 3D object manipulation or navigation. However, 6-DOF desktop input devices are rarely used because they need a lot of training and even when they are operated by experts, it is hard to precisely control them [21].

(2) Tracking devices gather information about the 3D position and orientation of the human body, parts of the body or other physical objects. These can be used to provide tracked stereoscopic output, or for interaction. Changing the position or orientation of the body or object might change the virtual camera or virtual objects in the scene. There are different tracking approaches: motion tracking, eye trackers and data gloves. There exist a variety of commercial motion tracking systems, that are suited as input devices for VR (e.g. tracking solutions from Vicon⁵ or Optitrack by Naturalpoint⁶).

³<https://www.apple.com/magictackpad/>

⁴<http://www.3dconnexion.de/products/spacepilot-pro.html>

⁵<http://www.vicon.com/>

⁶<http://www.optitrack.com/>

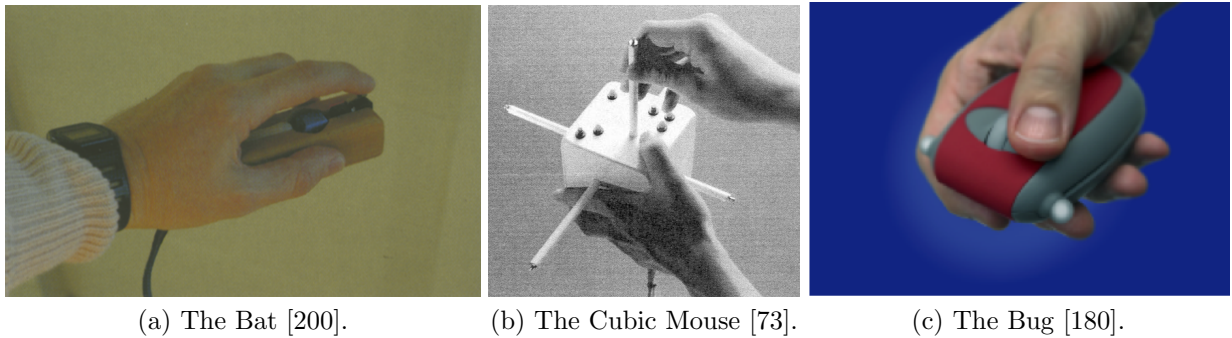


Figure 2.15.: Early 3D mice.

(3) 3D mice are tracking devices that are equipped with additional physical device components (e.g. buttons, sliders, knobs). Handheld 3D mice are mice- or joystick-like devices that are tracked in 3D space. Early examples of 3D mice are the Bat [201], the Cubic Mouse [73] or the Bug [180] (see Figure 2.15). 3D mice also exist as wearable devices, for example FingerSleeve [214].

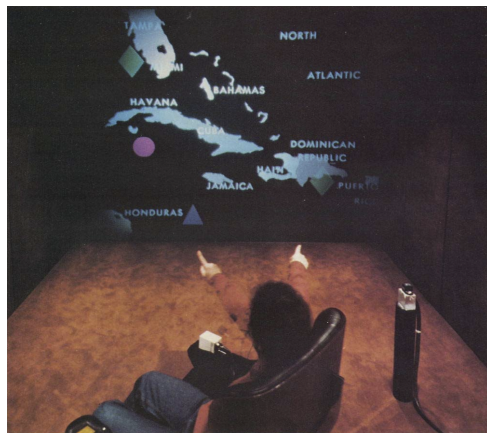
(4) Special-purpose input devices for 3DUI exist in various forms ranging from shape-changing tape to sensor-equipped shoes (see Bowman et al. [21] for more examples). According to Bowman, touch-enabled handheld and mounted tablet devices fall in this arbitrary class of input devices. This classification might have already changed due to the fact that most of these input devices are not intended for special purposes anymore, but rather general 3D input. Various interaction techniques that are proposed in this thesis are assumed to be general 3DUI input.

(5) Direct human input refers to interaction based on physical senses (i.e. modalities such as speech, muscle nerve, or brain signals) as another possible but uncommon way to interact with 3DUI. Motivated by put-that-there [14], voice input has been studied for 3DUI as well [122] (see Figure 2.16). However, as in other HCI-related domains, speech input has not become a popular research topic.

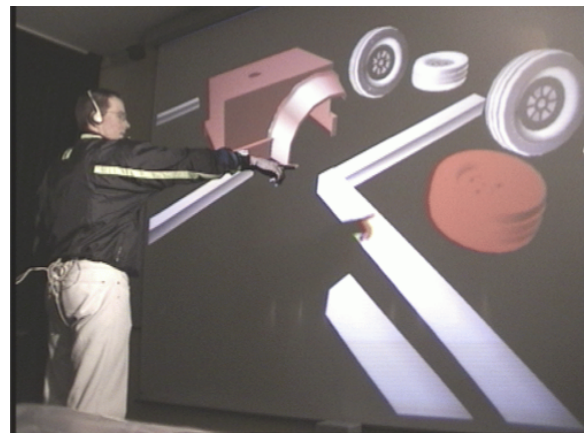
Commodity Hardware for 3D Input

Commodity hardware such as interactive tabletops, smartphones and depth sensors is affordable and already used as input devices in our everyday lives. Moreover, current 3D user interfaces, as for example provided by VR systems consist of stereoscopic projection and tracked input devices. But these are often expert systems with complex user interfaces and high instrumentation. For instance, travel in virtual environments is a universal interaction task and has been an intensive research topic. However, it is still a challenging task even in VR-based environments because six DOF need to be controlled to carry out this task.

2. Perceptual and Technical Foundations of Interactive Stereoscopic Surfaces



(a) Put-that-there [14].



(b) Speech and gestural input in VR [122].

Figure 2.16.: Speech input examples.

In order to address these challenges in VR research but also to introduce interactive 3D applications to the living room, such commodity devices can be used (cf. [169, 124]). Since commodity tracking hardware is of main interest as 3D input within the scope of this work, multi-touch, mobile and low-cost 3D tracking systems are briefly introduced here.

Multi-touch surfaces and interaction (especially on 2D user interfaces and interaction) have been a research topic in the last few decades (cf. Buxton’s history of multi-touch surfaces and interaction⁷). In the last years, inspired by Han’s work that relies on the principal of frustrated total internal reflection [84], several hardware solutions have been created that allow multi-touch input on surfaces of different sizes [165, 62, 164, 84, 105, 94, 102].

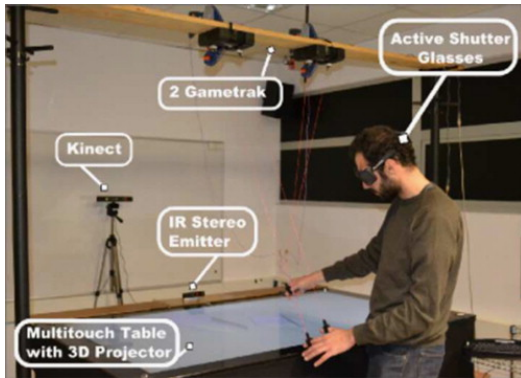
Little work has been done so far to combine multi-touch input and stereoscopic output. Grossman et al. [81] studied multi-touch input on volumetric displays.

The *Cubtile* by de la Rivière et al. [58] is a multi-touch device that enables 3D spatial interactions through its cubic shape of five touch-interactive surfaces. De la Rivière et al. [57] further introduced the *Iligh 3D Touch*, a multi-user multi-touch tabletop with shutter-based 3D.

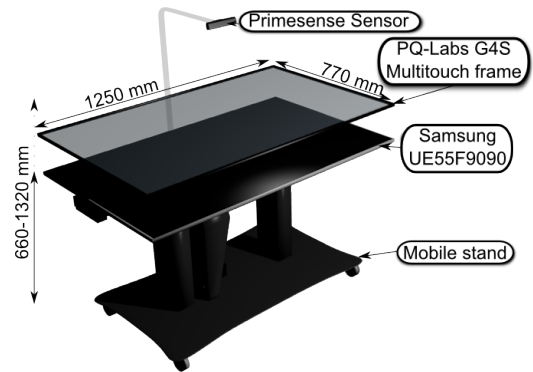
Coffey et al. [34] combined a horizontal multi-touch tabletop with a head-tracked vertical VR display. Coffey et al. [35] further investigated 3D interaction with volumetric data by applying a world-in-miniature (WIM) metaphor (i.e. using the multi-touch display as WIM in combination with a large stereoscopic projection).

A wall-sized stereoscopic multi-touch surface setup was built by Valkov et al. [197] in order to explore where people are actually touching a stereoscopic projection. The setup consists of a polarized projection for visualization, head-tracking, and multi-touch tracking based on diffuse illumination (DI) [173].

⁷<http://www.billbuxton.com/multitouchOverview.html>



(a) The Mockup Builder system [56].



(b) The interactive SPAtial SurfaCE (iSPACE) [132].

Figure 2.17.: Examples for multi-touch stereoscopic devices.

Strothoff et al. [185] equipped a traditional 3D workbench with multi-touch tracking (similar to [197]) and explored 3D manipulation techniques.

De Araujo et al. [56] built an extensive framework to track gestures on and above a multi-touch stereoscopic display. They discuss and actually use a lot of different sensors to track gestures on and above an stereoscopic tabletop display (see Figure 2.17a).

Lubos et al. [132] present the interactive SPAtial SurfaCE (iSPACE), a system that combines multi-touch and mid-air interaction with stereoscopically projected data. The system consists of a cost-efficient depth camera, a multi-touch frame and a large state-of-the-art 3D TV screen (see Figure 2.17b). This setup also allows the tracking of gestures on and above a multi-touch stereoscopic display.

To sum up, the few approaches for stereoscopic multi-touch technology are research prototypes. But most of them are light weight systems with low user instrumentation and based on commodity hardware components. Our studies on stereoscopic multi-touch also primarily rely on combinations of commodity hardware parts.

Handheld 3D Input Technologies are moving into focus because smartphones are now powerful enough to process complex tasks like graphics or image processing. Most recently, the release of handheld devices equipped with an autostereoscopic display fosters the development of stereoscopic 3D applications for mobile devices. Bringing stereoscopic 3D to mobile devices is assumed to allow a more realistic perception of AR and VR. Smartphones can be used as 3D input devices in different ways. In general, they can be used as a single device for 3D output and input or as a remote to operate a 3DUI. In either case, the sensors of the device can be used to generate input data [93].

Depending on the input modality, different DOF can be achieved. The combination of these modalities even allows the implementation of intuitive 6-DOF input methods. In a device-only setup, input that consists of direct touch and orientation sensors allows users

2. Perceptual and Technical Foundations of Interactive Stereoscopic Surfaces



Figure 2.18.: Examples of commodity 3D input devices.

to easily navigate by moving the device in space and manipulate it by direct touching of the objects.

In a VR scenario, the use of a physical mobile device serves as a passive haptic prop to support the user's spatial orientation and control (cf. passive real-world interface props by Hinckley et al. [90]). A passive haptic prop is a tangible object that serves as a surrogate for its virtual counterpart. The physical prop is used as an interface to manipulate the virtual object.

3D Tracking Technologies have evolved from large and expensive systems to affordable commodity tracking devices (see Figure 2.18) that are even appropriate for 3DUIs. Devices that rely on inertial and optical tracking proved to be the most promising approaches and are widely used for 3D gestural tracking in everyday devices such as game controllers or mobile phones.

The Nintendo Wiimote⁸ and Playstation Move⁹ are very early examples from the gaming domain that have also been used as spatial input devices for 3DUIs. These devices have a great potential for video gaming that is informed by VR research [124] but also for 3DUIs in general [169, 210]. Recent low-cost RGB-D sensors (e.g. Kinect for Windows¹⁰) have re-inspired a lot of 3D interaction research projects. Yet another class of lightweight near-field depth sensors has newly been created (e.g. Leapmotion¹¹, NUI DUO¹²) which also have the potential to inform new 3D interaction techniques.

As already mentioned above, mobile devices such as smartphones that have many in-built sensors (e.g. cameras and inertial sensors) are capable of 3D gesture tracking. Besides multi-touch gestures that are universally adopted, spatial gestures gained attention as well, and many applications use sensors to track such gestural input. On the other hand, it might

⁸<https://store.nintendo.com/ng3/browse/productDetailColorSizePicker.jsp?productId=prod150198>

⁹<http://de.playstation.com/psmove>

¹⁰<https://www.microsoft.com/en-us/kinectforwindows>

¹¹<https://www.leapmotion.com>

¹²<https://duo3d.com>

also be feasible to use classic 3D tracking systems (e.g. Optitrack) to gather 3D location with high precision, high refresh rate and low latency. However, the tracking quality of commodity devices has noticeably increased which makes it reasonable to use them on their own because the users are accustomed to intuitively interacting with them.

Having introduced all relevant input and output 3D technologies with a particular focus on commodity devices that require less or no user instrumentation, we now investigate 3DUIs and 3D interactions for this class of devices. In the next section, 3DUIs are presented with respect to their basic tasks, followed by a thorough discussion of related work on 3D interaction.

2.3. 3D User Interfaces and 3D Interaction

Although VR research has a long tradition, 3DUI research (i.e. the investigation of VR interfaces with HCI methods) is a fairly young field. Seminal work was done by Sutherland [186] who was far ahead of his time, and for many decades VR research was steadily growing but mainly with a technology-driven focus. Very few applications existed, and they were mainly expert systems that required a lot of training (e.g. flight simulators). Bowman et al. [21] ironically claim that a *Scientific American* article by Foley [71] that attracted the first public awareness of VR was illustrated by a 3D input device, namely a 3D glove. However, this did not mean that human aspects and HCI research in VR increased after this article. But with the progress of both VR technology and HCI research, the need for 3DUI research slowly evolved by the application of HCI methods in VR interfaces research.

In the following, 3DUIs are introduced. An overview on canonical 3D tasks and the related classic interaction techniques is given, and the experimental evaluation of these universal tasks is discussed. Then, related work on the combination of multi-touch and gestural interaction with stereoscopic output is discussed for the following areas: (1) gestural 3D interaction research with large displays (tabletops and walls), (2) hand postures and the *Reach to Grasp* interaction, (3) gaze-based interaction with stereoscopic data, and (4) handheld 3D interaction. Finally, the presented concepts and related work are set within the context of the current state of research of this thesis.

2.3.1. Canonical 3D Tasks

The canonical manipulation task consists of selection (object selection), positioning (object translation) and rotation (object rotation) [21]. Selection, as the most fundamental task, can be seen separately and is often investigated independently. Deformation (such as object scaling) is another manipulation task that is often excluded from the canonical manipulation tasks due to the simplicity of the task. The most relevant aspect that characterizes a 3DUI is the set of universal 3D tasks that are needed to control a 3D interface. These are selection, manipulation, travel and system control, which are briefly introduced here (for an extensive overview of virtual environment interaction techniques see also Mine [148]).

2. Perceptual and Technical Foundations of Interactive Stereoscopic Surfaces

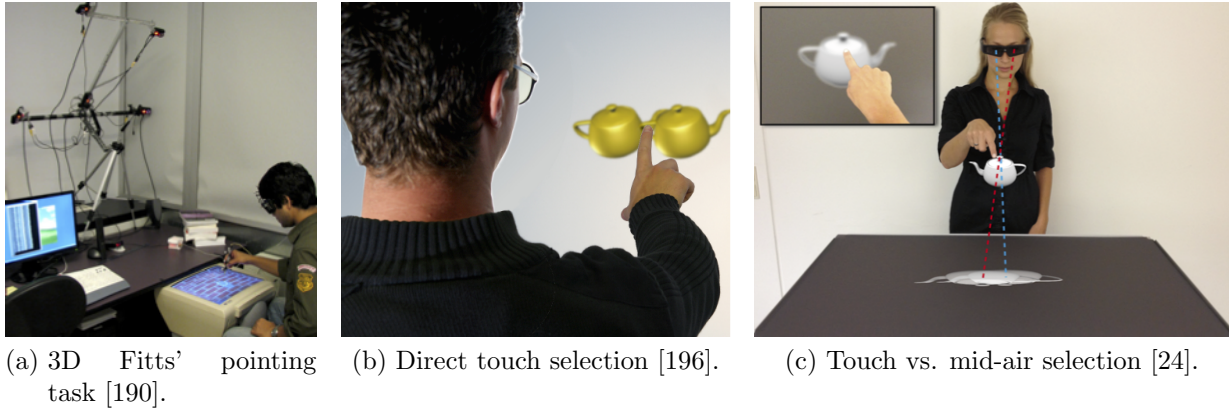


Figure 2.19.: 3D selection tasks.

In general, more DOF are needed for 3DUI tasks than for standard 2D graphical user interface (GUI) tasks. This requires more elaborated evaluation methods than most of the 2D tasks. For example for 2D pointing the ISO 9241-9 standard [99] exists that is based on Fitts' law [69]. Common experimental tasks exist only partially for the universal 3D tasks. In the following the universal 3D tasks and state-of-the-art experimental designs for those tasks are discussed.

Selection

Selection is the most fundamental universal interaction task. However, selection is already fairly complex in 3DUIs. In contrast to 2D, the objects are distributed in 3D space. Thus objects need to be selected that are projected in depth or that are (partially) occluded by other objects. Selection techniques can be categorized as local, at-a-distance, gaze directed, voice input and list selection [148].

According to Bowman [21], common selection techniques that use 3D input devices are the simple virtual hand, ray-casting [91], occlusion [159] and Go-Go technique [160]. Bowman and Hodges [18] compared Go-Go and ray-casting techniques in a qualitative study (Go-Go, Fast Go-Go, Go-Go stretch, indirect stretching Go-Go technique, ray casting and ray-casting with reeling). Besides the fact that none of the tested techniques were significantly preferred over the others, Bowman and Hodges found that grabbing (for selection) needs to be treated separately when evaluating manipulation techniques. The results of that study provide the basis of the differentiation of selection and manipulation in this thesis (see Chapter 4).

3D selection has been studied for stereoscopic display. For instance, Boritz and Booth [17] conducted a 3D pointing study that compares four different visual feedback modes (fixed viewpoint monoscopic perspective, fixed viewpoint stereoscopic perspective, head-tracked monoscopic perspective and head-tracked stereoscopic perspective). The participants had to select six targets that were located on the three axes, 10cm away from a fixed starting position. The results indicate that stereo outperforms the monoscopic

condition performance and that asymmetries exist both across and within axes. Notably, head tracking had no significant effect on performance.

Although this has been such an active field of research, a common methodology to study 3D selection has not been pursued by the VR community. To close this methodological gap, a few models including the application of Fitts' law have been proposed for 3D as well.

Zhai et al. [215] proposed the "Silk Cursor", a semi-transparent volumetric cursor for dynamic 3D target acquisition. They proposed a variation of Fitts' law and showed in an experiment that the volume occlusion is effective in both monocular and stereoscopic conditions.

Grossman and Balakrishnan [79] investigated pointing on volumetric 3D displays but they also studied how the physical movement angle affects the pointing performance. Results showed that target size dimension along the primary axis of movement has a greater impact on performance than the other two dimensions. As a result they proposed a model that describes pointing at trivariate targets.

Jota et al. [106] performed a comparison of ray pointing techniques for large displays. They showed that techniques based on "rotational control" perform better for targeting tasks, while techniques with low parallax are best suited for tracing tasks. They further showed that a Fitts' law analysis based on angles better approximates the ray pointing performance.

Teather and Stuerzlinger investigated 3D pointing in stereoscopic head-tracked 3D [190]. In a series of experiments they evaluated various pointing techniques in a 3D Fitts' law experiment and also compared them to 2D pointing techniques (see Figure 2.19a). They were able to show that the 2D version and their adaptation of Fitts' law holds for planar pointing tasks but badly predicts 3D motions using the ray- and mouse-based techniques.

However, in stereoscopic 3D environments selecting objects via touch is a crucial perceptual task. Valkov et al. [196] investigated the selection of stereoscopic objects that were projected with positive, negative or zero parallaxes. They studied how far a user perceives direct touching objects displayed with a different parallax as if they are at zero parallax. Their study further revealed that users tend to touch somewhere between the projections for the two eyes with an offset towards the projection for the dominant eye (see Figure 2.19b).

Bruder et al. [24] compared 2D touch and 3D mid-air selection in a Fitts' law experiment for objects that are projected with varying parallax (see Figure 2.19c). Their results show that the 2D touch performs better close to the screen, while 3D selection outperforms 2D touch for targets further away from the screen.

Manipulation

Manipulation connotes the modification of objects with respect to position, orientation and size. Therefore, the manipulation of objects is one of the most fundamental tasks in both physical and virtual environments. To control 3D position and orientation requires at least six DOF which can be controlled by appropriate 3D input devices (see above

2. Perceptual and Technical Foundations of Interactive Stereoscopic Surfaces

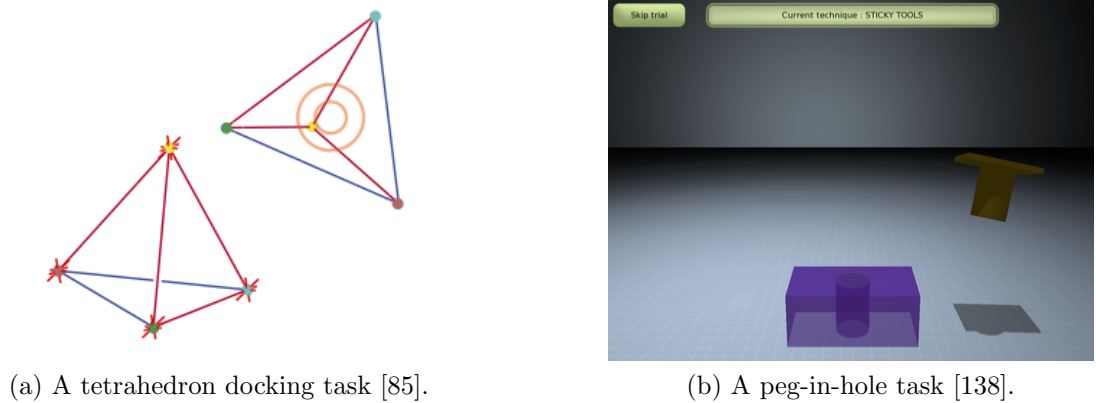


Figure 2.20.: Simple docking task examples.

in Chapter 2.2.2). The manipulation of size is often handled separately because of its simplicity [21].

As the term manipulation already implies by its etymological origin (latin *manus*: hand), the human hand is an ideal tool for direct manipulation. Humans are able to use their hands effortlessly together with other body parts and senses (e.g. 3D vision, audition and kinesthetic memory) and without consciously thinking about the underlying physical task [205]. Thus it seems natural to also use the hand as a 3D input device. By equipping the dominant hand (DH) with a 6-DOF sensor, its position can be directly mapped onto the position of a virtual object in an immersive world, creating the illusion that the user is able to move and rotate this object using her own hand. This allows a natural and intuitive interaction with virtual objects.

Common manipulation techniques are the simple virtual hand, hand-centered object manipulation extending ray-casting (HOMER) [18], scaled-world grab [149, 159] and WIM [184].

HOMER [18] is a hybrid manipulation technique that combines ray-casting with hand-centered manipulation. The user selects the object via ray-casting and manipulates it with the virtual hand metaphor.

In the scaled-world grab technique [149, 159] the whole virtual world is scaled in accordance to the object the user manipulates. The user selects an object with an image-plane selection technique. After the selection, the whole scene's zoom level is adapted with respect to this object and it can then be manipulated.

In the WIM [184] the user does not manipulate the actual 3D scene but rather a placeholder object, i.e. a miniature handheld model that is an exact copy of the scene.

Due to its naturalness and efficiency, hand manipulation is the dominant means of interaction in virtual worlds. Thus, research on other 3D manipulation modalities such as speech, gaze and whole-body interaction relatively little has been done.

3D manipulation can be experimentally evaluated in a docking task. In a docking task an object is manipulated (moved, rotated and scaled) until it “docks” into the corresponding

target object (see Figure 2.20). Several docking task studies have been conducted to test the object manipulation capabilities of various input devices (e.g. [16, 30, 119, 139, 217]). The docking task is a well-established method to evaluate input devices and interaction techniques in 3DUI [21]. However, there is no standardized docking task that is used in the literature.

An important prerequisite for the docking task is to define a metric in order to measure efficiency in terms of speed and accuracy. 3D rotations are the most crucial problem when defining such a metric. Zhai and Milgram [217] quantified the coordination in multiple DOF movement and evaluated this metric with 6-DOF input devices in a docking task. Masliah and Milgram [139] proposed the m-metric as a measure for the allocation of control in a 6-DOF docking experiment. More recent work suggested avoiding Euler angles in the definition of metrics for complex rotations (for details on metrics for 3D rotations see for example [98, 117]). In contrast to the related work, we have chosen a reasonably complex docking task to achieve a more realistic scenario and use a quaternion based metric as recommended by Huynh [98] (see Chapter 5).

Travel

Travel is a universal task and is a prerequisite for navigation. Navigation or way finding refers to the process of determining a path through an environment to reach a goal [20]. Travel tasks are one of the most fundamental human tasks in our physical environment as well as universal interaction tasks in 3DUIs. Travel plays an important role in virtual environments in general, like navigation in the world wide web, through many layers in a spreadsheet or in a virtual world of a computer game.

A travel task is defined as a task of performing the action that moves the viewport from a current to a target location [21]. Once a goal is formulated, muscles are triggered by the brain to perform the correct movements in order to achieve the goal. Turning a wheel, pressing a pedal or flipping a switch are examples for interfaces mapping various physical movements. In contrast to the real world, simple physical motions in virtual environments are only effective in limited space and at limited speed, which results in more or less natural mappings of the actions.

Travel is often only seen as a secondary task or as supporting other tasks, like picking up treasures or fighting enemies. Thus, if the user needs to focus on travel, this might result in a distraction from the primary task. Therefore travel techniques must be unobtrusive, intuitive and easily controllable. While developing the interaction metaphors, we attempted to keep each metaphor as simple and natural as possible, although we also wanted to study the differences, advantages and drawbacks of each form.

Besides other classifications, travel in a 3DUI can be categorized as active versus passive and physical versus virtual [21]. Active travel techniques enable the user to directly control viewport movement and orientation. In passive travel, the viewport is controlled by the system. Physical travel requires the user to perform motions that are tracked in the real world and projected in the virtual space. Physical travel is only effective in a very limited space. In virtual travel the user moves in the virtual space by steering a virtual vehicle.

2. Perceptual and Technical Foundations of Interactive Stereoscopic Surfaces

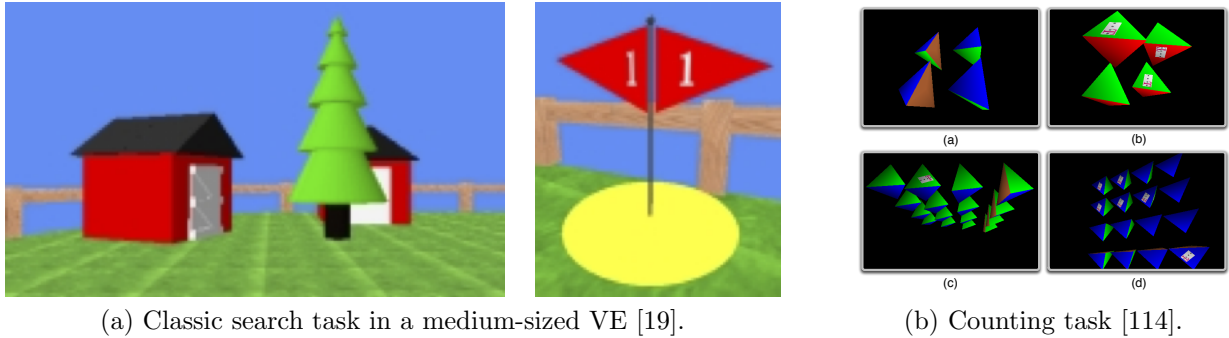


Figure 2.21.: Travel task examples.

The user's body remains stationary. Virtual vehicles provide only visual motion cues, not vestibular cues.

From a technological perspective, travel can be thought of as the control of the virtual camera in 3D environments. Since the early years of 3D graphics, camera movement has been widely studied. The control of the virtual camera in 3D environments requires at least six DOF, which can be directly controlled by means of 6-DOF input devices using established metaphors, like the scene-in-hand, eyeball-in-hand and flying vehicle metaphors proposed by Ware and Osborne [202]. In a qualitatively oriented study, users were requested to perform three navigation tasks in three different scenes. Ware and Osborne concluded which of the different metaphors is best suited depends on the particular task.

One of the classic 3D travel techniques is the Grabbing-the-Air technique [136]. In this technique, the entire world is viewed as an object to be manipulated, while the viewpoint remains stationary. To initiate the travel interaction, the user performs a grabbing gesture. Then she moves her hand to move the entire world. In the Camera-in-Hand technique the tracker is imagined to be a virtual camera looking at the world or the scene [21].

Bowman et al. [20] investigated travel techniques for first-person motion control and proposed an evaluation framework for the quality of different techniques with respect to specific virtual environment tasks. Results from their quantitative user studies show that pointing techniques are advantageous relative to gaze-directed steering techniques for a relative motion task, and that motion techniques which instantly teleport users to new locations are correlated with increased user disorientation.

Physical motion techniques have also been studied. One research direction is locomotion, which studies physical walking to travel in virtual environments (cf. Cohn et al. [37] for an overview and comparison of virtual environment locomotion interfaces). Also of interest for this work are “lean-based” techniques (e.g. [66, 195]). While Fairchild et al. [66] proposed a travel technique that was specifically designed for novice users, Valkov et al. [195] explored the virtual transporter metaphor.

3D travel has been widely studied in 3DUI. Travel is often evaluated in a search task (see Figure 2.21a) and can be separated into two main phases: exploration and search [21]. In

the exploration phase, travel is performed without an explicit goal. It is just for exploring the environment and obtaining local knowledge about the (virtual) environment. Thus, the exploration phase is best suited to give the user orientation in a travel task while getting familiar with a certain travel technique and input device. In the search phase, travel is required to move to a target location in the virtual environment, i.e. to achieve a specific goal. A previously performed exploration task might help the user to orient herself and support an effective search. Two types of search tasks exist: naive and primed. In the naive search, the user has no prior knowledge about the environment she is about to navigate through. The search task is denoted as primed search if the user has already gathered local knowledge and is thus aware of her location. Travel is often treated as only a secondary or supporting task. This requires travel to be intuitive and unobtrusive so as not to distract the user from the primary task. Travel techniques need to be carefully designed and evaluated with respect to these requirements.

In contrast to the common search task, a slightly different approach to evaluate virtual camera control on a mobile 3D display has been proposed by Decle and Hachet as well as Kratz and Rohs. Decle and Hachet [59] investigated 3D camera manipulation on touch-based mobile phones. Their study on direct versus planned navigation used a complex block structure as a VE. The task was to count markers within this 3D structure. Although the participants had a better completion time with the direct control, they preferred using the planned version of the trackball because it limited disorientation.

Kratz and Rohs [114] extended the virtual trackball metaphor to rear touch input and evaluated their input techniques in a 3D rotation task (see Figure). This rotation task was inspired by the travel task of Decle and Hachet and has the advantage that the task is highly parametric and thus provides an easy way to control the study design. We evaluated our interaction techniques in Chapter 6 in a travel task with an extended version of Kratz and Rohs' [114] approach.

System Control

System control is another crucial but often ignored task for 3DUI. As in almost every application, 3DUIs require menus and other tools to control the application (e.g. print) or toggle mode-switches. As in WIMP-based UIs, these controls are often arranged on a (floating) plane in 2D, but some are projected in 3D or shallow-depth 3D in the scene. The ring menu [175], the TULIP menu [22] and 3D widgets [39] are examples for more elaborated 3D menus. Marking menus are a seminal concept for efficient menus and a variant of pie or radial menus [120]. Grosjean and Coquillart [78] extended the marking menu by a third dimension. They developed a 3D widget called *Command and Control Cube*, a cube-shaped menu with $3 \times 3 \times 3$ items, omitting the center, which is reserved for the cancel operation. The menu was designed for a 3D workbench and controlled by a “pinch glove”. Selecting a menu item is performed similarly to the marking menu by drawing a line but in 3D space. Nevertheless, little research has been conducted in the field of 3D menus (c.f. Dachsel and Hübner [41] for a survey and taxonomy of 3D menus).

Conclusion

In the studies and experiments that are presented in this work, the above-mentioned approaches and guidelines were taken into consideration wherever possible. Of course, the literature on user interfaces in VR/AR research has been studied and discussed within the scope of this work. Further, HCI research on 2D user interfaces has also been taken into account when applicable to 3DUI.

We performed different user studies and we predominantly used participants without 3D experience in our studies. Still we could not completely remove the prior knowledge of 3D games and movies they are exposed to in their daily lives. We also have taken 3D games into consideration, as they explore the space of 3DUI very well and can serve as a potential testbed for the evaluation of universal 3D interactions. The design of basic 3D tasks that are very constrained to a few variables reflects this. Lots of qualitative data was gathered in order to close the gap between constrained experiments and more open-minded subjective feedback that might inform applications that make use of the studied 3DUIs and 3D interactions.

2.3.2. 3D Interaction

The combination of multi-touch and gestural interaction with stereoscopic output as a new paradigm for 3DUI is a relatively novel research field and little work has been done in this direction. In the following a variety of related work is presented that has influenced this field of research.

Touch and Gestural 3D Interaction

Since the beginning of mankind, humans have used gestures. Gestures are non-prehensile skilled hand movements generally accompanying (gesticulation) or substituting for speech (sign language) [104]. The expressiveness of human gestures makes them attractive for HCI although, from a technical perspective, gestures are challenging to recognize and process reliably as an input modality. Thus gestural commands as input have a long research tradition in HCI and 3DUI. Gestural interfaces in particular are expected to be good candidates for post-WIMP interfaces [92].

Seminal work on gestural interaction as an input modality was conducted in the early 1980s by HCI and VR pioneers such as William Buxton, Myron Krueger and others. Buxton worked on hand gesture input [26] and bimanual interaction [27]. Krueger et al. [115] introduced Videoplace, a wall-sized display that uses video cameras and image processing to track gestural input.

In the research domain of multi-touch interaction, much work has been carried out on the definition of frameworks and taxonomies for such gesture-based multi-touch input. Wu et al. [213] defined the principle of “Gesture Registration, Relaxation and Reuse”. Wobbrock et al. [211] investigated user defined gestures and developed a taxonomy of gestures for surface computing.

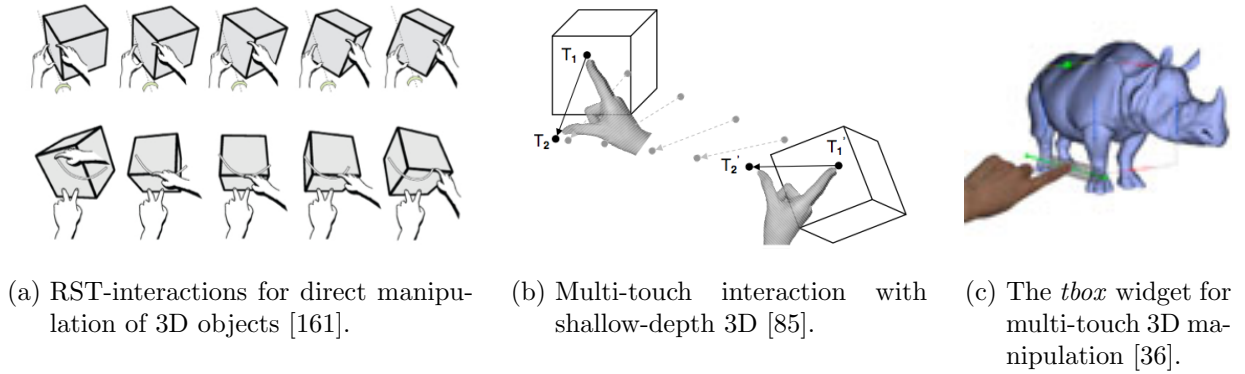


Figure 2.22.: 6-DOF multi-touch interaction examples.

But only a few researchers have addressed the problem of 3D interaction on a 2D multi-touch surface so far. Grossman and Wigdor [80] explored 3D interaction on interactive tabletops. From a survey on 3D tabletop systems they derived a taxonomy and provide design guidelines.

Wilson et al. [209] suggested including physics in surface interaction and on that basis Wilson [208] simulated grasping behavior on interactive surfaces.

Some distinct interaction methods have been proposed that allow 6-DOF multi-touch interaction with 3D data. These techniques either allow direct input that differentiates between fingers, or use a widget to enable 6-DOF multi-touch input.

Reisman et al. [161] defined interactions for direct manipulation of 3D objects through rotation, scale and translation (RST) (see Figure 2.22a). Inspired by 2D RST interactions, they proposed interaction techniques that allow the user to directly manipulate 3D objects with three or more (touch) points.

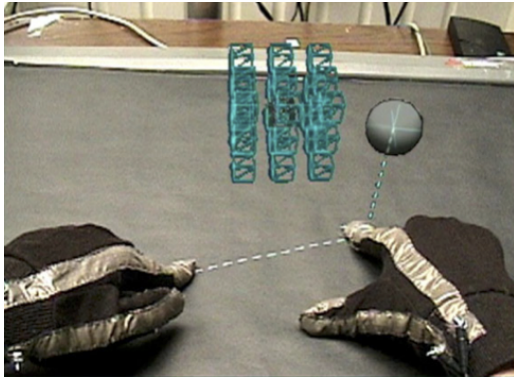
Hancock et al. [85] presented the concept of shallow-depth 3D (i.e. 3D interaction with limited depth) in order to extend interaction with digital 2D surfaces (see Figure 2.22b). They [86] further introduced force-based 3D interactions for multi-touch tabletops.

Hachet et al. [83] presented a widget-based approach for 3D navigation through 2D input for different (touch-enabled) devices. The technique was motivated by the Point-of-Interest (POI) technique introduced by Mackinlay [135]. But in contrast to the classic POI technique, the whole trajectory is interactively defined by the user, instead of just the endpoint of a trajectory.

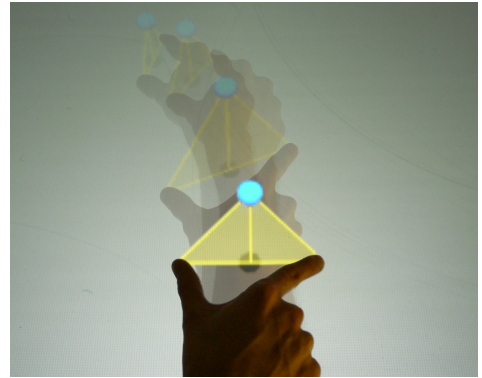
Cohé et al. [36] used widgets to provide 3D manipulation on touch screens. They proposed *tBox*, a cubic widget that enables 3D object transformations through multi-touch input (see Figure 2.22c).

A limitation of all these approaches is the constraint of the interaction and visualization to almost zero parallax because the plane of the interactive surface limits the interaction space more or less to the 2D surface [174].

2. Perceptual and Technical Foundations of Interactive Stereoscopic Surfaces



(a) Balloon selection [6].



(b) The triangle cursor [185].

Figure 2.23.: Multi-touch 3D selection.

Hilliges et al. [88] tried to overcome the restriction of the interaction and visualization to the 2D surface by adding depth tracking as input to interactive tabletops. Through the tracking of gestures on and above the surface, interactions in the air were made possible.

Zilch et al. [218] investigated 2D and 3D GUI widgets for stereoscopic multitouch displays. They classified 2D widgets and derived a set of 3D GUI widgets with strong mental models of real-world interactions. Their studies revealed differences in touch behavior with and without stereoscopic displayed 3D widgets.

Due to their natural and non-conflicting depth cues, direct touch interaction with monoscopic 3D objects can be assigned to the image plane selection techniques [159]. In a stereoscopic multi-touch environment they are practically the same, but conceptually similar to ray-casting methods, with a ray emitted into the negative and positive parallax space [148].

With today's technology it is now possible to apply the basic advantages of bimanual interaction [27, 92] to 3D interaction. Mine [149] proposed the two-handed flying technique, a pointing selection method where direction and speed of the pointer is specified by the vector between the user's hands.

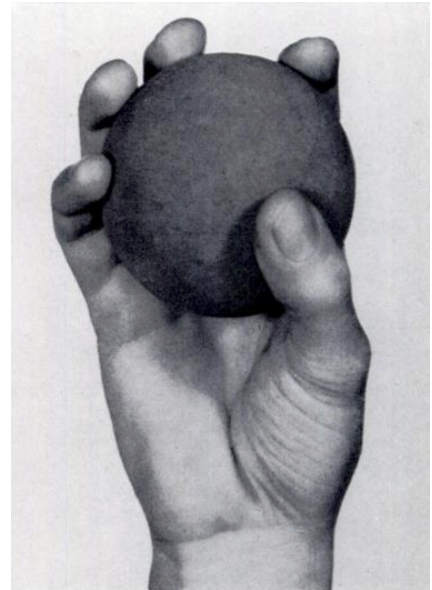
Slice WIM [35] is a multi-touch application for the exploration of volume datasets. The classic WIM metaphor is used in a VE that combines a multi-touch tabletop (as WIM) and a wall-size display.

The *Balloon Selection* approach by Benko and Feiner [6] is a multi-touch technique in an augmented reality setting that allows indirect selection in the 3D space above the tabletop (see Figure 2.23a). The selection pointer is manipulated via touch by multiple fingers with a balloon metaphor, i.e. the position and height of a helium balloon is manipulated with a string.

Strothoff et al. [185] presented *TriangleCursor*, an interaction technique similar to *Balloon Selection* (see Figure 2.23b), and compared it to an extended version of Benko and Feiner's approach in a manipulation task. Both techniques, the *Balloon Selection* and the *TriangleCursor*, are restricted to negative parallax (see Figure 2.23).



(a) Power grip.



(b) Precision grip.

Figure 2.24.: Prehensile movements of the human hand [151].

Although many 3D interaction techniques have been proposed and studied, very few efficient intuitive interactions exist. This might be due to the fact that an in-depth understanding of how users can be enabled to naturally interact with high DOF is still missing. Thus there is a need for further investigations on how to interact with complex 3D data in particular stereoscopic rendered data. We therefore investigate interaction with stereoscopic 3D data using gestural input, in particular with respect to the parallax spaces introduced above. Following this direction, we investigate indirect selection techniques that allow seamless selection and manipulation of stereoscopic objects displayed with different parallax on a multi-touch display without high instrumentation (see Chapter 4 and 5).

Hand Postures and Grasp Gestures

Multi-touch technology can be used in order to allow a rich set of interactions without any instrumentation of the user, but the interaction is often limited to near-zero parallax [174]. Although the combination of multi-touch technology, depth cameras and stereoscopic displays promises interesting and novel user interfaces, the benefits, possibilities and limitations of using this combination have not been examined in depth and are so far not well understood [182]. Psychological research on the *Reach to Grasp* task has shown that the pre-shaping phase of the human hand allows a prediction of the object a human is going to grab. Multiple studies were conducted in this direction including physical objects as well as memorized and virtual objects that had to be reached and grasped [32, 142, 167]. Research in this direction has shown evidence that not all DOF of the grasping hand, but rather only a few, have an impact on that prediction [142, 166, 168, 191].

2. Perceptual and Technical Foundations of Interactive Stereoscopic Surfaces

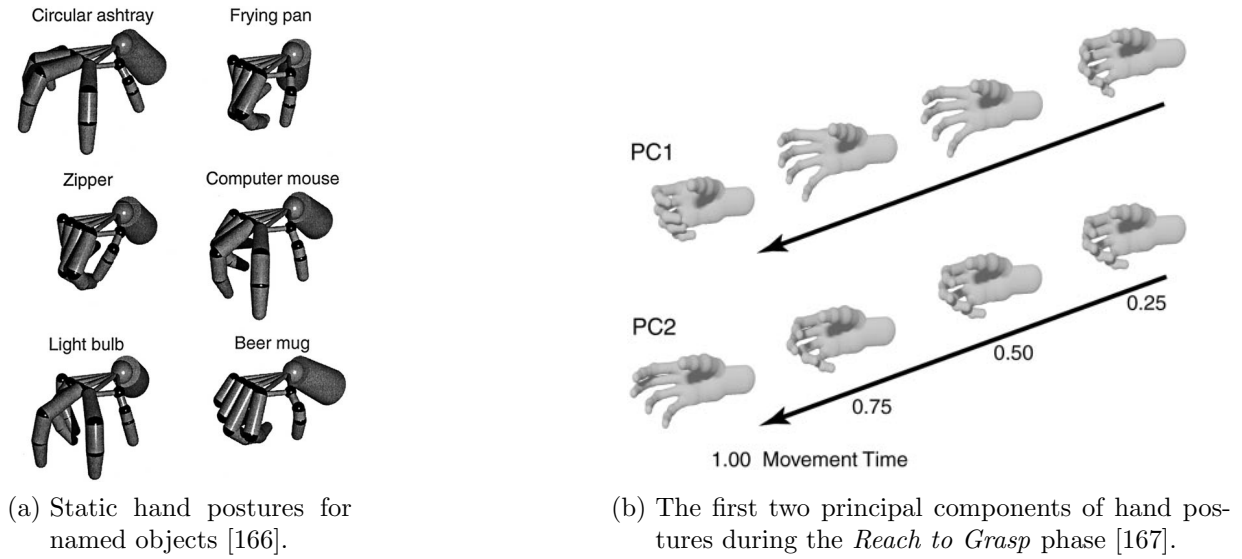


Figure 2.25.: Hand postures in the *Reach to Grasp* phase.

The *Reach to Grasp* task has been investigated by many neuropsychological and robotic research groups over the last decades, as the following body of research shows. In his seminal work Napier [151] distinguished between two basic patterns of grasp movements that he termed precision grip and power grip (see Figure 2.24).

Chieffi and Gentilucci [32] investigated the influence of the transport on the grasp component of prehension movements. One important finding of their work is that early in the transport phase, the grasp aperture remains fixed relative to the size of the object.

Various studies have been performed on motor coordination while reaching for and grasping objects. By using principal component analysis Santello et al. [168] concluded that all but two degrees of freedom of the hand are controlled as a unit. In their follow-up work, Santello et al. [167] also showed that the visibility of an object during the transport phase had no influence on the kinematics and the shape of the hand (see Figure 2.25). Further experiments focused on the question of when the shape of the hand (dependent on object shape as well as on object size) stops changing in the transport phase before contact.

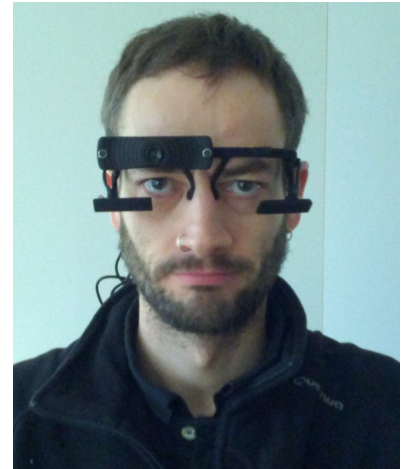
Thakur et al. [191] came to similar conclusions in a more general study where multidigit movement synergies in an unconstrained haptic exploration task were investigated.

Following this direction, Maycock et al. [142, 143] also investigated grasping of virtual and real objects and discuss applications in robotics. In robotics, the main aspect lies in grasp formation with respect to the object.

These insights from neuropsychological and robotics research are promising and we believe that the information for the *Reach to Grasp* phase can substantially improve interaction with stereoscopic multi-touch displays. In contrast to the research presented above, our approach investigates the grasp posture for object prediction that is used as interaction



(a) Early example of gaze-based interaction in VR (left) [188].



(b) A recent mobile bifocal eye-tracker.

Figure 2.26.: Gaze-based interaction.

context (see Chapter 7). Thus, our approach can be seen as an application of the findings of *Reach to Grasp* research to 3DUIs.

While reaching and pointing tasks have a long tradition in the HCI field, the hand pre-shaping has rarely been investigated. However, due to the availability of low-cost algorithms, simple off-the-shelf hardware and low instrumentation are now sufficient to track the human hand above the interactive surface. Depth cameras provide the possibility to recognize hand gestures and postures. Furthermore, when tracking the users grasp postures above a multi-touch display, her intended interactions might be predicted before she actually touches the surface. Such knowledge has the potential to improve the user interface of stereoscopic multi-touch surfaces, for example, by snapping desired objects to the touch surface. Motivated by these research challenges, we investigated grasp as interaction context to enable multi-touch interaction with stereoscopic data projected with differing parallax (see Chapter 8).

Gaze-Based Interaction

Eye gaze interaction is very natural because humans habitually use their eyes for communication with each other. Eye movements are very fast and require little physical effort [177]. With the development of inexpensive, unobtrusive desktop and wearable eye tracking solutions, this technology has become very popular in HCI.

Eye Tracking technology can be stationary (remote) or mobile (head-mounted) [64]. Current research on gaze-interactive interfaces mostly rely on stationary eye trackers. However, advances in mobile eye-tracking equipment have potential as a pervasive interface in everyday life [25]. With improvements in bifocal eye-tracking, gaze-based interaction has also made some progress in 3DUI research (see Figure 2.26).

2. Perceptual and Technical Foundations of Interactive Stereoscopic Surfaces

Eyes are good at (quickly) processing and perceiving visual information, but eye gaze is not necessarily well suited as an explicit input modality, e.g. eye-blinking to activate buttons, etc. [9]. Nevertheless, gaze-based interactions could serve as an additional input mode when the user needs both hands for other tasks [176].

Kumar and Winograd [118] investigated gaze-input based scrolling techniques including a technique for map panning with gaze.

Gaze can be used to track the user's visual behavior. In particular, it can be determined where the user is looking. One crucial aspect of gaze interaction is the so called *Midas touch* that stands for accidental interaction with everything the user is looking at [103].

To avoid the *Midas touch* effect, additional modalities are often used together with gaze input. Stellmach et al. [183] for example combined gaze and touch for target acquisition. They formulated the design principle that “gaze suggests, touch confirms”.

Gaze-based interaction can be also extended to multi-display environments. Turner et al. [193, 194] studied various approaches for gaze-based cross-device content transfer.

Other application examples for eye tracking are video game control [177], 3D interaction in VE [188] and text processing [9]. Holman [95] investigated gaze-based interaction techniques for tabletops, mainly focusing on co-located collaboration tasks.

Having these issues of gaze-based interaction in mind, we examined gaze-based interaction by mainly treating gaze as an additional input mode or interaction context that supports other interaction modalities such as multi-touch gestures (see Section 3.2 and Chapter 7).

Handheld 3D Interaction

Recent smartphones are equipped with various sensors (e.g. camera, accelerometer, gyrometer, GPS, etc.) and can be defined as “ubiquitous input devices” [4]. A lot of research has been done in the field of sensor-based mobile interaction (cf. Hinckley et al. [93]).

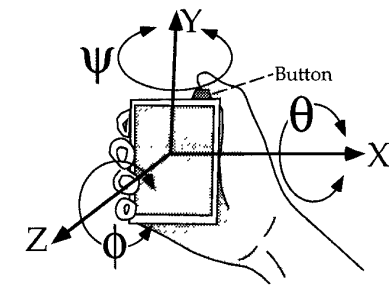
Boring et al. [15] introduced three interaction concepts to remotely control a pointer via scroll, tilt and move gestures with a mobile phone. Their approach enables an intuitive and easy-to-control remote pointing mechanism for distant displays.

However, only a few researchers have addressed the problem of 3D interaction on mobile devices so far. Steinicke et al. [182] discussed possibilities and limitations for using multi-touch interfaces with mobile multi-touch enabled devices to interact with stereoscopic content.

In an early work, Rekimoto [162] proposed a handheld device with one button and orientation sensors, where 3D interaction is mentioned as one application for such a device (see Figure 2.27a).

Capin et al. [28] presented a camera-based approach to navigate VEs on mobile devices (see Figure 2.27a). They used a feature-based tracking algorithm to track movements of the device, and based on this mapped the physical motion of the physical device to motion of the virtual camera.

Benzina et al. [8] investigated phone-based motion control for travel in VR. In their approach, the virtual camera was manipulated through touch input for translation, and



(a) Rekimoto's smallscreen device concept [162].



(b) camera-based navigation in VEs on a mobile phone [28]

Figure 2.27.: Handheld 3D interaction.

different device rotation strategies were investigated as well. Recently, they extended their approach in a thorough study by focusing on DOF reduction and the mapping function (i.e. the mapping between user action and mobile device) [7].

Liang et al. [127] investigated how mobile devices could be used as input for distant large 3D displays. In an exploratory study, they asked participants to propose interactions for 3D tasks and they applied their findings to a prototypical application for 3D object manipulation.

As already discussed above, Decle and Hachet [59] studied direct versus planned 3D camera manipulation on touch-based mobile phones, and Kratz and Rohs [114] extended the virtual trackball metaphor to rear touch input.

Following the above directions, we investigate remote gestural and mobile manipulation and navigation techniques for large stereoscopic displays (see Chapters 5 and 6). We further extend the interaction space by investigating multi-touch and sensor-based interaction with 3D stereoscopic data on a mobile device (see Chapter 9).

2.4. Conclusion

The interaction space that is investigated within this thesis ranges from touches and gestures on to mid-air gestures and postures above the interactive surface. Gestures can be applied to 3D interaction with different levels of directness of the interaction. The interaction with 3D data can either be direct, indirect or remote.

Direct manipulation means that the user is directly manipulating the object by (literally) touching it. Like touching an object in the real world the user is partially occluding it with her finger(s) in order to move it, rotate it, etc. Direct touch manipulation is assumed to be very natural due to the directness of the interaction. The concepts of direct manipulation can also be applied to stereoscopic 3DUIs especially when objects are displayed with shallow parallax (i.e. close to zero parallax, e.g. [85, 161]). However, direct manipulation might lead to perceptual problems (e.g. accommodation-convergence conflict) when objects that are displayed at extreme parallax are manipulated (see also Section 2.1 for a more extensive

2. Perceptual and Technical Foundations of Interactive Stereoscopic Surfaces

discussion of perceptual issues). We further explore direct manipulation by investigating multi-touch and sensor-based interaction with 3D stereoscopic data that is displayed on a mobile device (see Chapter 9). An approach that extends the metaphor of direct interaction and avoids perceptual confusion is the concept of interaction context and the *Reach to Grasp* interaction. As discussed later, in this approach the user interface is adapted with respect to interaction context in order to allow direct manipulation on the interactive surface (see Chapter 7).

A different approach to object manipulation is indirect manipulation. Objects are manipulated indirectly by using, for example, widgets. This approach is often used in desktop 3DUIs to cope with low-DOF input devices. Indirect manipulation has already been used for multi-touch 3D interaction (e.g. *Balloon Selection* [6, 45, 185] and *z-technique* [137]), and widget-based approaches also exist (e.g. [36, 83]). We explore this approach by investigating indirect selection techniques that allow a seamless selection of stereoscopic objects displayed with differing parallax on a multi-touch display (see Chapter 4).

Indirect interaction can be seen as a variation of the remote manipulation approach. Remote interaction as one class of traditional VR interaction techniques is also of interest for multi-touch 3D interaction. For mobile interaction with 3D displays in particular, we investigate remote gestural and mobile manipulation and navigation techniques for large stereoscopic displays (see Chapters 5 and 6).

To sum up, this chapter gives an overview of technologies and concepts that are of interest for this thesis. First, stereoscopic vision and depth perception was discussed, with a special focus on depth perception on touch-enabled stereoscopic displays and handheld stereoscopic AR. Depth perception is a critical task when designing interactions for stereoscopic displays, and further research is needed regarding these issues. Second, 3D input and output technologies were discussed that not only allow an intuitive touch and gestural interaction with stereoscopic data, but also avoid heavy user instrumentation. Stereoscopic displays that rely on glasses to perceive 3D tend to be best suited for multi-touch workbench setups, even though they require (little) user instrumentation. Autostereoscopic display technology tends to be the best solution for mobile display in particular. While few autostereoscopic solutions exist for large displays, affordable mobile devices exist that are equipped with autostereoscopic display. Third, 3DUIs, including the universal 3D tasks and related 3D interaction, were introduced. The main conclusion that can be drawn is that the 3DUI community is still discussing how to design effective user evaluations. This is carefully taken into account within the studies conducted for this thesis. 3D interaction was considered with a focus on the combination of multi-touch and gestural interaction with stereoscopic output as a new paradigm for 3DUI. A variety of related work is presented that has influenced this field of research, including gestural 3D interaction with large displays, mobile 3D touch interaction and gaze-based interaction with interactive surfaces.

The related work clearly shows that there is a need for further investigations on how to interact with complex spatial, three-dimensional data, in particular stereoscopic rendered data. In this thesis we therefore investigate interaction with 3D data using gestural input on and above interactive surfaces. The focus of this work lies on gestural interaction techniques allowing users to interact in a direct manner without further instrumentation.

3. Classification of Multi-Modal 3D Interaction on Interactive Surfaces

In multi-touch research much work has been carried out on the definition of frameworks and taxonomies for gesture-based multi-touch input (see Chapter 2). While general taxonomies exist for 3D input, little research on the classification of 3D interaction with 2D surfaces via touch input or mid-air gestures exists (see Chapter 2). In this chapter a taxonomy for multi-modal 3D interaction on and above interactive surfaces is proposed. First, a framework for multi-touch interaction is developed. The multi-touch framework consists of three main parts and defines the commands and controls that are needed to manipulate the interaction space. Starting with an initial study that investigated the relationship between those multi-touch gestures and geospatial operations, the framework is incrementally extended by focusing on additional modalities. As a result, the taxonomy for multi-modal 3D interaction on and above interactive surfaces is presented and discussed. The taxonomy aims to inform the design of the interaction techniques of this work. Thus, in the remainder of this work the taxonomy is used to specify the developed interaction techniques. The contributions of this chapter have been partially published in [52, 53, 50, 171, 170].

3.1. Multi-Touch Interaction with Spatial 3D Data

In this section a conceptual framework for multi-touch interaction with spatial 3D data is presented. The framework is based on three key components: physical interactions, interaction primitives (IPs) and interaction space. Altogether, these three parts define the commands and controls that are needed to manipulate the interaction space (at different scales). As shown in the following, the interactions that are needed to navigate in and manipulate spatial data such as geographic data are clearly specified by these components.

Geographic data is usually manipulated in a geographic information system (GIS). A GIS is an information system for the mapping, retrieval, storage and analysis of geo-referenced data [33]. Novel UI paradigms such as multi-touch have a great potential for natural GIS interaction (see Figure 3.1), especially for map interactions, since they can be grounded on interaction with physical maps and thus enable strong metaphors [172].

3. Classification of Multi-Modal 3D Interaction on Interactive Surfaces

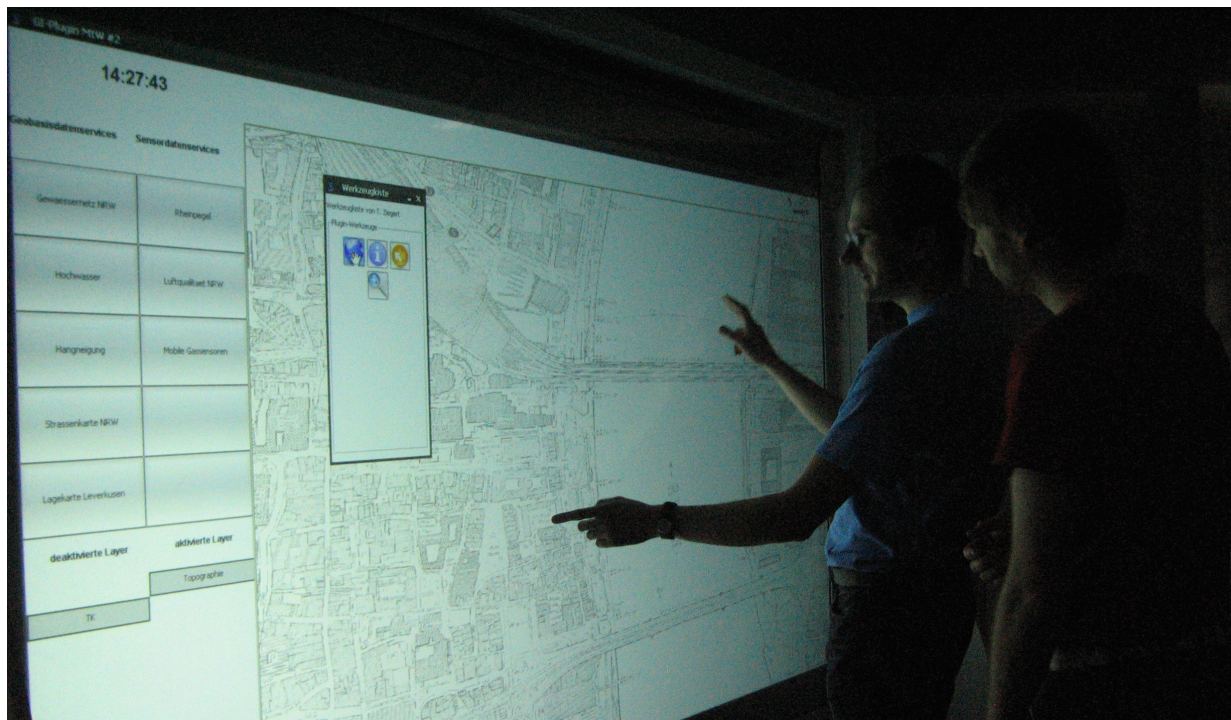


Figure 3.1.: Multi-touch GIS Interaction.

3.1.1. Multi-Touch Interactions

Three key interactions have been identified as introduced above: physical interactions, IPs and interaction space. These interaction components define the commands and controls that are needed to manipulate the interaction space.

Physical Interactions

As a first step towards the multi-touch framework, a set of simple physical interaction patterns for multi-touch input was generated (see Figure 3.2, inspired by [209, 212]).

These physical gestures can be classified in two dimensions: uni- vs. bimanual and finger vs. hand gestures. For the latter, there are three classes of these patterns: simple fingertip (F), palm-of-the-hand (H) and edge-of-the-hand (EH) input. Unimanual gestures are simple single hand gestures (gestures with the suffix 1 and 2). Bimanual gestures are simple two-handed gestures (3 – 5) as well as the combination of unimanual gestures (ones and twos) that result in more complex two handed gestures. Gestures $F1 - F5$ are based on one or two single-finger touches. Interacting using one or two whole hands is performed with gestures $H1 - H5$. The main idea behind the F and H interaction classes is the direct manipulation of region shaped objects. To interact with line-like objects and to frame or cut objects, the edge of the hand provides another class of gestures (EH). Each interaction class contains the following gestures: single pointing touch (1), single moving touch (2) (not limited only to linear movement), two touches moving in the same direction (3),

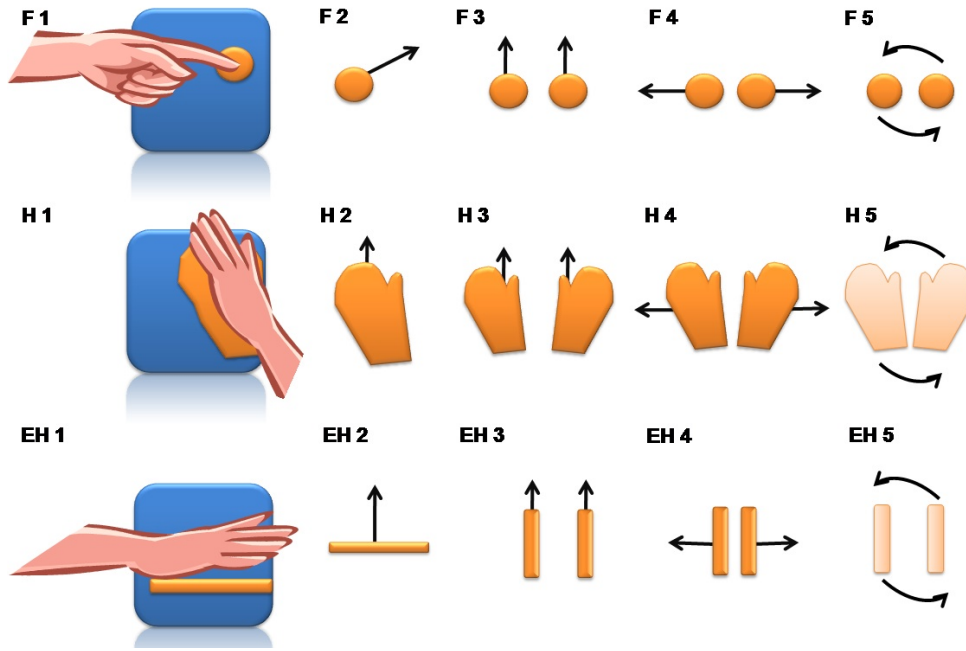


Figure 3.2.: Set of physical multi-touch gestures.

two touches moving in opposite directions (4), and moving of two touches in a rotational manner (5).

Interaction Primitives

A set of IPs for interaction with 3D geospatial data is defined. These commands and controls (such as pointing or zooming [77]) are needed to manipulate the geographic interaction space (at different scales) as well as to select, modify and annotate 2D and 3D objects. The universal GIS tasks are pointing, zooming, panning, rotating, tilting and cutting as described in [77].

Interaction Space

The interaction space contains a set of graphical views and representations for spatial objects. The view (spherical globe and plain map view), spatial objects (features), symbols and layers can be manipulated using the IPs, e.g. zooming, panning, and rotating the view, manipulate feature symbolization, or showing/hiding layers or features.

3.1.2. Study

We conducted an initial study to get a better understanding of the relationship between gestures and geospatial operations. 12 participants (five female and seven male) were asked to fill out a matrix, that assigns one or more IPs or a combination of IPs to certain

3. Classification of Multi-Modal 3D Interaction on Interactive Surfaces

	World		(Geo-) Objects			Symbols		Layer
	Globe	Plain	Point	Line	Polygon	Point-Symbols	Labels	
POINT	$F1$	$F1$	$F1$	$EH1$	$H1$	$F1$	$F1$	$F1$
ZOOM	$F4$	$F4$	-	$F4$	$F4$	$F4$	$F4$	$(F4)$
PAN	$H2$	$H2$	$F2$	$EH2$	$H2$	$F2$	$F2$	-
ROTATE	$F5$	$F5$	-	$F5$	$F5$	$F5$	$F5$	-
TILT	$H1 + F2, H1 + H2$	$H1 + H2$	-	-	-	$F6$	-	-
CUT				$EH1, EH3$	$EH1, EH3$			-

Table 3.1.: Framework for physical multi-touch interaction with geospatial data.

tasks. Five participants were employees of our institute, four graduate students outside of the subject area of geoinformatics and three domain experts (cartographers from a GIS company) who completed the subject pool. The participants were asked to fill out a matrix with either (1) one or more physical interaction possibilities, (2) an indication that a cell makes no sense (e.g. zooming a point object), or (3) combinations of primitives, e.g. pointing with two fingers ($F1 + F1$). They completed 149 different matrix cells, which include 12 proposed combined interactions. The preferred gestures got an average of 3.59 votes.

Comparing simple gestures (e.g. $F1$, $H1$) against more complex ones (e.g. $F5$, $H1 + F2$), the participants predominantly tended to prefer simple gestures, i.e. they tried to use the point-gesture (click) first and subsequently did something with the selected object. 52% of all interactions (66% as in the conceptual framework (CF)) were physical F-gestures (see Figure 2), 11% of all (10% in CF) were H gestures, 17% of all (17% in CF) were EH gestures and 26% of all (7% in CF) were combined gestures. Comparing one-handed against two-handed gestures, participants tended to prefer both hands (45% of all, 44% in CF) instead of just one hand (55% of all, 56% in CF). At this point, not all possible gestures are used in the framework because we started the framework with basic interactions, not more complex spatial interactions like buffering, intersecting two or more layers and so on. Interestingly, the participant from outside of the subject area had nearly 80 percent of her proposed gestures as simple one handed F-gestures.

3.1.3. Multi-Touch Framework

Based on the evaluation above, a framework has been built (see Table 3.1). The participants assigned physical gestures for the IPs to the interaction space. An interaction style was inserted in the framework if three participants agreed on the same interaction primitive. In the resulting framework the rows represent the IPs (a selection of the most common that are needed for geospatial tasks) and the columns of the table the interaction space (view, features, symbols and layers).

3.2. Extended Framework for Whole Body Interaction

Based on this framework, interaction (selection and manipulation) with geo-objects can be distinguished according to their geometric properties: point, line, and polygon. Interestingly, the geometric property of the interaction is reflected in the physical nature of the proposed multi-touch interaction. For example, single point-like objects are referred to with a single pointing gesture (*F1*), while rotation of a globe or panning of a 2D map is more likely to be performed by a wiping-style gesture (*H2*). The selection of geo-objects can be improved by referencing to their geometric properties. For example, the selection of a street on a map could be more precisely performed by moving a finger along that street (*F2*) instead of just pointing to it. This helps to reduce the ambiguity of the gesture as pointed out in [203].

This framework helped us to gain a basic understanding of multi-touch interaction with spatial data and thus can be seen as a first step towards a general framework. As mentioned above, not all of the primitive gestures of Figure 3.2 are listed in Table 3.1. For example, the two-hand gesture (*EH4* and *H4*) seems to be of no use. However we believe that if we look at more complex operations such as intersecting two polygons, these operations will become useful. In the following, other modalities will be explored that will also build the foundation for a general framework, i.e. a taxonomy for multi-modal 3D interaction with interactive surfaces.

3.2. Extended Framework for Whole Body Interaction

Even though multi-touch interaction has gained a lot of attention in the last few years the question remains how physical multi-touch gestures in combination with other modalities can be used in spatial applications. In the following we explore two modalities that go beyond touch: foot input and eye gaze. Those modalities allow intuitive interactions and extend multi-touch input by additional DOFs.

In the last few years, the possibilities of using feet for input were not considered in depth. Some researchers have done relevant work in the area of foot input for interactive systems, e.g. Pearson and Weiser [157] identified appropriate topologies for foot movement and presented several designs for realizing them. They showed in an exploratory study [158] that novices can learn to select fairly small targets using a mole. Pakkanen and Raisamo [154] highlight alternative methods for manipulating graphical user interfaces with the foot and show the appropriateness of foot interaction for non-accurate spatial tasks. Recently, also inspired by multi-touch technologies, Augsten et al. [1] explored foot interaction with floors. This work was extended to whole rooms where users and their poses were tracked by a pressure sensing floor [23].

This section describes two extensions of the multi-touch framework presented above. First, an approach for multi-touch and foot interaction is presented and second, gaze-based interaction is additionally included in the framework.

3. Classification of Multi-Modal 3D Interaction on Interactive Surfaces

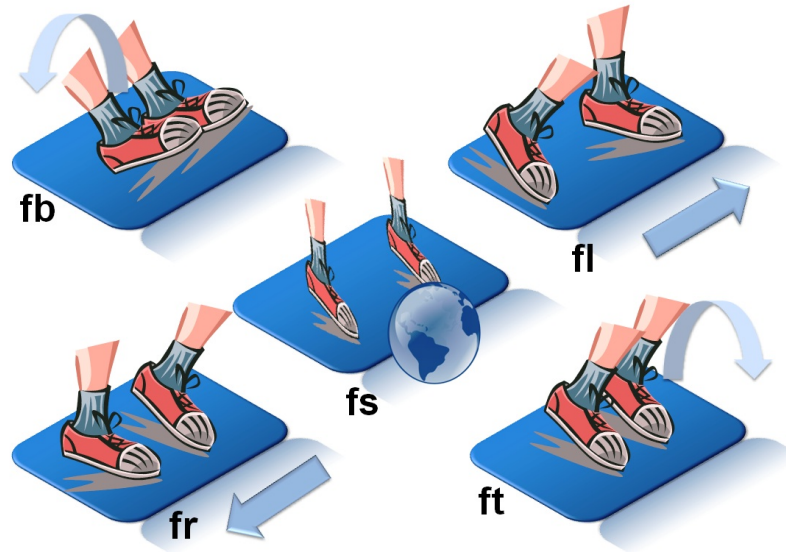


Figure 3.3.: Physical foot gestures.

3.2.1. Multi-Touch and Foot Interaction

A second iteration of interaction primitive design has been performed that is based on the multi-touch interactions above. A set of simple physical foot interaction patterns was developed that can be performed by a user standing on a Wii Balance Board. Up to now five different patterns (named with lower case letters) were taken into account: fb = “stand on balls of feet”, ft = “stand on tiptoe”, fr = “balance center on the right”, fl = “balance center on the left”, fs = “stand on sides of feet”. Most of those gestures are self-explanatory. For example ft means that the user is moving the balance point forward and just stands on tiptoe. fs denotes an action (user standing on sides of feet) people often perform while they are waiting (see Figure 3.3). Again, as in the multi-touch framework, these interactions can be assigned to specific tasks. Since those interactions are hands-free they provide additional DOFs and can be performed in parallel, e.g. together with touch input.

3.2.2. Gaze-based Interaction

Since gaze-based interaction is best as additional input mode [176], it is interesting to investigate how to integrate the user’s gaze information as another modality in spatial applications. This section describes the extension of the multi-touch framework presented above with the focus on gaze input. Eye gaze interaction allows hands-free interaction and therefore it is best suited as additional input for bimanual multi-touch interaction. In combination with another modality, gaze-based interaction is well suited for selection tasks [183]. Since selection of spatial objects is a crucial task in geospatial applications our extended framework for gaze interaction will mainly focus on selection tasks of the interaction space.

3.3. Taxonomy for Multi-Modal 3D Interaction On and Above Interactive Surfaces

	World		(Geo-) Objects		
	Globe	Plain	Point	Line	Polygon
POINT	$F1, GP$	$F1, GP$	$F1, GP$	$EH1, GM$	$H1$
ZOOM	$F4, F4 + GP$	$F4, F4 + GP$	-	$F4$	$F4$
PAN	$H2, fr, fl, GP$	$H2, fr, fl, GP$	$F2, fr, fl$	$EH2, fr, fl$	$H2, fr, fl$
ROTATE	$F5$	$F5$	-	$F5$	$F5$
TILT	ft, fb	ft, fb	-	-	-
CUT				$EH1, EH3$	$EH1, EH3$

Table 3.2.: Framework extensions for foot and eye gaze interactions.

In a first iteration, the following physical interactions are proposed to extend the basic framework with gaze input (see Table 3.2). Eye pointing on a specific location of the screen is denoted as GP . This can be used to highlight a point of interest or retrieve further information about objects (feature info). Another potential use is panning by looking at the horizon or the edges of the map to initiate “scrolling” [118]. Gaze motion (denoted as GM) can be used to retrieve information about objects of a certain geometrical form (e.g. line-objects). Furthermore, the combination of gaze and touch additionally offers novel and intuitive interactions, e.g. the user stares at a location where he or she wants to zoom in.

3.2.3. Extended Framework for Foot and Gaze-based Interaction

While IPs and the interaction space stay nearly the same (see Section 3.1.1) some IPs can be (additionally or exclusively) controlled by the feet or by eye gaze.

The proposed interaction styles for various selection and manipulation tasks are summarized in Table 3.2. The table is organized as described in section 3.1.3, but now filled up with physical hand and foot gestures and/or gaze input to interact with geo-objects. For example, panning can be accomplished by using the physical multi-touch interaction $H2$ or the foot interactions fr, fl and pointing by using eye gaze.

3.3. Taxonomy for Multi-Modal 3D Interaction On and Above Interactive Surfaces

A generic taxonomy for 3D interaction on and above interactive surfaces can be derived from the results above. The taxonomy is inspired the seminal work of Card et al. [29], but it extends in particular the taxonomy of 3D manipulation techniques by Martinet et al. [138]. While the taxonomy of Martinet et al. focuses on multi-touch (i.e. multi-finger 3D interaction techniques), our taxonomy consists of canonical 3D tasks that can

3. Classification of Multi-Modal 3D Interaction on Interactive Surfaces

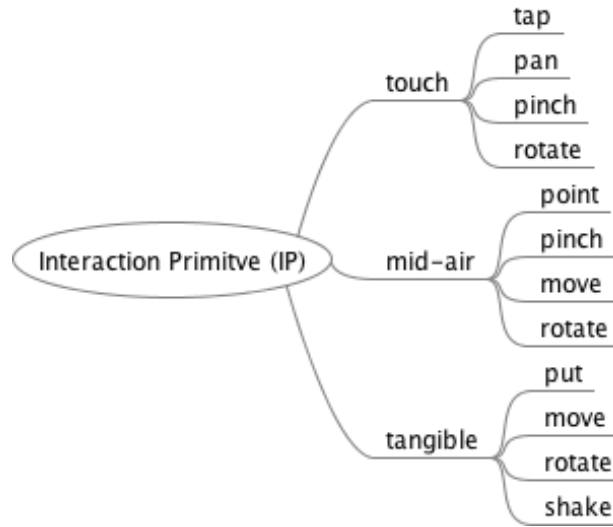


Figure 3.4.: Main IPs for the interaction with stereoscopic data on and above the interactive surface: touch, mid-air and tangible IPs

be performed with a more general set of input modalities. The taxonomy further takes different dimensions into account, including tasks, IPs, devices and modalities.

Canonical 3DUI tasks that are covered in the taxonomy are selection, manipulation and navigation. Touch, mid-air and tangible IPs are IPs that are of general interest for 3DUI. Examples for IPs are depicted in Figure 3.4. Another important design issue for 3DUI is separation of DOF. This basically means that some DOF are controlled not integrally, but separately. DOF separation is denoted by a connecting line between the IPs that control a task together.

This results in a taxonomy that basically consists of IPs and tasks (see Table 3.3). For a specific interaction technique, the table simply needs to be filled out accordingly by assigning IPs to tasks. Depending on the number of tasks an interaction technique covers, the table contents give an indication of completeness and universal applicability of the depicted technique. The number of devices, modalities and IPs of an interaction technique can be easily identified from the table. An IP can range across one or more DOF which is denoted by a horizontal rectangle that includes all DOF that an IP covers. A control that covers more than one IP is denoted by a dotted line that connects the corresponding IPs. This DOF markup also shows if DOFs are separated or not. Integrated DOF control by different IPs is denoted by line connecting the corresponding IPs. Altogether the number of IPs that are used gives an indication of complexity. This can be used to rate the complexity of an input technique.

In the following, two examples for applications of the taxonomy are presented. These examples are taken from interaction techniques that are presented in later chapters. Different input and output modalities were used to realize these 3D interaction techniques reaching from gestural and mobile interaction with large displays to fully mobile setups.

3.3. Taxonomy for Multi-Modal 3D Interaction On and Above Interactive Surfaces

Device/Modality	IP	Selection			Manipulation						Travel						Control	...
		translation			translation			rotation			translation			rotation				
		x	y	z	x	y	z	r_1	r_2	r_3	x	y	z	r_1	r_2	r_3		
$Device_1$	IP_1																	
	IP_2																	
$Device_2$...																	
	IP_n																	

Table 3.3.: Taxonomy for multi-modal 3D interaction.

3.3.1. Example 1: 3D Interaction on Mobile Devices

The interaction space can be extended by mobile and gestural interactions that rely on mobile devices. Built-in sensors can be used to allow 6-DOF input. The mobile properties of the device enable a variety of gestural input possibilities that go beyond pure touch input. With this approach, no external tracking system is needed to track how the user is holding, moving and touching the device. Further, the use of mobile stereoscopic displays requires specifically designed interaction styles for this class of 3D output devices. In the following, these interactions styles (proposed in Chapter 9.1) are presented as an example instance of the taxonomy above. The interaction techniques are briefly introduced together with the instance of the framework (see Table 3.4).

The mobile interaction techniques consist of object selection, manipulation and navigation. Objects selection and manipulation is realized by direct touch interaction. Selection is performed by a simple touch (IP: tap). To enable the selection of occluded objects a tilting mechanism allows “flipping” through space in the z-direction (as in navigation below). This results in 2-DOF tapping together with 1-DOF flipping for selection. Manipulation is also realized by direct interaction in combination with tilting and rotating the mobile device. Translate (IP: pan), rotate (IP: rotate), and scale (IP: pinch) are supported as direct manipulation tasks. In order to touch and select an object, the user needs to navigate through the scene. Navigation was realized through rotation (IP: rotate) and flipping (IP: flip) gestures of the mobile device. The virtual camera can be physically moved in every direction to change the field of view, its focus, etc. and thus behaves similarly to a real camera.

This results in an instance of the extensible framework as depicted in Table 3.4. The IPs are denoted by rectangles that span over one or more dimensions indicating how many DOF an IP is able to cover. In this example, some of the tasks are realized by DOF separation, which is denoted by a connecting line. However, they are tracked by different

3. Classification of Multi-Modal 3D Interaction on Interactive Surfaces








Device/Modality	IP	Selection			Manipulation						Travel					
		translation			translation			rotation			translation			rotation		
		x	y	z	x	y	z	r_1	r_2	r_3	x	y	z	r_1	r_2	r_3
Touch and acceleration	Tap															
	Flip															
Touch and orientation	Pan															
	Tap															
	Rotate															
Orientation and acceleration	Rotate															
	Flip															

Table 3.4.: Framework for sensor-based 3D interaction on handheld stereoscopic devices.

sensors and can be thus performed in parallel, e.g. touching and rotating the device for object rotation.

3.3.2. Example 2: Gestural and Mobile 3D Interaction Above the Interactive Surface

The second framework example uses mobile and gestural 3D interactions above (or in front of) interactive surfaces and is depicted in Table 3.5. In combination with 3D tracking, robust 6-DOF input methods can be designed. The mobile device can be seen as a physical placeholder for the virtual object to interact with and is thus well suited for manipulation and navigation tasks. The device then serves as a passive haptic prop that supports the user's spatial orientation and control while bi-manual gestures enable a rich set of interactions. Performing selection with the non-dominant hand (NDH) and manipulation with the DH results in an intuitive yet effective bi-manual interaction technique.

In this example object selection is realized by a grip gesture of the NDH which serves as a simple toggle mechanism. Manipulation is realized by changing the position and orientation of the mobile device with the DH. Translation is realized by moving the mobile device. Rotation is performed by a shake metaphor in a non-isomorphic manner, changing the orientation of the mobile device. Indirect mapping is preferable due to the physiological constraints of the human hand for rotating a physical object. The scaling of the 3D scene

Device/Modality	IP	Selection	Manipulation					
		translation			translation			rotation
		x	y	z	x	y	z	r_1 r_2 r_3
Gesture Tracker	Grab	NDH						
	Move				DH			
Orientation	Rotate							DH

Table 3.5.: Framework for mobile and gestural 3D manipulation.

can be directly or indirectly mapped. In the ladder case, a manipulation tasks require clamping in order to move an object within the reference frame.

In the resulting instance of the framework (see Table 3.5) the IPs are again denoted by rectangles that span over one or more dimensions. This indicates how many DOF an IP is able to cover. The two IPs that allow integrated control of the translation and rotation are connected by a line. To denote which hand controls an IP, the rectangles are labeled with the according hand (i.e. DH or NDH). As the table clearly shows the presented bi-manual interaction technique supports integrated 6-DOF manipulation.

3.4. Conclusion

In this chapter a taxonomy for multi-modal 3D interaction on and above interactive surfaces is presented. The taxonomy is based on a framework that was iteratively developed from multi-touch to multi-modal interaction. Different modalities were investigated, namely the whole-body, the feet, multiple touches, and the eye gaze. The taxonomy for multi-modal 3D interaction on and above interactive surfaces was deduced from these frameworks.

In the initial approach, two steps of a framework for geospatial operations are presented. In the first step, multi-touch gestures to navigate and manipulate spatial data are derived from a usability inspection test. Based on the results of the multi-touch framework a first concept and implementation of the combination of multi-touch hand and foot interaction is provided. The combination of direct, position-controlled (hand) with indirect, rate-controlled (foot) input is proposed and evaluated in an initial user study. While hand gestures are well suited for rather precise input, foot interactions have a couple of advantages over hand interactions on a surface: (a) they provide an intuitive means to input continuous input data for navigation purposes, such as panning or tilting the viewpoint; (b) foot gestures can be more economic in the sense that shifting one's weight over from one foot to another is less exhausting than using one or both hands to directly manipulate the application on the surface, e.g. when trying to pan a map over a longer distances: (c) it provides additional mappings for evocative gestures, for single commands. In a multi-

3. *Classification of Multi-Modal 3D Interaction on Interactive Surfaces*

touch wall setup, foot interaction provides an orthogonal horizontal interaction plane to the vertical multi-touch hand surface and can be useful for improving the interaction with large-scale interaction multi-touch surfaces.

The resulting taxonomy for multi-modal 3D interaction on and above interactive surfaces is a generalized version of the above presented frameworks and at the same time outlines an approach to extend and adapt the framework for novel interactions. The taxonomy consists of different dimensions that can be filled with appropriate tasks, devices and input modalities, and IPs. The approach enables designers to easily and consistently integrate new modalities for 3DUI. The framework approach was further used in this work. The insights also informed other work on multi-touch and whole-body 3D interaction. Kerber et al. [108], for example, investigated 3D navigation in a first-person shooter with a specialized setup (a Wii Balance Board and the cubic multi-touch device *Cubtile* [58]) to enable whole-body interaction.

Part II.

Canonical 3D Tasks

4. Indirect 3D Selection

This chapter discusses selection of stereoscopically displayed 3D objects on interactive 3D surfaces. On stereoscopic displays objects, can be displayed with different parallax resulting in different stereoscopic effects. Objects can appear behind (positive parallax), on top of (zero parallax), or in front of (negative parallax) the screen. As mentioned in Chapter 2, interaction with objects that are displayed with different parallaxes is still a challenging task even in VR-based environments [182]. Multi-touch technology might be a good tradeoff to overcome this limitation by allowing a rich set of interactions without high instrumentation. However, objects displayed with positive parallax cannot be accessed by direct touch interaction, since the screen surface limits the user’s reach. Indirect selection techniques for such objects can be a feasible way to address this problem.

In this chapter, we therefore address the question of how users can select stereoscopic objects when the interaction is restricted to a two-dimensional multi-touch surface. Motivated by *Balloon Selection* [6], two indirect multi-touch selection techniques will be presented and studied. The proposed techniques allow seamless selection of stereoscopic objects displayed with differing parallax on a multi-touch display. The techniques are evaluated in an experiment that addresses multi-touch selection of stereoscopic objects displayed with varying parallax and position. The experiment gives insights on how the different parallax paradigms as well as the position of objects determine multi-touch selection techniques. The results of the user study indicate that the selection of elements in the negative parallax space is more difficult to perform than in positive parallax. A second result is that at regions at which the occlusion as well as the ambiguous depth cues are mitigated selections are not as difficult to perform. The contributions of this chapter have been partially published in [45, 67].

4.1. Multi-touch 3D Selection Techniques

Stereoscopic effects on screens are achieved by showing each eye of an observer a different image (see Section 2.1). The effect of objects floating in front of the screen is produced while the depth cues the brain obtains are ambiguous. The eye’s convergence presumes that two different images are seen, but the eyes need to focus on the screen instead of the objects in front of it. This leads to an accommodation contradictory to the convergence. Therefore it is expected that the selection of objects with negative parallax is more complex than the selection of objects with positive parallax. Indeed, for negative parallax, additional degrees of ambiguity exist. Objects appearing in front of the screen are clearly behind the user’s hand while touching the screen. Furthermore, by providing a selection pointer

4. Indirect 3D Selection



Figure 4.1.: Interaction with stereoscopic data on a multi-touch surface with anaglyph display (with the *Balloon/Fishnet Selection* technique).

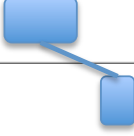
that needs to be moved with respect to the static 3D objects, motion parallax and relative object size depth cues are added to the interface. However, a result of the inconsistency between accommodation and convergence is that if the user's hand is focused on, the stereoscopic effect gets lost. To resolve the eye focus problem, the selection tool must be blindly manipulated so that the focus can remain steadily on the scene objects.

In its original version, the *Balloon Selection* used an HMD for the rendering of the stereoscopic 3D content. This hardware configuration avoids several ambiguities of depth cues. For stereoscopic multi-touch environments, indirect selection methods have not yet been considered. The aim of this work is therefore to determine the effects of parallax and ambiguous depth cues during the use of the widget-based selection methods. Two selection techniques for stereoscopic touch displays are proposed to investigate these effects: (1) *Balloon/Fishnet Selection* and (2) *Corkscrew Selection*. *Balloon Selection* has already been shown to be adequate for this kind of task, while *Corkscrew Selection* is expected to allow a less rigid manipulation since it can be used bi- or unimanually. Taking these two methods into account, more general conclusions can be deduced from the study obtained in this chapter.


4.1.1. Ballon/Fishnet Selection

In the original version of the *Balloon Selection* technique, users wear pinch gloves and perform gestures on a multi-touch tabletop. The user controls a 3D spherical cursor (balloon) above the surface by multi-touch gestures on the surface using a virtual string. *Balloon selection* is performed as follows [6]: To instantiate the balloon, the two index fingers (the

4.1. Multi-touch 3D Selection Techniques

Device/Modality	IP	Selection
		translation x y z
Touch	1st Finger Pan	
	2nd Finger Pan	

(a) Framework for *Balloon/Fishnet Selection*.

Device/Modality	IP	Selection
		translation x y z
Touch	Pan	
	Rotate	

(b) Framework for *Corkscrew Selection*.

Table 4.1.: Frameworks of the multi-touch 3D selection techniques.

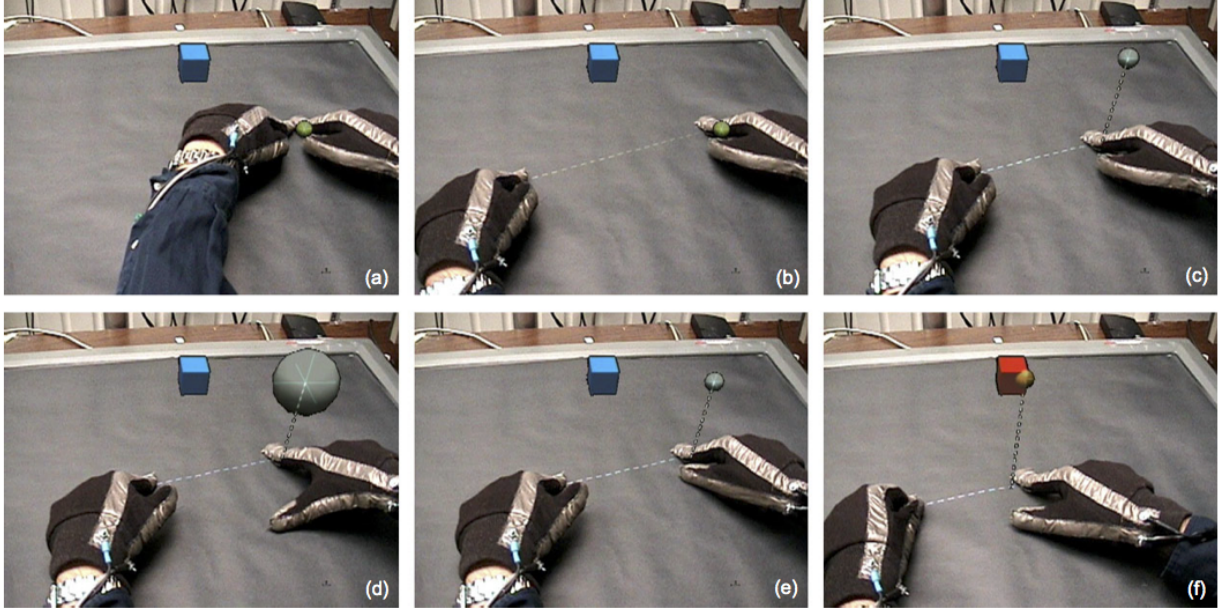


Figure 4.2.: The original version of the *Balloon Selection* technique [6]: users wear pinch gloves and perform gestures on a multi-touch tabletop. a) Instantiation of the balloon; b-c) Stretching the string and raising the balloon; d-e) Pinch to scale the balloon; f) Select an object.

4. Indirect 3D Selection

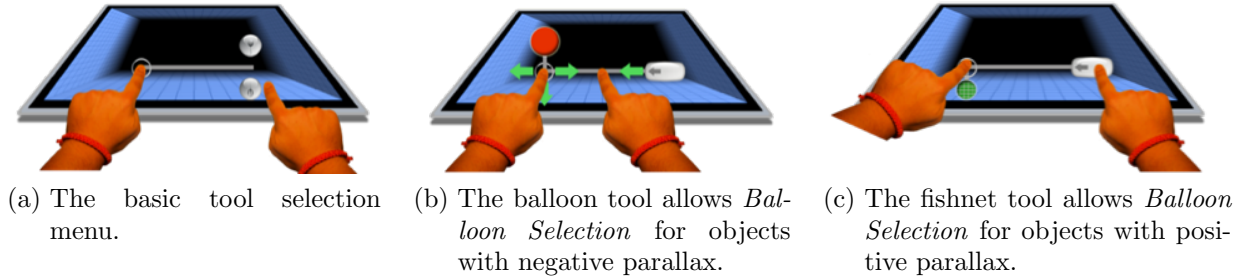


Figure 4.3.: Initialization of the *Balloon/Fishnet Selection* tool.

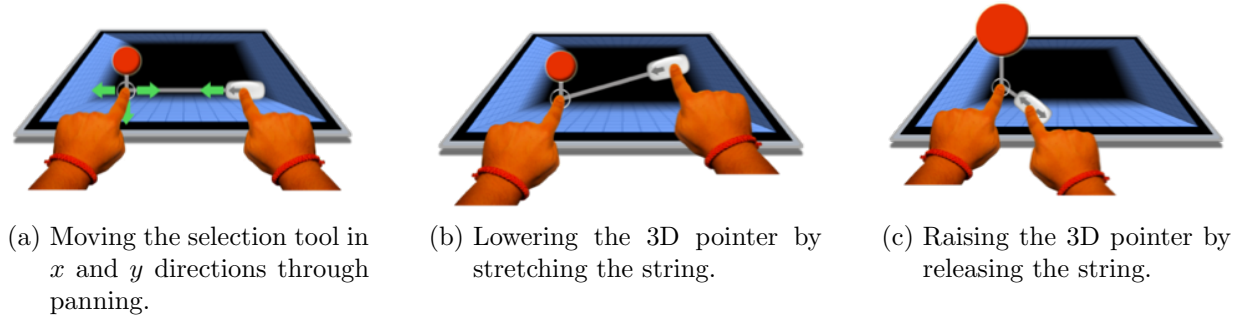
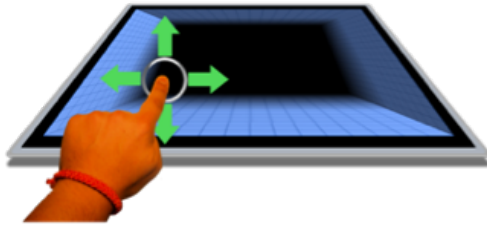


Figure 4.4.: Controlling the *Balloon/Fishnet Selection* pointer.

anchor and the stretching finger) are placed adjacent to each other on the tabletop (see Figure 4.2 a). Moving the stretching finger away from the initial position stretches the string between the two fingers (see Figure 4.2 b). Moving the index fingers closer together raises the balloon from the surface (see Figure 4.2 c). The balloon can be scaled by the thumb of the anchor's hand (see Figure 4.2 d-e). Moving the anchor on the surface translates the balloon parallel to the plane of the table while varying the distance between the anchor and the stretching finger determines the balloon's height (see Figure 4.2 b-c). The selection of the target is triggered by placing the thumb of the stretching finger's hand adjacent to the stretching finger (see Figure 4.2 f).

In order to allow interaction with a stereoscopic multi-touch display, the original *Balloon Selection* technique needs to be customized. Besides the balloon metaphor that is still used to select objects with negative parallax, a fishnet metaphor is used to support the selection of objects with positive parallax. Touching with one finger initiates the *Balloon/Fishnet Selection* tool (see Figure 4.3a). The initial state of the tool requires the user to either select the balloon tool (see Figure 4.3b) or the fishnet tool (see Figure 4.3c). The manipulable balloon's string is of a fixed size to allow all objects visible in the scene to be reached. On its end a button widget indicates the possible interactions. The selection tool can be moved in x and y directions through panning (see Figure 4.4a). By moving the string, the 3D pointer (balloon or fishnet) can be raised and lowered in the z direction (see Figure 4.4b



(a) Dragging the circle widget allows x and y translation.



(b) A circular movement performs the z translation..

Figure 4.5.: Controlling the *Corkscrew Selection* tool.

and 4.4c). Finally, the target object can be selected when it is fully encompassed by the 3D pointer. The framework that specifies the *Balloon/Fishnet Selection* is depicted in Table 4.1a.

4.1.2. Corkscrew Selection

The *Corkscrew Selection* technique is a selection technique that is somewhat similar to the scalable selection pointer of the *Balloon/Fishnet Selection* technique, but it uses another metaphor to raise and lower the selection tool. Touching and dragging the widget performs a translation along the x - and y -axes (see Figure 4.5a). Performing rotation gestures on the circular widget enables the user to steer the selection pointer. Counter-clockwise rotation on the widget makes the pointer rise up, while clockwise rotation makes the selection pointer drop down (see Figure 4.5b). When the 3D pointer encompasses the target object, the actual selection can be performed by pressing a button. The *Corkscrew Selection* is specified by the framework in Table 4.1b.

4.2. Experiment

An experimental study has been conducted in order to compare the two interaction techniques and to investigate multi-touch selection of stereoscopic objects displayed with varying parallax and position. The study thus gives insights on how the different parallax paradigms as well as the position of objects determine multi-touch selection techniques. Therefore, the following questions are addressed:

1. Parallax: the selection of elements with negative parallax vs. the selection of elements with positive parallax.
2. Position: The selection of the objects on the lower screen half vs. the selection in the upper screen half independent from the parallax.
3. Interaction technique: *Corkscrew Selection* vs. *Balloon/Fishnet Selection*.

4. Indirect 3D Selection

Participants were therefore asked to select clearly visible cubes, located at the corners of a fictive cube, that were placed at extreme parallax and task completion time and error rate were measured and analyzed.

4.2.1. Participants

13 subjects (seven female, six male) between 19 and 36 years old ($M = 22.5$) volunteered for the study. The participants received no monetary compensation and all were naive to the experimental conditions. One male candidate could not wear the anaglyph glasses over his sight-corrective glasses and was thus excluded from the experiment. Two female candidates were excluded as they felt too uncomfortable with the multi-touch surface. Therefore, ten (five female, five male) participants were considered for analysis. All participants were students, but from different scientific fields (biology: 1; computer science: 2; economics: 4; design and arts: 3). All were right-handed, had normal vision, and none of the participants had experience with stereoscopic 3D content (except for occasional 3D movies). Furthermore, the participants had little experience with multi-touch technology, and only half of them reported experience with smartphones (three female, two male subjects).

4.2.2. Conditions

For the test scenario, the subjects were asked to perform a selection task with both selection techniques (within-subject design). The technique with which a participant began was constantly alternated to counteract the familiarity with the hardware and the task she would gain while performing the first technique of the test.

4.2.3. Task

Clearly visible cubes, located at the corners of a fictive cube, that were placed at the extreme parallax of each respective space, had to be selected by the participants. The center of the cube was on the zero parallax plane, on which another object to select was placed, and a zero-parallax element was provided for completeness. For the negative parallax, the elements were placed on an extreme parallax surface close to the screen edges. For the two upper elements, a better performance was expected due to the advantage the linear perspective offers at these spots. For the positive parallax, elements were placed at the same distance from the fictive cube's center as the elements with negative parallax. A successfully selected object vanished and the next one appeared. The selection of one element corresponded to one trial. All interactions were logged for later analysis.

4.2.4. Design

The experiment was a $2 \times 8 \times 4$ within-subjects mixed factorial design. The factors were interaction technique, parallax and position. In each interaction technique, nine elements had to be selected that were characterized by their parallax space and position. In total

the selection of the nine elements had to be performed six times per selection method. The objects appeared in a fixed order, alternating negative and positive parallax. The two first objects of both parallax spaces were in the upper screen half, followed by zero parallax, and the last four objects were again altering between the parallax spaces in the lower screen half. The total number of trials amounted to $2 \text{ interaction techniques} \times 9 \text{ objects} \times 6 \text{ cycles} = 108$.

4.2.5. Procedure

The experiment was structured as follows. After a short introduction, the subjects were asked demographic questions. Then their ability to perceive stereoscopic vision was tested. One of the two selection techniques was demonstrated (including a free trial phase of two minutes). In the free trial phase, the participant was given time to get used to the system and the selection technique in a scene with two selectable and translatable elements. Afterwards she had to perform the actual tasks and select objects with the first selection technique, and then after a mandatory resting phase of five minutes, continue with the second technique. After the completion of all trials, a post-study questionnaire had to be filled out before the participant was debriefed. The experiment took around 45 minutes overall for each participant.

4.2.6. Apparatus

The study apparatus was developed using the Ogre SDK¹ to render anaglyph stereoscopic 3D content and the TUIO Reference Client² to receive touch signals. The implementation of the gestures was realized according to the approach used by Benko and Feiner [6]. An All-in-One Medion Akoya P4010 (MD8850) touch computer running Windows 7 was used for the study. With a screen diagonal of 55 centimeters up to two simultaneous touches were supported, and a red-cyan anaglyph was used to generate stereoscopic 3D output. To obtain TUIO signals, the touches received from the hardware drivers and processed by the Windows operating system were converted to the TUIO format and forwarded to the test application through the use of the Touch2TUIO³ software. The device stood on a bar table and was tilted by 45 degrees. The table's height was adaptable and adjusted the way so that the participant could comfortably stand in front of it and work with arms inclined by 90 degrees. A standard HD camera was placed on a tripod to record the task completion for later video analysis.

4.2.7. Independent and Dependent Variables

Parallax, position and technique were treated as independent variables, while task completion time and error rate were measured as dependent variables.

¹<http://www.ogre3d.org>

²<http://www.tuio.org>

³<http://dm.tzi.de/touch2tuio/>

4. Indirect 3D Selection

4.2.8. Hypotheses

According to the expected difficulties mentioned in the previous sections, the hypotheses to be verified by the experiment are:

- The selection of elements with negative parallax is more difficult than the selection of elements with positive parallax, since for the latter, more natural and unambiguous depth cues are provided (H1).
- The selection of the objects on the lower screen half is more complex than the selection in the upper screen half, independent of the parallax (H2).
- Corkscrew Selection performs better than Balloon/Fishnet Selection (H3).

4.3. Results

In the following the performance of the different selection techniques is investigated, with special focus on different parallax paradigms and positioning within the parallax space, and afterwards the techniques are compared to each other. Finally, subjective feedback from the questionnaires is provided.

Parallax and position were treated as independent variables, while task completion time and error rate were dependent variables. The data was evaluated using one-way and two-way ANalysis Of VAriance (ANOVA), and t-tests under the assumption of a confidence interval of 95% for all tests. The error metric was defined as follows:

$$errortimerate = \frac{errortime}{taskcompletiontime}$$

An error is committed when between two consecutive image rendering frames, the user increases the distance between the balloon pointer and the object to select instead of diminishing it. The error time rate indicates the precision with which a single selection task is performed.

4.3.1. Balloon/Fishnet Selection

Testing trials with negative parallax ($M = 7.13$, $SD = 1.62$) against trials with positive parallax ($M = 6.23$, $SD = 1.64$), the t-test results in $t(23) = 2.91$, $p < .05$ for task completion time. The t-test of the error time rate of trials with negative parallax ($M = 0.30$, $SD = 0.06$) against trials with positive parallax ($M = 0.03$, $SD = 0.03$) shows the following results: $t(23) = 16$, $p < .001$. Therefore, for both metrics a strong significant difference between negative and positive parallax exists. The average task completion time as well as the average error time rate is considerably higher for objects with negative parallax than for objects with positive parallax.

To evaluate the object positioning within the different parallaxes, the values for task completion time and error time rate from the four objects of positive and negative parallax

were separately evaluated in a one-way ANOVA. For objects with positive parallax, no significant effect was found between the different object positions, neither for the task completion time metric, nor for the error time metric (ANOVA results for task completion time: $F(3, 20) = 3.1$, $p = 0.34$, and error time rate: $F(3, 20) = 3.1$, $p = 0.07$). Position is not significant for task completion time of elements with negative parallax. However, for the error time the position is significant. Trials with objects on the upper part of the display in negative parallax space have a lower error time rate (task completion time: $F(3, 20) = 3.1$, $p = 0.38$; error time rate: $F(3, 20) = 3.1$, $p = 0.03$).

4.3.2. Corkscrew Selection

The results of the t-test analyzing the effect of parallax for the Corkscrew Selection show the same significant effects as the *Balloon/Fishnet Selection*. For *Corkscrew Selection*, the t-test also shows a strong significant difference between the two parallaxes for both metrics (task completion time: negative parallax ($M = 6.23$, $SD = 1.47$) vs. positive parallax ($M = 4.30$, $SD = 0.64$), $t(23) = 3.33$, $p < .05$; error time rate: $t(23) = 19.5$, $p < .001$). Similar to the *Balloon/Fishnet Selection* technique, selecting objects with negative parallax took more time and was less precise than for objects with positive parallax. This supports the previously introduced hypothesis that the selection of elements with negative parallax is more complicated than the selection of objects from the positive parallax space.

Regarding object position for positive parallax objects, no significance is found (task completion time: $F(3, 20) = 3.1$, $p = 0.81$; error time rate: $F(3, 20) = 3.1$, $p = 0.58$). Similar to *Balloon/Fishnet Selection*, the task completion time shows no significance, whereas for the error time rate there is a significant difference between the positions. Fewer errors were committed during the selection of the objects near the upper screen edge.

To determine if a selection had a better performance and a stronger impact on the parallax problems, the selection techniques are evaluated against each other in the following section.

4.3.3. Balloon/Fishnet vs. Corkscrew Selection

The effects for the different parallaxes can be determined for both selection techniques. Selections of objects with negative parallax take longer and are more error-prone than in the positive parallax condition. Object positioning within each parallax also has a great impact. For the evaluation, two consecutive two-way ANOVA tests are used, one for the task completion time and one for the related error time rate, in order to determine if one of the two methods has an impact on the selection in different parallax spaces. The independent variables in this test are the parallax spaces and the selection methods.

The aforementioned significance for the selection of elements in different parallax spaces is also visible ($p < .001$). A strong significance between the selection methods also exists ($p < .001$). Furthermore, the selection methods in conjunction with the parallax spaces show no significant difference ($p = 0.08$). The *Corkscrew Selection* ($M = 5.25$, $SD = 1.49$) significantly outperforms the *Balloon/Fishnet Selection* ($M = 6.68$, $SD = 1.71$) regarding

4. Indirect 3D Selection

task completion time. The average overall task completion time for the *Balloon/Fishnet Selection* amounts to 320.8 seconds while for *Corkscrew Selection* it amounts to 252.5 seconds, more than a minute faster.

For the error time rate analysis, a strong significant difference exists between the selection methods ($p < .001$) as well as a significance in the conjunction between the selection method and parallax variable. The values show that *Corkscrew Selection* ($M = 0.29$, $SD = 0.17$) is less precise than *Balloon Selection* ($M = 0.17$, $SD = 0.14$). For negative parallax (*Balloon Selection*: $M = 0.30$, $SD = 0.06$; *Corkscrew Selection*: $M = 0.46$, $SD = 0.1$) the error time rate is 50% higher with *Corkscrew Selection*, while for positive parallax (*Balloon Selection*: $M = 0.03$, $SD = 0.03$; *Corkscrew Selection*: $M = 0.13$, $SD = 0.004$) it is more than three times higher.

The evaluated results of the tests show that the selection of objects with negative parallax is harder to perform, in terms of task performance time and error time rate, than the selection of objects with positive parallax. Within the negative parallax space, the position of the object also influences its selection. Objects positioned above the horizontal middle of the screen are selected faster and more precisely. To sum up, the results of the post-study questionnaire are presented in the following subsection.

4.3.4. Post-Study Questionnaire

After the test, the participants were asked to fill out a questionnaire containing answers in a seven point rating scale, as well as free text forms. All questions had a positive connotation, so that full agreement is always expressed with the value seven (vs. one for complete disagreement). The questionnaire contains questions about the stereoscopic 3D effects of the scene and about the selection methods in order gain insights on selection of stereoscopic 3D elements, usability, learnability and joy of use.

The results of the general questions about the stereoscopic effect show that the parallax which is most difficult to recognize is the zero parallax, so negative and positive parallax are clearly distinguishable. The average answers to the questions concerning the parallaxes in conjunction with the selection task are shown in Table 4.2. The results from the logs are consistent with the participants' answers. The values for the questions concerning the selection of objects in negative parallax space are clearly lower than for zero and positive parallax. Indeed, selection in negative parallax appears more difficult to the test subjects. Most subjects chose the *Balloon/Fishnet Selection* as their favorite method, explaining that it was faster. The logs from the experiment, however, show that every participant performed better with the *Corkscrew Selection*.

4.4. Discussion

The study revealed that object parallax and object positioning within a parallax space have a strong impact on indirect multi-touch selection in stereoscopic environments. The selection of objects within the positive parallax space outperformed selection within the

Question	Parallax	Balloon/Fishnet	Corkscrew
Easy to select	Negative	5.7	6
	Positive	6.3	6.2
	Zero	6.2	6.1
Recognizable during task	Negative	5.9	6
	Positive	6.4	6.7
	Zero	6.4	6.3

Table 4.2.: Average results of the questions concerning selection in different parallax spaces (on a 7-point scale, with 7 being the highest score).

negative parallax space. This leads to the conclusion that the selection of elements with negative parallax is more difficult than the selection of elements with positive parallax. It takes more time and is less precise. This supports the previously introduced hypothesis that the selection of elements with negative parallax is more complicated than the selection of objects from the positive parallax space. It is most probably due to the conflicting depth cues the user perceives for negative parallax. The complexity of selection in negative parallax space was also noticed by participants themselves, as reported in the post-study questionnaire. For the object position, the task performance time shows no significant effect for the selection of objects with different positions, while the error rate shows a strong significant difference between the selection techniques. Objects placed at spots implying lower occlusion by hands or the selection tools could be selected more accurately.

In direct comparison, the same effects for the different parallaxes could be detected for both selection techniques. As supposed in the first hypothesis (H1), selections in negative parallax space take longer and are more error-prone than in the positive parallax condition. But object positioning within the parallax space is also of importance (H2). Due to the linear perspective of the 3D scene, the selection performance can be improved. Overall, *Corkscrew Selection* performs better in task performance time (H3). However, the *Corkscrew Selection* is less precise with respect to error time rate. The better performance for the task completion time of the *Corkscrew Selection* could be the result of a more linear selection process with a more accentuated DOF separation, since it is visible on the videos where users preferred a single-handed manipulation. But even if the *Corkscrew Selection* method performs better in terms of task completion time, the precision is lower. By inspecting the video footage, it can be observed that with the *Corkscrew Selection* method the subjects often started rotating their finger around the widget without previously worrying about the right direction. This might be due to the fact that when the wrong tool is chosen with the *Balloon/Fishnet Selection*, the process must be restarted in order to select the right tool, whereas with the *Corkscrew Selection* tool a flawless switch is possible. So the low precision of the *Corkscrew Selection* tool compared to the *Balloon/Fishnet Selection* might be counteracted by providing better guidance for the *Corkscrew Selection* tool (e.g. symbols on the flat buttons that indicate direction).

4. Indirect 3D Selection

The results of the post-study questionnaire underpin the results of the experiment as they also justify the hypothesis that selection in negative parallax is more difficult to perform. In contrast to the measured task completion time, the participants subjectively judge the *Balloon/Fishnet Selection* technique as faster than the *Corkscrew Selection* technique. The subjects' preference might therefore be related to a more dynamic handling offering a greater joy of use.

4.5. Conclusion

In this chapter we discussed selection techniques for stereoscopic data on multi-touch displays. Two indirect selection techniques were presented that enable users to seamlessly select stereoscopic objects displayed with different parallax on a multi-touch display. To study the indirect multi-touch 3D selection techniques, an experiment was conducted. The main goal of the experiment was to gain insights on how the different parallax paradigms as well as the position of objects determine multi-touch selection techniques.

The results of the experiment indicate that the selection of elements in the negative parallax space is more difficult to perform. This might be related to the ambiguous depth cues that touch interaction with such stereoscopic content involves. However, we found out that even if selection in negative parallax space is difficult, it still remains feasible. A second result of the study is that at regions at which the occlusion as well as the ambiguous depth cues are mitigated, the selection task was less difficult to perform.

5. 3D Manipulation

This chapter discusses the manipulation of 3D objects as a canonical 3DUI task (see Chapter 2.3.1). The manipulation of 3D objects requires at least six DOF that can be directly controlled by means of 6-DOF input devices (see Chapter 2.2.2). In the following, an interaction technique for the manipulation of 3D objects is proposed and evaluated in a docking task experiment with varying display conditions (monoscopic vs. stereoscopic display).

The proposed interaction technique is specifically designed for commodity 3D input devices that are widely available and affordable for everyday use. A sensor-based gestural interaction technique for stereoscopically displayed 3D data is presented and evaluated in a docking task experiment on a stereoscopic display. The experiment further aims to investigate how stereoscopy affects the precision of 3D manipulation. The results give insights on how commodity input and output devices can be used for complex 3D interaction. The conclusion is two fold. The interaction technique has proven to be usable and fast for users that are more experienced with 3D, but very difficult to use for novice users. Object translation precision is higher in stereo. Furthermore, the overall conclusion is that stereoscopy leads to more precise rotation for docking tasks, especially when simultaneous manipulations on all three axes are required. Contributions of this chapter have been published in [49, 179].

5.1. Bimanual Gestural and Mobile 3D Manipulation

Again, the goal was to design an interaction technique that relies on affordable commodity input devices, namely mobile devices and depth cameras. We expect that the combination of these two device classes allows the implementation of robust 6-DOF input methods. Furthermore, the physicality of mobile devices supports the user's spatial orientation and control, as it serves as a passive haptic prop.

A bimanual interaction technique is proposed for object manipulation. Object selection is realized by a simple toggle mechanism that is activated by a grip gesture of the NDH (see Figure 5.1). The user can manipulate the object only if the interaction is active. Translation is realized by moving the object with the DH (i.e. the mobile device). Rotation is performed by a shake metaphor in a non-isomorphic manner changing the orientation of the mobile device (again with the DH). The framework that specifies this interaction technique was already introduced in Table 3.5 in Chapter 3. An indirect mapping approach was chosen for rotation due to the physiological constraints of the human hand. To stay consistent, the 3D scene's dimensions were chosen in such a way that the manipulation task requires clutching.

5. 3D Manipulation



Figure 5.1.: Mobile and gestural interaction with stereoscopic 3D user interface in a docking task. The left hand (NDH) toggles the object selection with a grip gesture. The right hand (DH) controls the 6-DOF object manipulation. The input is captured with the Kinect and the orientation sensors of a mobile device.

To evaluate this interaction technique, we conducted a docking task experiment. In this experiment the participants were asked to perform a 3D virtual docking task with a 3D object (the *Utah teapot*) on a stereoscopic display in two conditions (stereoscopic versus monoscopic). The experiment is introduced in the next section.

5.2. Experiment

An experiment was conducted to evaluate the above-presented interaction technique. Besides the evaluation of the general usability and appropriateness of the interaction technique for commodity devices, the goal of the experiment was to investigate how stereoscopy affects the precision of 3D manipulation. Therefore, the participants were asked to perform a three-dimensional virtual docking task with a 3D object (the *Utah teapot*) on a stereoscopic display in two conditions (stereoscopic versus monoscopic) in a within-subjects experiment. As quantitative metrics, the elapsed time for completing a trial (task completion time), the precision of the object translation (translation task precision) and the precision of the ob-

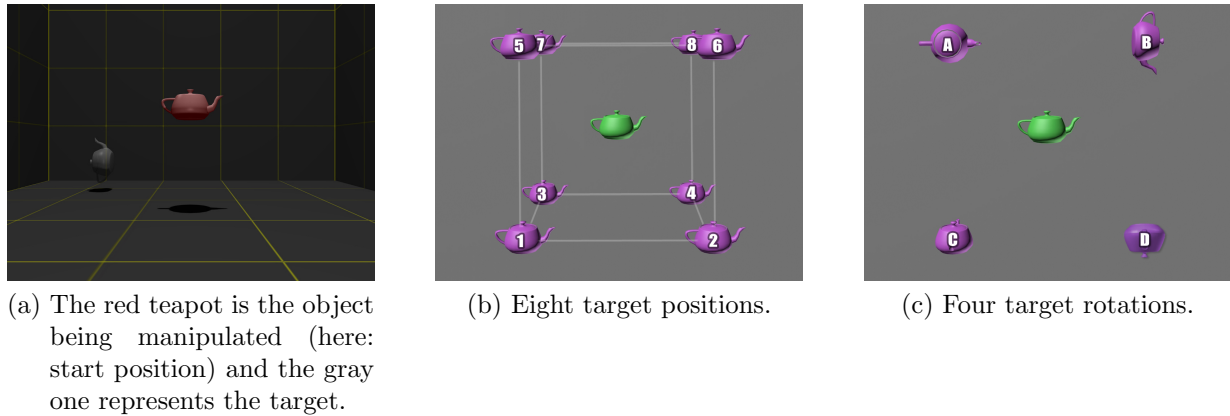


Figure 5.2.: Experiment setup conditions and tasks (c.f. [179]).

ject rotation (rotation task precision) were measured. In addition, subjective feedback was gathered using the NASA TLX [87] rating scale.

5.2.1. Participants

14 participants (four female and ten male) volunteered for the experiment. All subjects were right-handed and aged between 21 and 35 ($M = 26.57$, $SD = 3.3$). They had varying experience with computer science, 3D modeling and graphic software. All subjects claimed to have former experience with stereoscopic visualization (3D movies, television, other studies, etc.) and all of them owned smartphones and were very familiar with using them (usage of more than six hours per day on average). Two subjects were rejected as being unable to complete the docking tasks in stereo. The participants received no monetary compensation for their participation in the study.

5.2.2. Conditions

The proposed interaction method was tested under two conditions: a monoscopic and a stereoscopic condition. In both conditions, the interface, task and interaction technique stayed the same. In the stereoscopic condition the participants were required to wear polarized glasses to perceive the stereoscopic 3D content.

5.2.3. Task

The proposed interaction technique was evaluated in a docking task. In the docking task, the participants were asked to change the position and orientation of a virtual object (the *Utah teapot*) until it fits in a gray semi-transparent target object (see Figure 5.2a). The red teapot is the object being manipulated and the gray one represents the target. Eight

5. 3D Manipulation

different positions (see Figure 5.2b) and four rotations (see Figure 5.2c) resulted in 32 different tasks.

The *Utah teapot* model was chosen because it is unique in its dimensions and only symmetric in one axis (i.e. from lip to handle). By using this model, a fairly complex docking task was realized, unlike the manipulation of a standard cube or tetrahedron. The interaction technique described above was used to perform the task. The selection of the teapot with the grip gesture of the NDH is indicated by a color change of the teapot (yellow = toggled, red = not toggled). To keep the scene as simple as possible and let the participant focus on the task, only one selectable object was placed in the center of the scene at a time. Further, the participants were not required to scale the object to fit the target. The size of each model stayed the same throughout the whole experiment. The selection itself was a target acquisition task and was performed by direct target selection.

A trial was successful when the difference between the manipulated object and the target was lower than five units (Euclidean) distance for translation precision and 30° for rotation precision.

5.2.4. Design

The experiment was a $2 \times 8 \times 4$ within-subjects factorial design. The factors were display mode, position and rotation. The trials for each display mode were the product of eight position (see Figure 5.2b) and four rotation conditions (see Figure 5.2c). Three simple rotations (around only one axis) and one complex rotation (around all three axis) had to be conducted. Hence, each participant performed 32 trials in each of the two conditions of the display mode (monoscopic and stereoscopic). The order of these conditions was counterbalanced across participants, while the order of trials was randomized. Altogether $12 \times 2 \times 8 \times 4 = 768$ trials were conducted.

5.2.5. Procedure

After a brief introduction to the experiment, the participants were asked to fill out a questionnaire with demographic data. Then the task was introduced to the participants and they were instructed how to perform the interaction technique. Only minimal instruction was given to the participants on how to use the input devices and thus to perform the interaction technique. The participants were told to decide by themselves how to hold and use the mobile device as it felt natural to them.

The participants were asked to select an object by a grip gesture with the NDH that instantly toggled the manipulation phase. In this phase they could move and rotate the virtual object with the device in the DH as long as it was selected. The selection state was visualized by the color yellow for selected objects and red for unselected objects. Before the actual manipulation, the device parameters (position, velocity and acceleration) of the initial position and rotation needed to be specified. The velocity was determined by the movement speed of the mobile device, while the acceleration of each interaction was defined by a constant value. The distance and direction from the initial position to the

target position were essential parameters to determine the translation distance. The initial and final orientation, and the amount of rotation per axis were the crucial parameters for measuring the rotation accuracy.

When the participant decided that the object's position and orientation fits into the target, the next trial was started from the initial position. The participants were asked to perform each trial as precisely as possible within 60 seconds. After all trials for one display mode the participants were required to take a mandatory break of five minutes. Then all trials of the second mode followed.

Finally the participants were asked to fill out a post-study questionnaire and then debriefed on the experiment. The whole experiment took 45 minutes on average per participant, including the introduction, the actual experiment, the mandatory break and the questionnaires.

5.2.6. Apparatus

The study setup was developed using a scalable rendering environment for the rendering of stereoscopic 3D content (cf. [49] for more details on the framework). The same apparatus was also used in the travel experiment in Chapter 6. A Windows system on an Intel Core i5 4x 3.20GHz with 8GB of RAM and an nVidia GeForce GTX 660 Ti graphics card was used and the software was written in C++ and DirectX. A projection wall for polarized stereoscopic display with a size of $5 \times 3m^2$ and a screen diagonal of circa 5.5 meters was used. All participants were placed in front of the projection at a distance of 2.5 meters. As input devices a mobile device (iPod Touch) and a depth camera (Microsoft Kinect) were used. The Apple iPod Touch 4G was used for touch and orientation tracking. The sensor data of the device was streamed via a wireless network. The Kinect was used for whole-body skeleton tracking of the 3D positions of the user's hands, shoulders, head, etc. Predominantly hand gestures were tracked, but head and body motions were used as well.

5.2.7. Independent and Dependent Variables

Target position, target rotation and display mode were treated as independent variables while task completion time, translation task precision and rotation task precision were measured as dependent variables.

5.2.8. Hypotheses

Our main hypotheses to be verified by the experiment were:

- The participants' translation task precision is significantly higher in the stereoscopic than the monoscopic condition, particularly regarding the depth axis (H1).
- The participants' three-dimensional rotation task precision is significantly higher in the stereoscopic condition than in the monoscopic one (H2).

5. 3D Manipulation

- The participants' precision in the tasks that require rotations around all three axes is worse than in tasks that require only one-dimensional rotations (H3).

5.3. Results

In the following the results of the task completion time, translation task precision and rotation task precision for both display modes are reported. In the results section, the following encoding scheme is used for all charts: numbers denote the position condition (1-8), whereas letters encode the rotation condition (A-D).

5.3.1. Task Completion Time

Task completion time was measured from the first to the last user interaction of a task. The overall task completion time was on average 18.53s ($SD = 9.99$) in the monoscopic condition and 19.66s ($SD = 10.44$) in the stereoscopic condition.

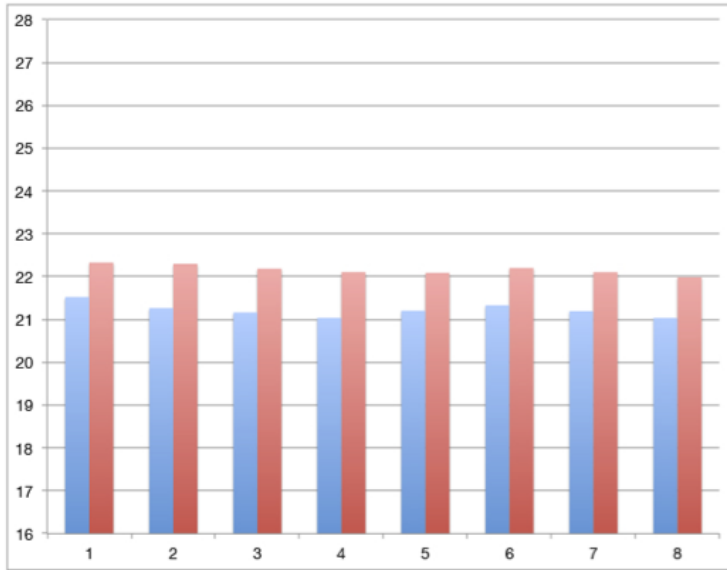
Univariate ANOVA analyses showed no significant difference in the display mode condition for task completion time ($p = 0.056$) while target position had a significant effect on task completion time ($p < 0.01$). In addition, univariate ANOVA analysis for the rotation task condition also showed a significant effect on task completion time ($p < 0.01$). In summary, the average task completion time was worse in the stereoscopic condition for both target position and target rotation, except for rotations around the y-axis (see Figure 5.3).

5.3.2. Translation Task Precision

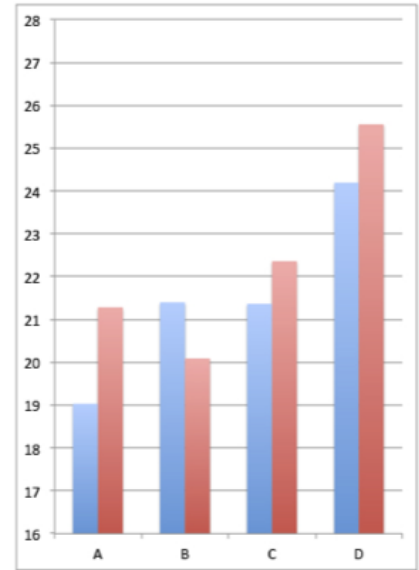
As translation task precision the Euclidean distance p_{trans} between the object position \vec{P}_1 and the target position \vec{P}_2 was calculated as follows:

$$\begin{aligned}\vec{P}_1 &= (x_1, y_1, z_1)^T \\ \vec{P}_2 &= (x_2, y_2, z_2)^T \\ \vec{P}_{dist} &= d(\vec{P}_1, \vec{P}_2) = \vec{P}_2 - \vec{P}_1 \\ p_{trans} &= \left\| \vec{P}_{dist} \right\|\end{aligned}$$

There was no significant effect in the display mode condition. Although there was no significant difference of the translation precision between target positions, in general the average overall precision was higher in the monoscopic than in the stereoscopic condition (see Figure 5.4). The fact that the two upper back target positions performed worse indicates a worse precision in translation along the y -axis, while translation along the x - and z -axis was performed better in the stereoscopic than in the monoscopic condition (see Figures 5.5a and 5.5b). Rotation around the z -axis performed the worst with an average precision of 12.13 units in the stereoscopic and 10.96 units in the monoscopic condition.



(a) Average completion time w.r.t. the eight target positions.



(b) Average completion time w.r.t. the four target rotations.

Figure 5.3.: Task completion time for position and orientation in the monoscopic (blue) and stereoscopic (red) conditions [179].

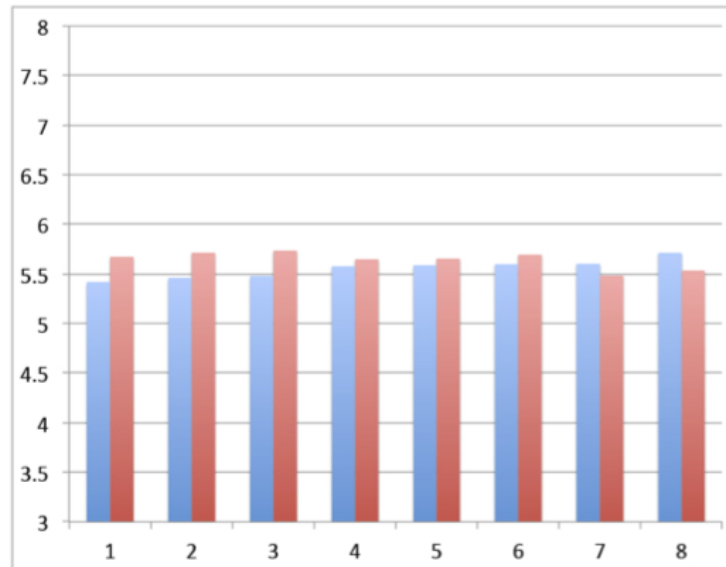


Figure 5.4.: Overall translation task precision in the monoscopic (blue) and stereoscopic (red) conditions of the eight target positions [179].

5. 3D Manipulation

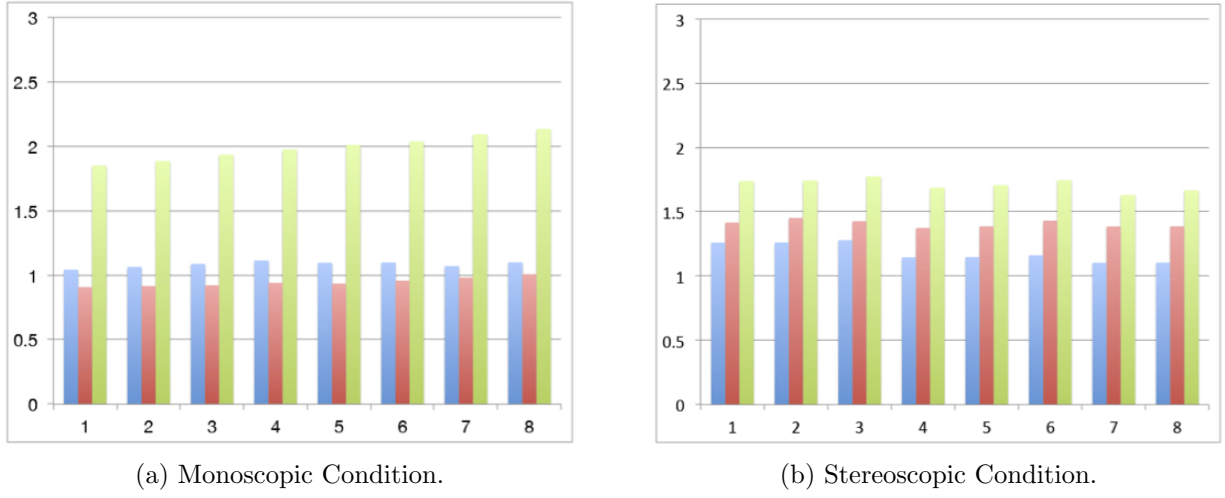


Figure 5.5.: Translation task precision for the eight target positions by axis (x, y, z → blue, red, green) [179].

5.3.3. Rotation Task Precision

Rotation task precision is defined by the quaternion metric (cf. [117]):

$$Q1 = (a1, b1, c1, d1), Q2 = (a2, b2, c2, d2)$$

$$p_{rot} = \sqrt{a_1 \cdot a_2 + b_1 \cdot b_2 + c_1 \cdot c_2 + d_1 \cdot d_2}$$

with $Q1$ and $Q2$ as object and target quaternions and p_{rot} as the resulting rotation precision.

The rotation precision was on average higher in the monoscopic condition, showing a significant effect for display mode, but no significant difference for target rotation (see Figure 5.6a). When inspecting the rotation precision per axis the rotation offset for the z-axis was better in the stereoscopic condition (see Figure 5.6b and 5.6c). In summary, simple trials that consists of only one-dimensional rotations (A,B,C) were performed more precisely than the complex trials with three-dimensional rotations (D).

5.3.4. NASA TLX

The NASA TLX questionnaire generally includes six sub-scales: Mental Demand (MD), Physical Demand (PD), Temporal Demand (TD), Performance (PE), Effort (EF) and Frustration (FR) [87]. Due to the high complexity of this task, the frustration sub-scale was not considered. The results of the remaining five NASA TLX sub-scales are illustrated in Figure 5.7a. Regarding the average overall workload for the respective sub-scales, effort and physical demand dominated, with the highest average values (5.81 and 5.88) in comparison to the other sub-scales. While mental demand and performance remained in mid-field (5.73 and 5.00), the temporal demand had the lowest scaling average of 4.04.

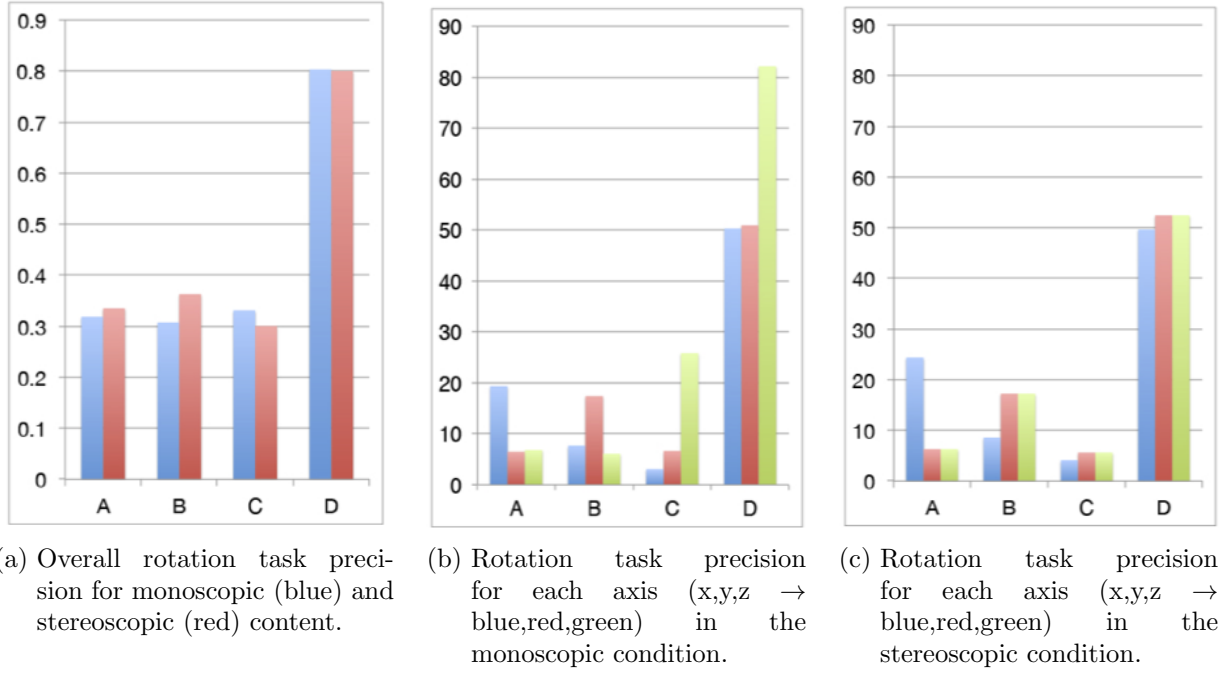


Figure 5.6.: Rotation task precision regarding the four target rotations (A,B,C: simple; D: complex) [179].

The average overall workload regarding the two display modes resulted in 5.22 ($SD = 2.02$) for the monoscopic condition and 5.03 ($SD = 1.70$) for the stereoscopic condition with an overall average of 5.09 ($SD = 1.79$) for both conditions (see Figure 5.7b).

In conclusion, the monoscopic display condition performed slightly better (5.17) than the stereoscopic condition (5.25). However, in all other sub-scales, in particular mental demand (4.67 vs. 4.25) and effort (6.25 vs. 6.00), the monoscopic condition performed worse than the stereoscopic condition. Interestingly, the stereoscopic condition (5.67) required a lower physical demand than the monoscopic condition (5.75).

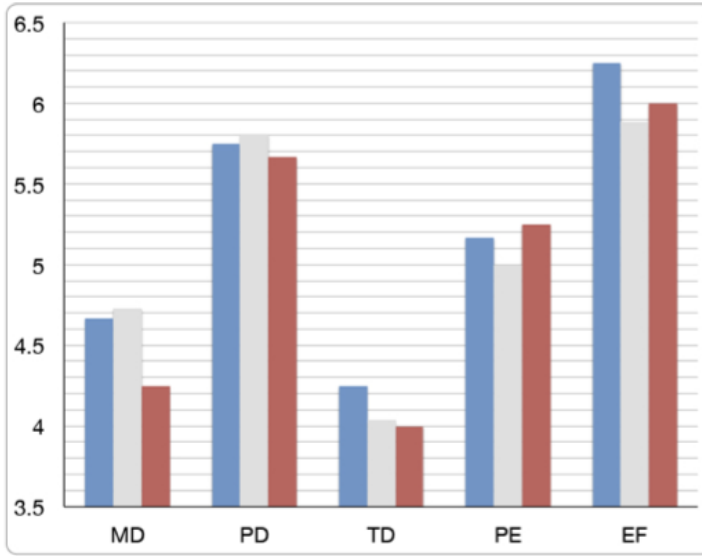
5.3.5. Observations During the Study

Only a few participants really used the mobile device as a physical prop. This could be observed by the fact that they utilized a certain point on the mobile device (e.g., the home button) to align the physical object with the virtual teapot.

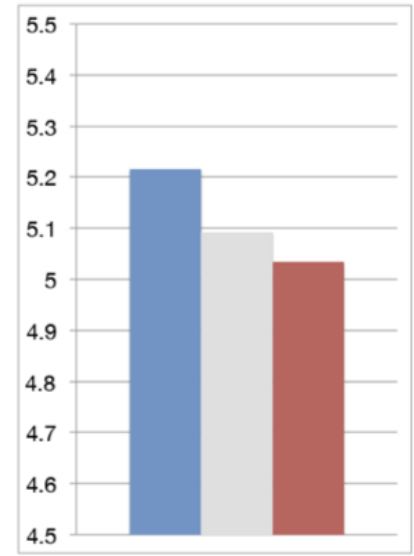
Another observation during the study was the consistently ordered subdivision of object manipulations performed by the subjects. Starting with a coarse rotation, followed by the translation of the object into the target, they finalized the task with a fine-grained rotation.

A general observation is that more experienced participants (e.g. experts in 3D modeling) performed fully integrated three-dimensional manipulations while the others did not. Novices, on the other hand, manipulated objects with fewer DOF at the same time than the more experienced ones.

5. 3D Manipulation



(a) NASA TLX sub-scales.



(b) NASA TLX overall workload.

Figure 5.7.: NASA TLX results: The blue color indicates the results for the monoscopic condition, the red color for the stereoscopic condition. The gray color indicates the overall results independent of display mode [179].

5.4. Discussion

The aim of this study was to explore the effectiveness and expressiveness of 6-DOF input with a mobile device that is used as a placeholder for the object to be manipulated in a stereoscopic VE. Mid-air gestural input is assumed to be natural and quick but at the cost of precision, due to the instability of the hand moving in the open air. Thus the experiment was mainly focused on precision rather than speed.

5.4.1. Interaction Technique

The main benefit of integrated 6-DOF interaction is that it enables the participants to manipulate a virtual object simultaneously without the need for mode changes. Although 6-DOF devices have shown to perform better for 3D manipulation (e.g. [217]) it is still unrefuted that they require a higher cognitive load (cf. [21]). This is reflected by the temporal results (measured task performance time and subjective temporal demand) as well as the perceived frustration. These effects had been observed in particular at the beginning of the task. The insights confirm other studies that show one has to carefully design 6-DOF interaction techniques when targeting novice users. Nevertheless, after a reasonable learning phase it could be observed that even the novices were improving. One solution might be to design interactions that comprise novice and expert modes that differ in separated and fully integrated DOF manipulations.

5.4.2. Monoscopic vs. Stereoscopic Display

Overall, the experimental results indicate that translation and rotation task precision was higher in the monoscopic than in the stereoscopic condition. However no significant effect could be shown and thus H1 had to be rejected. The contradicting results regarding the translation precision indicate that there are ergonomic and perceptual issues involved. Lifting the hand requires higher physical demand that might be compensated for by a lower cognitive demand on the depth perception. To comprehend these issues in more depth they need to be investigated in future studies that treat these dimensions separately.

The rotation precision was on average higher in the monoscopic condition, showing a significant effect for display mode which contradicts H2. The rotation precision results were highly dependent on the complexity of the rotation. The stereoscopic condition notably outperforms the monoscopic one in complex, simultaneous manipulations that require all DOFs (i.e. when the participants had to rotate the objects around more than one axis) which confirms H3. While the precision difference between monoscopic and stereoscopic might be negotiable for applications that only deal with simple single-axis rotations, complex manipulations are critical for monoscopic displays.

5.4.3. Docking Tasks

In 3DUI and VR research, no general docking task metric exists. In the research community there is also no common docking task that is used as a standard in experiments. Sometimes simple objects such as cubes or tetrahedrons are used, sometimes more complex 3D objects (e.g. the *Utah teapot* or the *Dragon*). We believe that specific (sub-)tasks, conditions and models are needed to make the experiments comprehensible and replicable. One reason for that might be the fact that the normalization of translation and rotation is very complex. If one takes the ambiguity of the rotation into account, it gets even more complex. Thus only a few metrics have been proposed [139, 217] in this direction. However, they are all based on Euler angles, which makes them at least questionable. Rotation metrics that rely on quaternions have been proposed as well (e.g. [98, 117]). However, to our knowledge no attempt to integrate translation with a quaternion-based metric has been undertaken.

To conclude, a somewhat artificial docking (without physics) was performed as in many other studies. While the stereoscopic condition showed a significant effect on precision in 3D rotation, the overall performance was bad. This could be improved, for example, by reducing the complexity of the rotation by allowing only single-axis rotations at a time. This could also be achieved by a separation into novice and expert modes. The introduction of a rotation snapping mechanism could also make sense for real application scenarios (but not all the time).

5.5. Conclusion

In this chapter mobile and gestural 3D manipulation was investigated. A mobile and gestural interaction technique for stereoscopically displayed 3D data was presented and

5. 3D Manipulation

evaluated in a docking task experiment. The results give insights into how commodity input and output devices can be used for the interaction with stereoscopic 3D data in everyday life. Our results show that the positioning precision is higher in stereo and furthermore indicate that stereo notably outperforms mono in tasks that require simultaneous rotation on all three axes.

A standardized docking task to evaluate 3D input devices and interaction techniques would be desirable, like ISO9241-9 [99] for selection tasks and should be followed up in future work.

The bimanual interaction techniques developed in this chapter inspired another set of novel mobile and gestural interaction techniques. In particular, they informed 3D interaction techniques for travel that also rely on well-established interaction metaphors and make use of commodity devices as spatial input. Thus, in contrast to the manipulation task above, the travel techniques in the following section were evaluated in a virtual search task on a large stereoscopic 3D display. Consequently, the same testing environment was used to study both the universal manipulation task above and the travel task in the next chapter.

6. 3D Travel

In this section we investigate travel, the canonical 3DUI task (see Section 2.3.1). More precisely, 3D travel techniques are proposed and evaluated that rely on mobile and gestural input. The proposed travel techniques are motivated by the bimanual manipulation techniques developed in Chapter 5 and consequently continue the investigation of universal 3D interaction tasks that make use of commodity devices as spatial input.

As mentioned previously, from a 3D graphics perspective, travel can be seen as camera movement and has been widely studied in the early years of 3D graphics (see also Section 2.3.1). However, most of these studies investigated 3D input devices that are still not affordable for most users. The control of the virtual camera in 3D environments requires at least six DOF which can be directly controlled by means of 6-DOF input devices using established metaphors. *Grabbing-the-Air* and the *Camera-in-Hand* technique are based on mid-air hand gestures. Since mid-air gestures are already established as input for commodity devices, such as Wiimote and Kinect, we used these metaphors to navigate the 3D scene. In 3DUI research, various input devices have been built that support up to 6-DOF input. Some of them are also of relevance for mobile input because recent mobile devices can be seen as surrogates for some of those customized 3D input devices. Thus, as already motivated in the previous chapter, we are mainly interested in the use of commodity hardware, i.e. mobile devices and depth cameras. Among others, Decle and Hachet [59] as well as Kratz and Rohs [114] investigated mobile 3D interaction and evaluated it in a promising rotation task. We evaluated our navigation techniques in an extended version of Kratz and Rohs' [114] approach.

In this chapter we investigate affordable consumer tracking devices that allow 6-DOF input. The main goal is to use controllers which can be easily understood by novices and should be as natural to use as possible with minimal need for assistance or manuals. In order to achieve this goal we propose four 3D travel techniques that rely on well-established interaction metaphors and make use of commodity devices as spatial input (i.e. a mobile device and depth tracking). We evaluate these techniques in a virtual search task experiment on a large stereoscopic 3D display. The study gives insights on user preferences for gestural and mobile 3D input. Our results show that the physical travel techniques outperformed the virtual techniques with respect to task performance time and error rate. These findings are also supported by subjective user feedback. Contributions of this chapter have been published in [49, 54, 179]

6.1. Mobile and Gestural 3D Travel

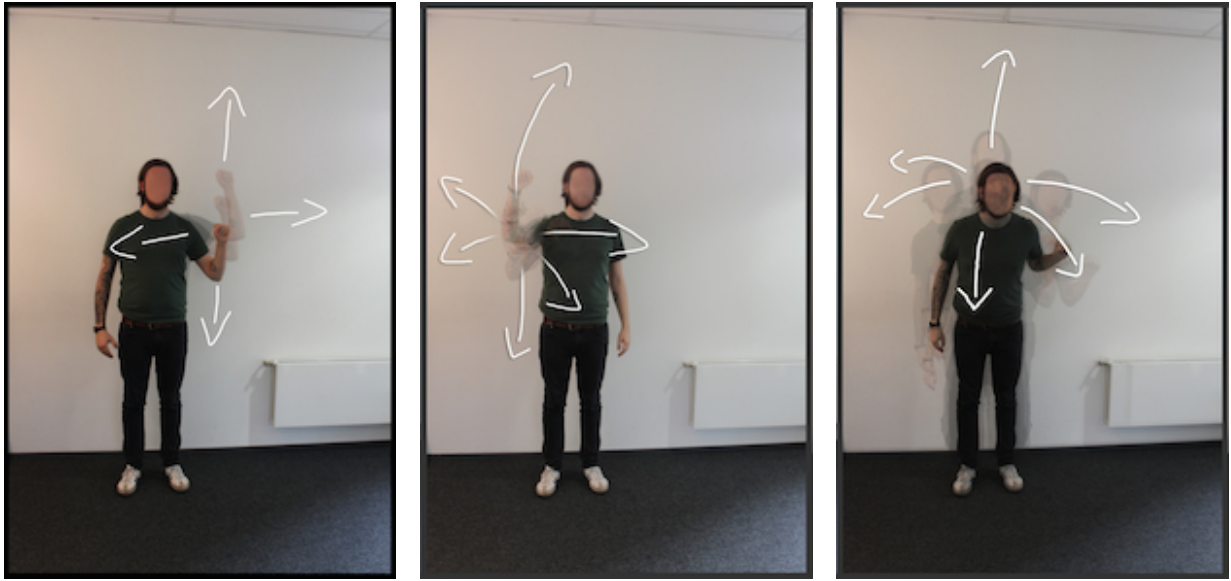
Travel tasks are one of the most fundamental human tasks in our physical environment as well as universal interaction tasks in 3DUIs. The travel task is a secondary task and should not distract the user from the primary task. Thus the travel task must be performed unobtrusively, intuitively and in a way that is easily controllable. While developing the travel techniques, we attempted to keep each metaphor as simple and natural as possible, although we also wanted to study the differences, advantages and drawbacks of each in its form.

In this section we discuss classic travel techniques and their potential use for mobile and gestural interaction with commodity devices such as mobile phones or depth cameras. First, we present two physical techniques using a depth camera, where the user physically changes the viewport with her whole body and both of her hands. Second, two virtual travel techniques are introduced, where the user controls the virtual camera using a mobile device. Moreover, all travel techniques are active, i.e. the viewport movement and orientation are directly controlled. In the following we describe the interaction metaphors regarding their classification and relate them to the corresponding classic travel techniques. Manual manipulation metaphors are designed to manipulate the viewport, instead of virtual objects (e.g. *HOMER* [18], *Go-Go* [160], etc.). The mobile travel techniques rely on gyroscope sensor data and touch inputs, which are provided by all common state-of-the-art touch-enabled smartphones. These virtual input techniques are also based on well-known smartphone gestures and metaphors. Besides the tilt metaphor, we chose the panning gesture, which is a standard direct interaction metaphor with two fingers for scrolling and one finger for navigating in mobile applications.

6.1.1. Bimanual Grabbing

The physical input technique *Bimanual Grabbing* is a manual manipulation technique. In particular the *Grabbing-the-Air* travel technique [21] treats the entire world as an object to be manipulated. The user performs a grabbing gesture to initiate the travel interaction and moves her hand in order to move the viewport. Although this kind of interaction requires a lot of arm motion, it is intuitive to use. The human hand is a remarkable device which is very useful for manipulating physical objects quickly, precisely and with little conscious attention. Therefore we have chosen this technique, which combines the *Camera-in-Hand* technique (see Figure 6.1a) and the *Grabbing-the-Air* technique (see Figure 6.1b). *Camera-in-Hand* pans with the DH to orientate (*yaw, pitch, roll*) the scene viewport. *Grabbing-the-Air* is performed with the NDH for translation (x, y, z) of the viewport. In this scenario each hand controls 3-DOF, respectively, resulting in a simultaneous 6-DOF input. A grabbing gesture toggles the interaction stream. The framework that specifies the *Bimanual Grabbing* technique is depicted in Table 6.1.

The bimanual interaction approach allows an intuitive and flexible control, i.e. the user can look around while moving the viewport. The high sensitivity of the tracking device can increase the user's effort for precise movements in small areas in contrast to travel



(a) Camera-in-Hand metaphor. (b) Grabbing-the-Air metaphor. (c) Whole-body tilt metaphor.

Figure 6.1.: Physical travel techniques.

larger distances. However, using this input technique might result in high physical demand because both hands need to be held in the air while performing the interaction.

6.1.2. Whole-Body Tilt and Grab

The design of the *Whole-Body Tilt and Grab* technique is inspired by the control of a Segway vehicle. Leaning and bending the head and torso in combination with a grab gesture of the NDH results in moving the viewport continuously in the desired direction. This can be seen as a variant of the semi-automated steering technique. The movement

Device/Modality	IP	Travel					
		translation			rotation		
		x	y	z	r_1	r_2	r_3
Gesture Tracker	Grab						
	Pan						

Table 6.1.: Framework for Bimanual Grabbing.

6. 3D Travel

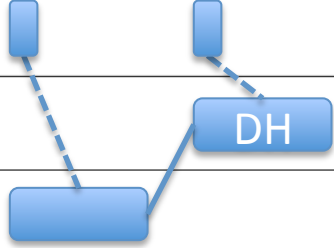
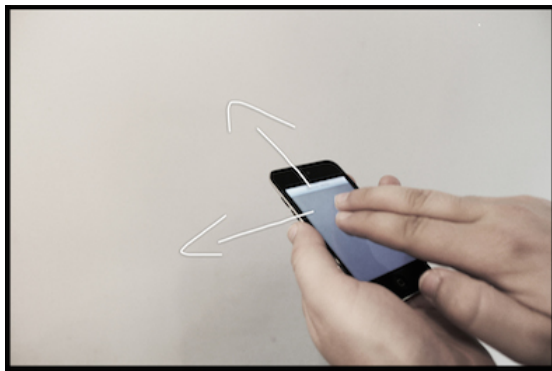
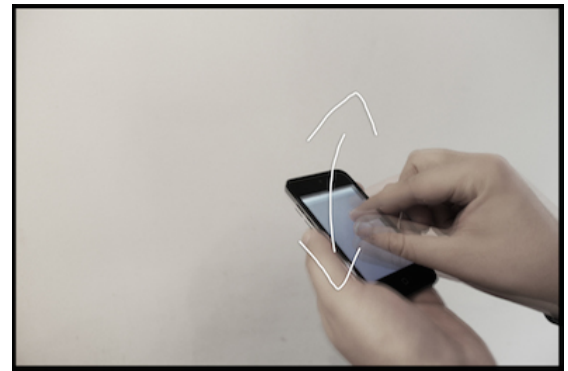
Device/ Modality	IP	Travel					
Gesture Tracker	Grab	translation			rotation		
		x	y	z	r_1	r_2	r_3
							
	Pan						
	Body Tilt						

Table 6.2.: Framework for Whole-Body Tilt and Grab.



(a) Panning gesture with two fingers.

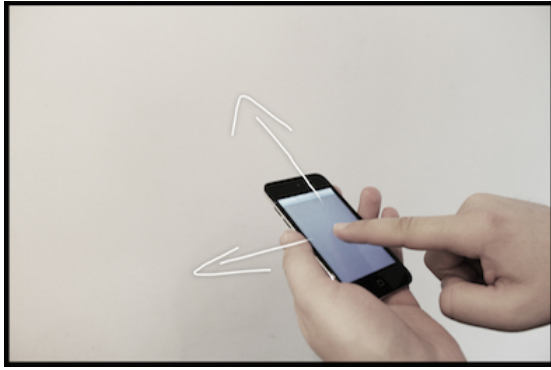


(b) Pinch gesture.

Figure 6.2.: Mobile gestures for virtual camera positioning.

speed can be controlled by the leaning angle (see Figure 6.1c). The user controls the viewport's orientation the same way as in the previously mentioned *Bimanual Grabbing* technique, again based on the *Camera-in-Hand* technique by panning with the DH (see Figure 6.1a). Because of the continuous rate-controlled movement, the user can focus on viewport orientation control. This results in an integrated 6-DOF input technique. The *Whole-Body Tilt and Grab* technique is specified by the framework in Table 6.2.

Furthermore, this input modality is less sensitive than the *Bimanual Grabbing* technique, because tracking head motions is very easy to implement and less error-prone than tracking hand motions. Although this technique has a very high physical demand due to the leaning and bending motions (particularly compared with the mobile variants), here the user can travel longer distances without high effort.



(a) Panning with one finger.



(b) Rotating with two fingers.

Figure 6.3.: Mobile gestures for virtual camera orientation.

6.1.3. Mobile Multi-Touch

The *Mobile Multi-Touch* technique is a combination of multi-touch gestures with two fingers for movement and one finger for camera orientation. The user moves the viewport on its horizontal and vertical axis by panning on the touchable surface of the mobile device with two fingers and using the pinch gesture to move forward and backward (see Figure 6.2). In order to control the three-dimensional camera orientation, the user can pan with one finger in combination with the rotate gesture (see Figure 6.3). In summary, the 6-DOF are composed by the mobile device panning gestures for translation (x,y), the pinch (z) gesture, the rotation (yaw, pitch) and the rotate (roll) gesture. The framework in Table 6.3 specifies *Mobile Multi-Touch*.

6.1.4. Mobile Tilt and Touch

The camera orientation control in the *Mobile Tilt and Touch* technique is analogous to the orientation control in the *Mobile Multi-Touch* technique. But in contrast, the movement is controlled by using the three DOF of the gyroscope. Tilting the mobile device is comparable to the leaning and bending of the *Whole-Body Tilt and Grab* technique. The interaction is toggled by the user's finger touch on the screen of the mobile device. In summary, the user touches the screen of the smartphone with her finger and tilts the device in order to set the move direction of the viewport. The movement is stopped by releasing the finger from the surface. The movement speed can be directly controlled by changing the tilting angle of the smartphone (see Figure 6.4). The framework for *Mobile Tilt and Touch* is depicted in Table 6.4.

6. 3D Travel





Device/Modality	IP	Travel					
		translation			rotation		
		x	y	z	r_1	r_2	r_3
Multi-touch	1-Finger Pan						
	Rotate						
	2-Finger Pan						
	Pinch						

Table 6.3.: Framework for Mobile Multi-Touch.

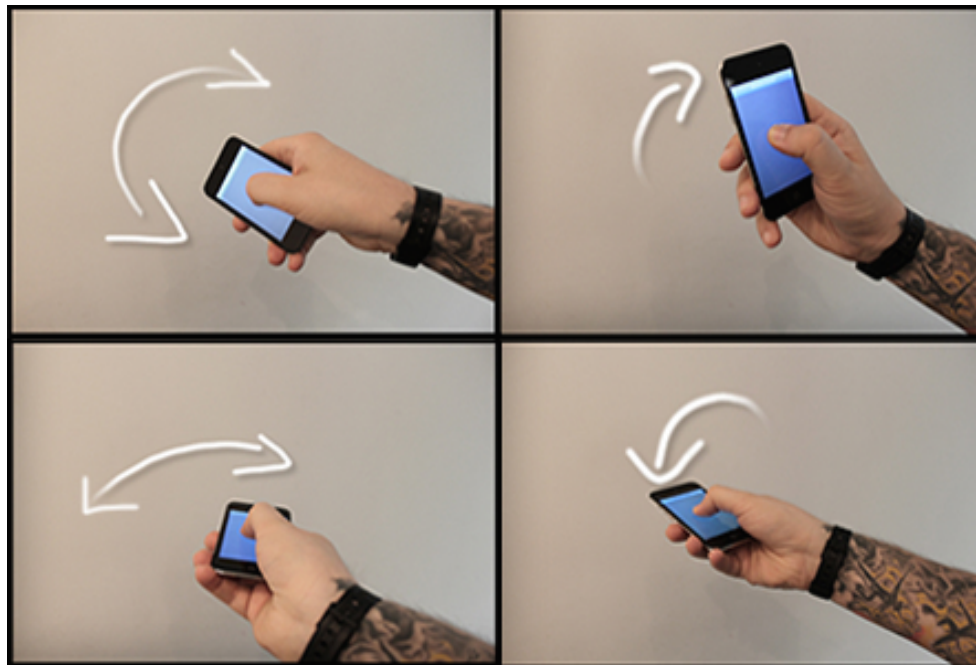


Figure 6.4.: Mobile tilt gestures for virtual camera positioning.





Device/Modality	IP	Travel					
		translation			rotation		
		x	y	z	r_1	r_2	r_3
Multi-touch	2-Finger Pan						
	Pinch						
Touch and Orientation	Tap						
	Rotate						

Table 6.4.: Framework for Mobile Tilt and Touch.

6.2. Experiment

We conducted an experimental study in order to evaluate our techniques in a 3D search task on a large 3D display. We designed a parametrized search task that provides an easy to control yet flexible experimental setup. As quantitative metrics, the elapsed time for completing a trial (task completion time) and percentage of correct reported textured faces for each task (error rate) were measured. In addition, subjective feedback was gathered using the NASA TLX [87] rating scale.

6.2.1. Participants

Ten participants (9 male, 1 female) from the university environment volunteered in the user study. The participants were between 20 and 35 years of age ($M = 27.8, SD = 4.13$). They all owned a touch-enabled smartphone, but had at most limited experience with depth cameras. Only three of the participants had prior experience with stereoscopic 3D applications and 3D user interfaces. According to the participants' self-assessment on a 10-point scale (1 = no experience, 10 = highly experienced), their experience level for 3D modeling ranged from 1 to 9 ($M = 5.4, SD = 2.55$). Furthermore, their experience in computer graphics ranged from 2 to 9 ($M = 5.9, SD = 2.47$) and in computer science from 4 to 10 ($M = 7.9, SD = 2.13$). The participants received no monetary compensation for their participation in the study.

6.2.2. Conditions

The four proposed interaction methods (*Bimanual Grabbing*, *Whole-Body Tilt and Grab*, *Mobile Multi-Touch* and *Mobile Tilt and Touch*) were compared with each other. According to the interaction technique, either a mobile device (virtual input condition: *Mobile*

6. 3D Travel

Multi-Touch and *Mobile Tilt and Touch*) or whole-body gestures (physical input condition: *Bimanual Grabbing*, *Whole-Body Tilt and Grab*) were used as input. Additional conditions were grid size (small vs. big) as well as textured face count (easy vs. difficult). The textured face count was randomly chosen in a ± 1 range around 3 (easy) and 7 (difficult) textured faces in order to prevent the participants from inferring the correct number of textured faces.

6.2.3. Task

While Kratz and Rohs [114] investigated 3D object rotation on a mobile device using a front and rear touch virtual trackball as well as tilt, we extended their experimental setup for travel tasks. According to previous work, each of the four faces of one tetrahedron object was colored in a distinct color to allow the participants to remember the sides of the objects and help them orient themselves in the scene.

Our approach provides a good control of the experiment conditions and thus even allows a reasonable way to compare future travel techniques. The experimenter was able to change each parameter (e.g. grid size or number of textured object faces) during the experiment remotely, as well as starting and stopping the trials. Therefore, the objects were not randomly chosen and the number of objects was defined programmatically, which enabled the experimenter to parametrize precisely the characteristics of the experiment.

Each travel task started with an exploration task followed by a search task. We chose an introductory exploration task without an explicit goal for movement in order to browse the environment and obtain information about objects, help the user to get oriented in the virtual world and build up spatial knowledge. Besides, in this training phase the user was able to get familiar with the travel techniques. Afterwards, the user was asked to perform the actual travel task, or more specifically a primed search task.

6.2.4. Design

The experiment had a $4 \times 2 \times 2$ within-subjects factorial design. Factors were interaction technique for navigation control (*Bimanual Grabbing*, *Whole-Body Tilt and Grab*, *Mobile Multi-Touch* and *Mobile Tilt and Touch*), grid size (small: $2 \times 2 \times 2$ and big: $3 \times 3 \times 3$) as well as textured face count (easy: 3 ± 1 and difficult: 7 ± 1). The textured face count was randomly chosen in a ± 1 range around 3 and 7 in order to prevent the participants from inferring the correct number of textured faces and forcing them to really count all textured faces in the scene presented to them.

The order of our four chosen input techniques was counterbalanced according to a Latin Square design, as was the order of grid size and textured face count settings. All trials for each input technique were conducted in sequences followed by a short break of two minutes before starting a new trial sequence. Each setting resulted in a total of $10 \times 4 \times 2 \times 2 = 160$ trials.

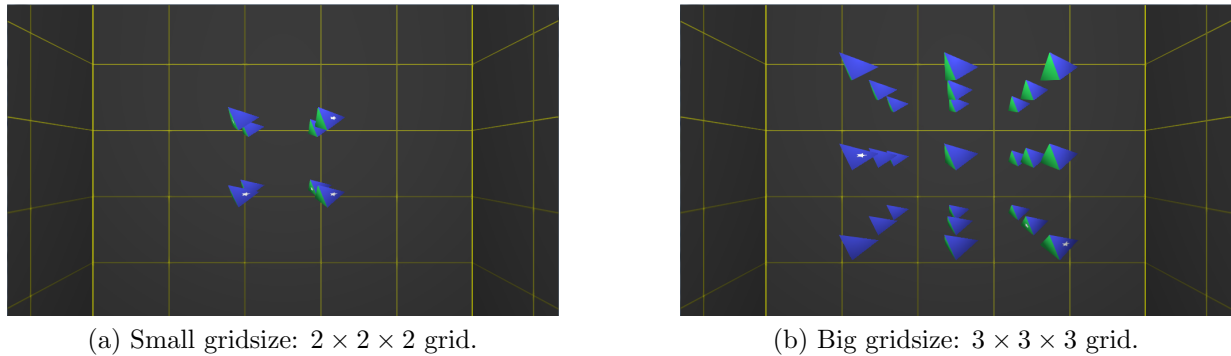


Figure 6.5.: The user's perspective on the 3D scene consisting of grids of tetrahedrons and textured faces.

6.2.5. Procedure

After the participants filled out a short questionnaire to gather demographic details (age, gender, etc.) and information about their level of experience with 3D graphics and computer science, they were placed in front of a 5×3 -meter projection wall at a distance of 2.5 meters during the trials. The freely navigable scene comprised a regular three-dimensional grid of tetrahedrons where the goal of each trial was to count the number of object faces textured with a white star logo (see Figure 6.5).

Each task started with an exploration task in a scene with a $3 \times 3 \times 1$ grid of tetrahedrons without any textured faces in order to let the participants focus on the exploration of the VE and acclimatization with the input technique. The task was explained to the participants, but only minimal instruction was given in how to use the input devices and thus perform the interaction techniques. Then they could try the techniques by maneuvering through the scene to get familiar with the device.

After the participants felt comfortable with the training task and decided to start the task, the experimenter initiated a trial by determining grid size and number of textured faces corresponding to the Latin Square design. A trial was completed after the participant reported the number of textured object faces found to the experimenter. Finally, the trial completion time and the number of textured faces found were recorded for each trial. After each sequence of trials for each input technique, the participants were asked to subjectively rate the workload of the just-finished input technique using the NASA TLX [87] rating scale.

6.2.6. Apparatus

The same apparatus was used for this experiment as the apparatus of the 3D manipulation study (see Section 5.2.6).

6. 3D Travel

6.2.7. Independent and Dependent Variables

Interaction technique, grid size and number of textured faces were treated as independent variables while task completion time and error rate were measured as dependent variables.

6.2.8. Hypotheses

Although mid-air gestural input is expected to require higher physical demand, it enables the user to intuitively and freely navigate in 3D space. We thus expect a good task performance for mid-air gestures. Therefore the hypotheses to be verified by the user study are:

- The task completion time of mid-air gestures outperforms mobile input (H1).
- The complexity of the scene (small grid vs. big grid) affects the task performance time (H2).
- The physical techniques require a higher workload than the virtual techniques (H3).

6.2.9. Improvement to Existing Methodology

The experimental setup is well suited to evaluate three-dimensional travel techniques. In order to find each textured face, the subjects need to look at all faces of each object. In order to solve this search task, they need to move and change their orientation to appropriate viewpoints. As already mentioned in the task description above, we investigated a 3D travel task instead of a rotation task. Thus, we extended the setup with a three-dimensional grid of tetrahedron objects. This allows a free and easy navigation control within a reasonable testbed environment. In previous work, colored faces were introduced to aid in orientation. We extended this approach by adding light at the top of the scene and placing the tetrahedron grid in the center of a virtual cube with grid pattern textures on the inner walls. This extension of the setup was meant to amplify the user's immersion.

The virtual objects were not randomly chosen and the number of objects was defined programmatically. This allowed a precise control of the characteristics of the experiment. This approach therefore provides better control of the experimental conditions than related work, and it may even allow a reasonable way to compare future travel techniques.

6.3. Results

In the following we present the results of the experiment with respect to interaction technique (*Bimanual Grabbing*, *Whole-Body Tilt and Grab*, *Mobile Multi-Touch* and *Mobile Tilt and Touch*), grid size (small or big) and number of textured faces (easy or complex) for the task completion time and error rate. Additional subjective feedback from a NASA

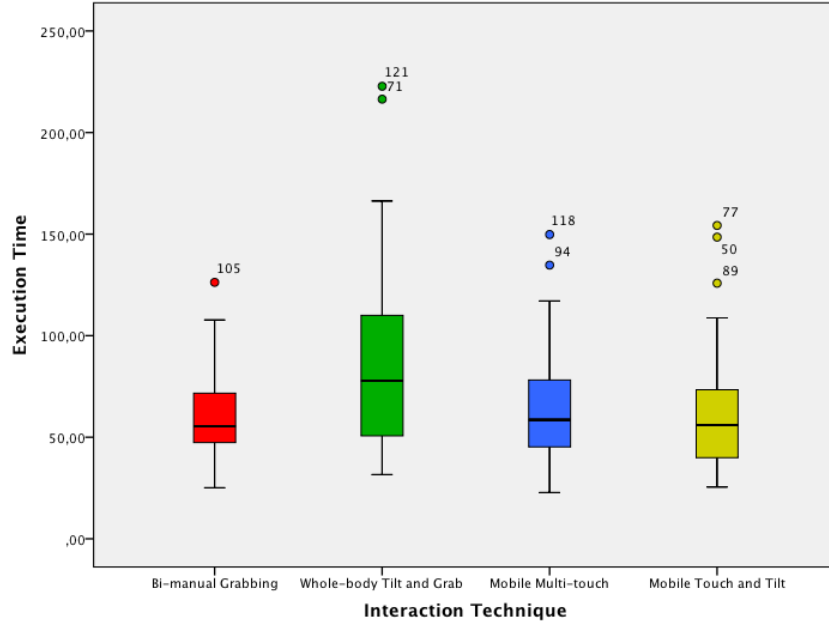


Figure 6.6.: Boxplots of the execution time in seconds w.r.t. interaction technique.

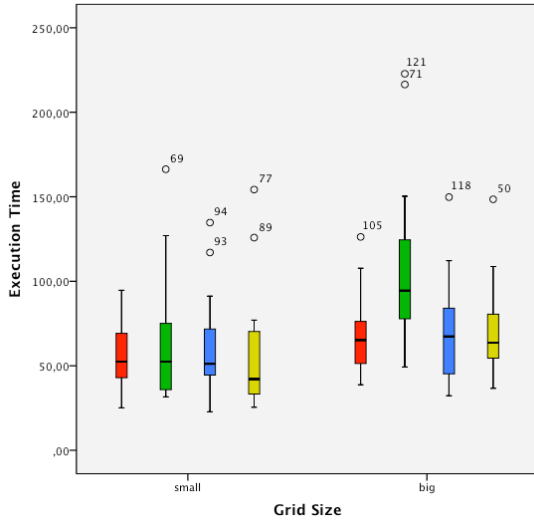
TLX test is also reported. All figures use the same color scheme for the interaction techniques (*Bimanual Grabbing*: red; *Whole-Body Tilt and Grab*: green; *Mobile Multi-Touch*: blue; *Mobile Tilt and Touch*: yellow).

6.3.1. Task completion time

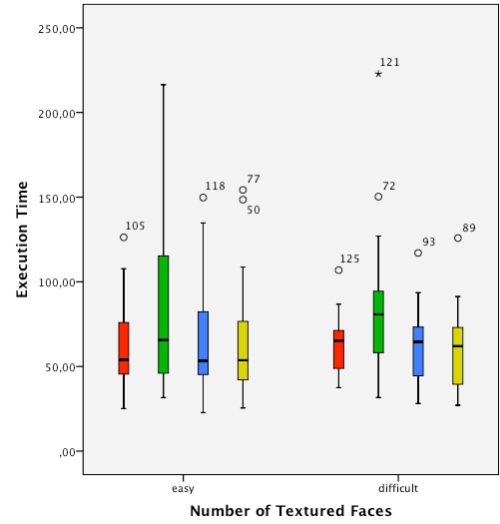
The results for task completion time are shown in Figure 6.6 and Figure 6.7. The mean completion time for *Bimanual Grabbing* was 62.24s, $SD = 22.05s$, for *Whole-Body Tilt and Grab* 85.08s, $SD = 46.98s$, for *Mobile Multi-Touch* 66.88, $SD = 28.46s$ and for *Mobile Tilt and Touch* 62.94, $SD = 30.94s$. The mean completion time regarding grid size was 58.70s, $SD = 29.25s$ for the small grid and 78.80s, $SD = 36.39s$ for the big grid. The mean completion time regarding the number of textured faces was 69.35s, $SD = 38.20s$ for easy and 68.16, $SD = 30.40s$ for difficult.

The univariate ANOVA test shows a significant effect on the task completion time depending on interaction technique ($F_{3,40} = 4.663$, $p < 0.05$) and grid size ($F_{1,80} = 15.729$, $p < 0.05$), but no significant effect from the number of textured faces ($F_{0,80} = .055$, $p = 0.815$). A Bonferroni pairwise comparison of interaction techniques shows a significant difference for *Bimanual Grabbing* vs. *Whole-Body Tilt and Grab*, *Whole-Body Tilt and Grab* vs. *Mobile Multi-Touch* and *Whole-Body Tilt and Grab* vs. *Mobile Tilt and Touch* ($p < 0.05$), but no significant difference could be found between the other techniques.

6. 3D Travel



(a) Execution time w.r.t. grid size.



(b) Execution time w.r.t. textured face count.

Figure 6.7.: The boxplots of execution time grouped by interaction technique.

6.3.2. Error rate

The error rate, i.e. the ratio of the number of incorrect responses to the total number of responses with respect to input method, was 20.0% for *Bimanual Grabbing*, 17.5% for *Whole-Body Tilt and Grab*, 25.0% for *Mobile Multi-Touch* and 17.5% for *Mobile Tilt and Touch*. In order to measure neutral error performance, the participants were not provided with feedback on whether they counted the right number of textured surfaces or not.

The responses for the physical techniques were closer to the actual numbers than the responses for the virtual techniques. This is reflected by the mean square error, i.e. the deviation of the reported count from the actual count with 0.2 for *Bimanual Grabbing*, 0.175 for *Whole-Body Tilt and Grab*, 0.475 for *Mobile Multi-Touch* and 0.4 for *Mobile Tilt and Touch*.

6.3.3. NASA TLX

The NASA TLX provided the following subjective results on the four input methods. Figure 6.8 refers to results of the six NASA TLX sub-scales with respect to the four interaction techniques: Mental Demand (MD), Physical Demand (PD), Temporal Demand (TD), Performance (OP), Effort and Frustration (FR). The average overall workload of each interaction technique is 9.55 ($SD = 2.75$) for *Bimanual Grabbing*, 11.83 ($SD = 2.85$) for *Whole-Body Tilt and Grab*, 6.46 ($SD = 3.99$) for *Mobile Multi-Touch* and 9.00 ($SD = 3.28$) for *Mobile Tilt and Touch* (see Figure 6.9). In conclusion, both physical input methods resulted in higher physical demand and effort but less frustration and temporal demand,

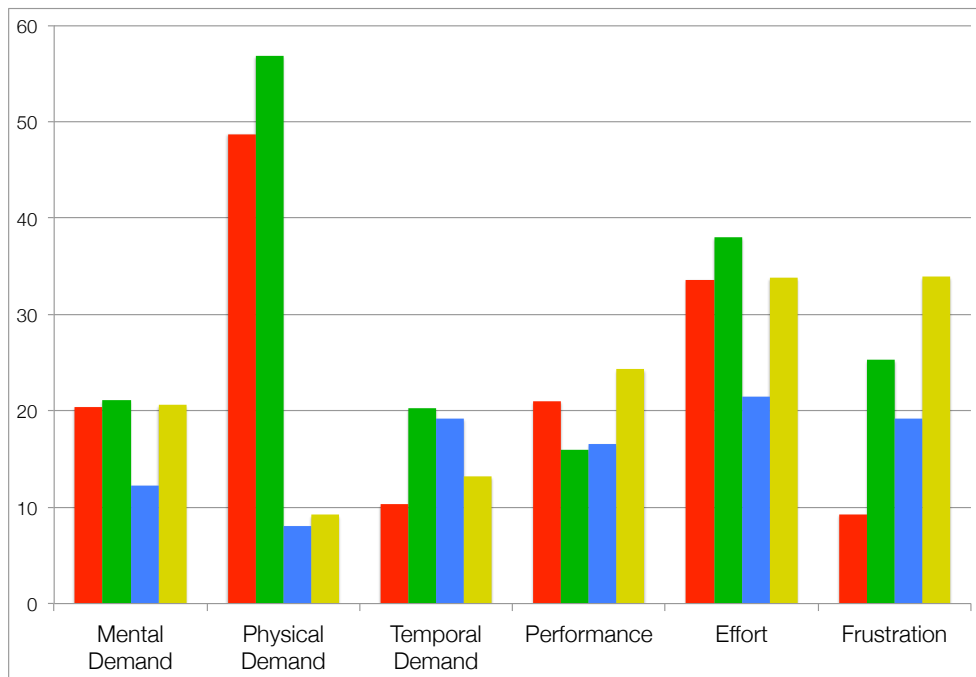


Figure 6.8.: Participants' averages concerning the four interaction techniques from NASA TLX.

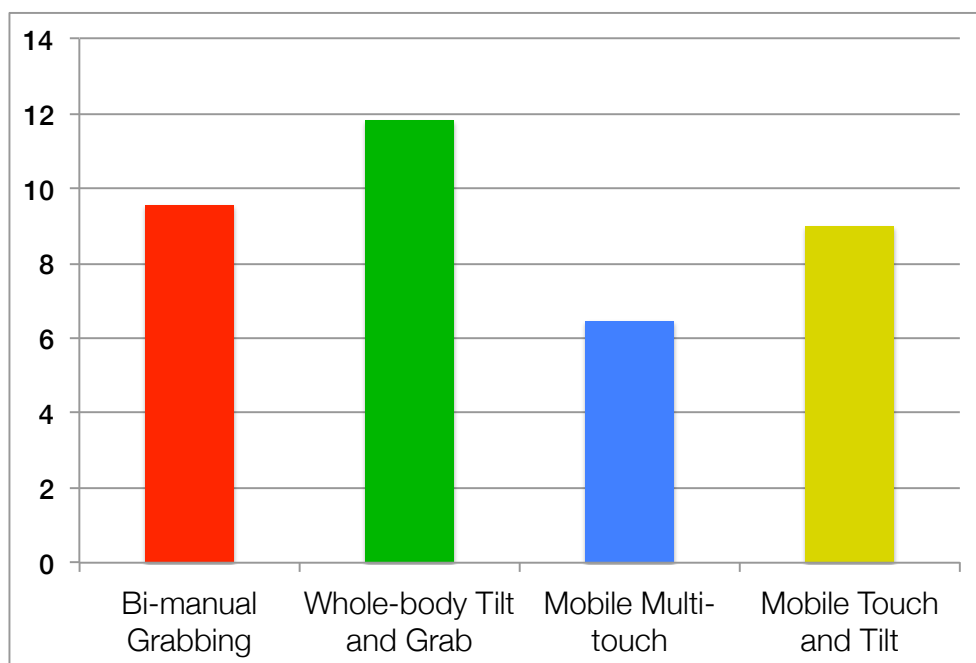


Figure 6.9.: Overall workload of all four interaction techniques from NASA TLX.

6. 3D Travel

while the virtual input methods could be performed with a very low physical demand and less effort but a high temporal demand and much more frustration.

6.4. Discussion

In this chapter, four input techniques were investigated. In the following the results will be discussed with a detailed consideration of the advantages and limitations of the techniques.

6.4.1. Experimental results

Bimanual Grabbing outperformed all other methods in task completion time and error rate. Regarding the remaining techniques, *Mobile Multi-Touch* performed well in task completion time but worst with respect to error rate, and *Whole-Body Tilt and Grab* has a low error rate (lowest mean square error) but bad task completion time, while *Mobile Tilt and Touch* was average for both metrics. In summary, those mobile interaction techniques that had an average performance on error rate evidently performed worse on mean square error. These high mean square errors lead to the conclusion that the physical techniques are superior compared to virtual techniques.

The results further show that the interaction technique and the size of the grid significantly affected the task completion time. This leads to the conclusion that outperforming interaction techniques also perform better when increasing the complexity of the scene (H2). The fact that the interaction techniques were not significantly affected by the number of textured faces indicates that the cognitive load of all navigation techniques is almost the same for all techniques. Regarding task performance time, all travel techniques are suited as secondary tasks since there is no significant effect regarding the number of textured faces, i.e. the primary search task.

The quantitative results of the experiment corresponded with the subjective results of the NASA TLX. As expected, the mid-air gestural input methods resulted in higher physical demand and effort, which is already a well-known issue. Nevertheless, the good task performance time of these input methods explained the corresponding TLX sub-scales (small temporal demand and low frustration level) very well. Overall the virtual techniques resulted in a lower workload (H3). The *Mobile Multi-Touch* technique performed best with respect to the overall workload. This might be due to the fact that this 3D travel technique is closely related to well-established direct touch interaction metaphors. Altogether, the task performance time and error rate of both mobile input methods were worse. This was also clearly reflected by the NASA TLX test that revealed low physical demand and effort but a high temporal demand and more frustration.

6.4.2. Advantages

One of the most important advantages of our approach is ease of use. Our goal was to keep the interaction very simple, intuitive and direct by guaranteeing that all parts in the 3D

scene are easily reachable. In order to keep the frustration level at a minimum, the users had free navigation control, i.e. users were free to travel without limitations in space and were able to control the movement speed. All input data was generic because we used the sensors of a depth camera and a mobile device instead of a joystick, mouse or keyboard. Furthermore, the interaction techniques were cognitively friendly. This is because of the continuous movement, i.e. the whole environment can be reached from the current position to the desired location, while the speed of camera movement gave an indication of the distance traveled.

The NASA TLX revealed that both physical input methods resulted in higher physical demand and effort but less frustration and temporal demand, while the virtual input methods could be performed with a very low physical demand and less effort but a high temporal demand and much more frustration. So even the physical methods perform well at least in some TLX subscales that make them worth taking into consideration for scenarios that focus on playful applications or time-critical tasks. To sum up, we were able to design natural interactions, which also proves our general design goal correct.

6.4.3. Limitations

Unfortunately, the accelerometer sensor of the mobile device is not usable due to errors based on the double integration of accelerometer readings [192]. However, it might be, for example, very usable for a shaking metaphor (as used in Section 9.1). Thus, we did not use this sensor in our input techniques. Although the Microsoft Kinect device works very well in normal-ambient and dark rooms, it is problematic and error-prone in places that are too bright. This ambient light dependence of the depth camera is a clear limitation. We solved that problem by performing the experiment in a darkened room to ensure constant lighting conditions over the course of the whole experiment. Finally, there is a general problem of power management using mobile sensors and wireless network traffic because they run down batteries quickly. Therefore we limited each task to a certain number of trials to avert the danger of power management complications.

6.5. Conclusion

The results of the study in this chapter give implications for the design of intuitive 3D navigation techniques that enable VR in the living room or in public places. One potential scenario is, for example, 3D gaming. The physical interaction techniques are very suitable candidates for such a scenario. The general drawback of physical demand and effort might even increase the complexity and thus the gaming experience. Another potential scenario in the living room is a 3DUI for a movie (or music) database.

The physical techniques might be inappropriate for exploring a movie database. On the other hand, the remote control metaphor is a well-known concept, and thus the mobile interaction techniques might be better for this kind of application. However, an appropriate mobile interaction technique needs to be carefully designed.

Part III.

Beyond the Multi-touch Surface

7. Interaction Context for Multi-touch 3D Interaction

Psychological research on the *Reach to Grasp* task has shown that the pre-shaping phase of the human hand allows a prediction of the object a human is going to grab. Multiple studies on the *Reach to Grasp* task have shown evidence that only a few variables have an impact on that prediction (see Chapter 2.3.2). The insights from neuropsychological and robotics research are promising and we believe that the information from the *Reach to Grasp* phase can substantially improve interaction with stereoscopic multi-touch displays. While reaching and pointing tasks have a long tradition in the field of HCI, the hand pre-shaping has rarely been investigated.

One of the main goals of multi-touch 3D interaction is to eliminate the restriction to near-zero parallax (see Chapter 2.3.2). This can be done by extending the interaction space with additional dimensions in space and time. The combination of multi-touch technology, depth cameras and stereoscopic displays promises interesting and novel user interfaces. However, the benefits, possibilities and limitations of using this combination have not been examined in depth and are so far not well understood [182].

In this chapter the concept of *interaction context for touch-enabled interactive surfaces* will be introduced. Several examples of this concept will then be discussed in more detail. In particular we investigated the *Reach to Grasp* phase in more depth in order to use it as interaction context. We performed experiments in which we analyzed hand postures above and on the interactive surface. These experiments aimed to examine whether the hand posture allows an early prediction of the objects the user is intending to interact with or not.

7.1. Interaction Context

While head tracking has a long history in subtle interaction (i.e. head tracked stereoscopic displays stimulate the motion parallax cue) other information about the user that can be used to inform the UI before the actual interaction happens have scarcely been investigated. The concept of interaction context goes one step further, and it is expected to inform the UI with additional information that can be instantly used in the user interface feedback loop. In our approach, the feedback loop is extended by another channel that adds context information to the loop (see Figure 7.1). Thus, the UI is affected by both explicit and implicit user input. User feedback is again returned to both loops, affecting

7. Interaction Context for Multi-touch 3D Interaction

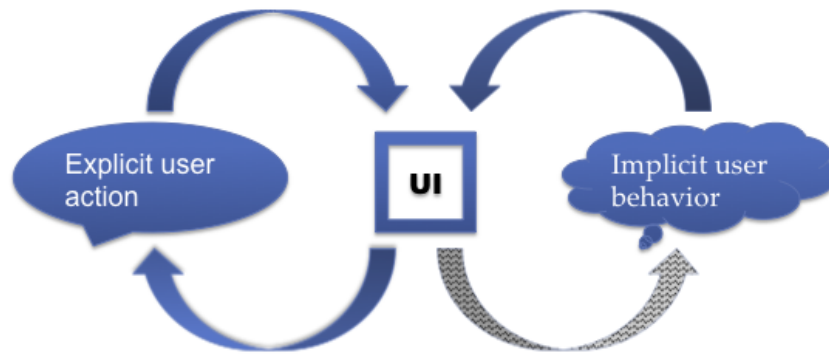


Figure 7.1.: Interaction context in the user feedback loop: the UI is affected by both explicit and implicit user input.

explicit interactions (e.g. moving a pointer to select or manipulate an object) as well as implicit user behavior (e.g. focusing on an object in order to manipulate it).

7.2. Modalities

In this section we briefly discuss potential candidates of modalities that can be used as interaction context. Context information of interest such as the position, orientation and posture of the user's head, the hands, fingers, etc. (see Figure 7.2) will also be discussed. The in the following chapter, grasp, or more specifically *Reach to Grasp*, is investigated in more depth as it radically exploits the concept of interaction context for 3DUI.

7.2.1. Whole-Body Postures

The position, orientation and pose of the whole body provides a variety of interaction context information. The position and orientation indicate the area of interest, i.e. the area on the interactive surface that the user is oriented towards and that is within the user's reach.

The whole-body part of this approach integrates well into the concept of interaction context. In tracked VEs various context information is tracked, and some is already used as input. For example, head-tracked stereoscopic setups currently make use of interaction context to adapt the scene to the user's head position and orientation. Some 3D interaction techniques exist as well that depend on the user's position and orientation and/or adapt the technique accordingly (e.g. extending the reach of the virtual hand in the Go-Go selection techniques).

7.2.2. Hand Postures and Grasp

In recent touch interactive systems the information about the location of the human fingers or hand is only tracked on the actual interactive surface. Since the interaction actually starts earlier, the UI misses a lot of information. One of the most obvious problems is that, in comparison to the mouse, almost no touch-based UI supports hover. But beyond the hover zone, the whole space above the interactive surface is not covered. However, this area is an important space for gathering interaction context that can be used to inform the current interaction. The interactive surface that restricts this space can be seen as the border between subtle interaction (interaction context) and actual interaction.

Hand postures and grasp are intended to be used as interaction context in order to enrich multi-touch input on interactive surfaces. The knowledge of the hand posture shortly before touching the multi-touch surface can be used to adapt the parallax such that the user can interact with the most probable object directly on the surface (with zero parallax).

7.2.3. Eye Gaze

Another dimension of interaction context is the user's eye gaze. Knowing where the user is looking might also help to deduce which objects are of interest at a certain instance. Further, gaze movements can be used to deduce whether the user is getting lost in an interface (e.g. in a travel task), and instant feedback can be provided. An important piece of information that is provided by eye gaze tracking is the content on the surface that is out of the user's reach (i.e. out of physical reach or displayed with extreme positive parallax).

Whole-body context in combination with gaze information can be used to adapt the UI so that every object in the scene is within the user's reach at all times. However, the navigation and orientation must be carefully guided in such a scenario to keep the user's cognitive load to orient herself in the scene at a minimum.

7.3. Conclusion

In this chapter the concept of *interaction context for touch-enabled interactive surfaces* was introduced. After introducing the general concept, several examples of it were discussed.

There are different potential application domains that can benefit from this UI and interaction concept. The most evident application example for our concept is 3D modeling. A 3DUI that enables direct touch together with interactions above the interactive surface enables direct manipulation for 3D modeling. Another good example for cluttered interfaces is a (virtual) mixer for DJs. Such a tool that emulates a real mixer console can be seen as a customizable expert interface. In such an interface, which consists mainly of buttons, knobs and sliders, the interaction concepts and virtual widgets of our studies can be directly applied. Browsing and interaction in large image databases visualized in 3D might offer other interesting scenarios to investigate.

We envision that the concept of interaction context enables the design of user interfaces that can be dynamically adapted based on predictions of the user's intention. Adaptation

7. Interaction Context for Multi-touch 3D Interaction

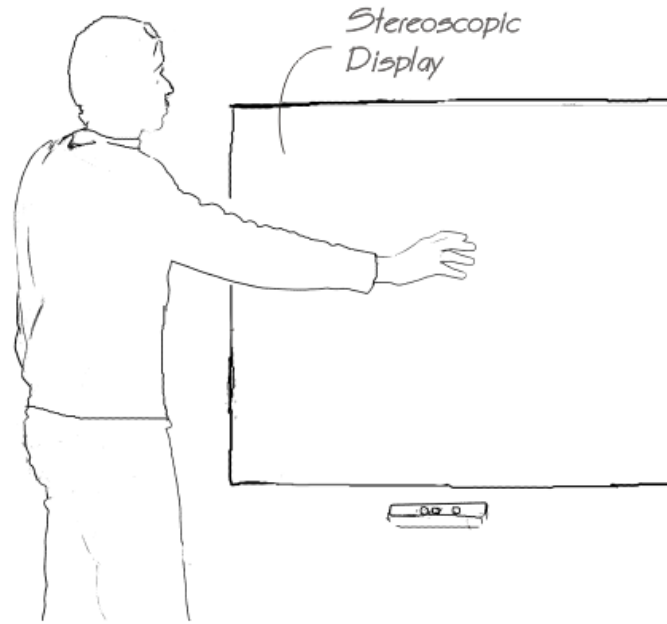


Figure 7.2.: Interaction context in the user feedback loop: position, orientation and posture of the user’s head, hands and fingers.

means that stereoscopically displayed 3D objects serve as virtually graspable objects of their real counterparts, which respond to the user’s grasping behavior before she actually touches the surface. Thus, an immersive interaction experience can be realized by “touching” virtual objects together with haptic feedback through the physical border of the interactive surface.

In the next chapter, the *Reach to Grasp* task will be investigated in more depth. We therefore present an initial study and a corpus acquisition study that analyze hand postures above and on the interactive surface. This results in design considerations that summarize the lessons learned from these studies.

8. Reach to Grasp Interaction

Psychological research on the *Reach to Grasp* task has shown that the pre-shaping phase of the human hand allows a prediction of the object a human is going to grab. Multiple studies on the *Reach to Grasp* task have shown evidence that only a few variables have an impact on that prediction (see Subsection 2.3.2). The insights from neuropsychological and robotics research are promising and we believe that the information from the *Reach to Grasp* phase can substantially improve interaction with stereoscopic multi-touch displays. While reaching and pointing tasks have a long tradition in the field of HCI, the hand pre-shaping has rarely been investigated. However, due to the availability of low-cost algorithms, off-the-shelf commodity hardware and low instrumentation are now sufficient to track the human hand above the interactive surface. Taking knowledge about the time dimension (i.e. the whole interaction phase before, during and after the actual touch interaction) into account has the potential to improve the user interface of stereoscopic multi-touch surfaces (e.g. by snapping desired objects to the touch surface). The contributions of this chapter have been partially published in [55, 197].

In this chapter, the concept of *interaction context for touch-enabled interactive surfaces* will be investigated for the *Reach to Grasp* task in more depth. With the knowledge of the user's intention, the touch-based UI can then be adapted before the user finally reaches the interactive surface (see Figure 8.1). We performed experiments in which we analyzed hand postures above and on the interactive surface. These experiments aimed to examine whether the hand posture allows an early prediction of the objects the user is intending to interact with or not. Finally, design considerations for the *Reach to Grasp* interaction are discussed.

8.1. Grasp Pre-Study

An initial study aimed to investigate how hand postures can improve multi-touch interaction. With this study we took a first step towards the application and evaluation of the concept of interaction context described above (see Chapter 7). In this study the participants had to grab user interface elements (buttons, knobs, etc.) on an interactive surface. We observed the participants' hand shapes and postures during the *Reach to Grasp* phase. The results of the study show first evidence that multi-touch gestures can be detected right before the user reaches the surface and starts to explicitly interact with the multi-touch surface.

8. Reach to Grasp Interaction

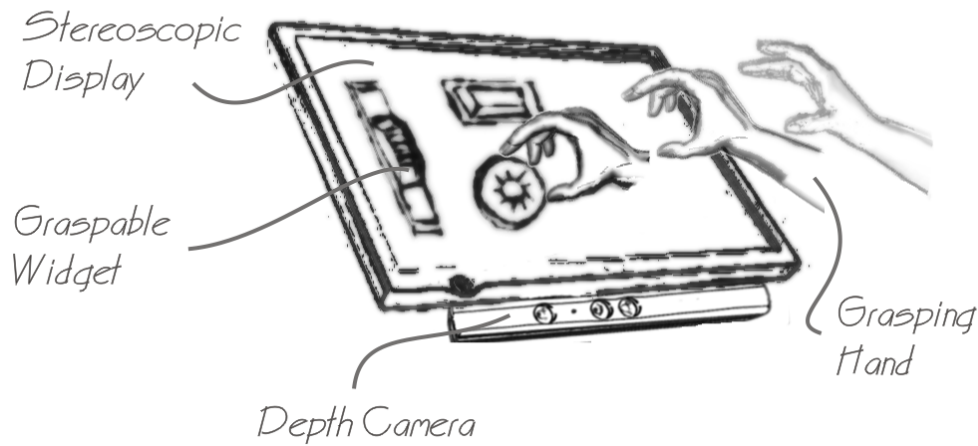


Figure 8.1.: Design concept of a multi-touch enabled stereoscopic surface equipped with additional depth sensors that can predict the user's intention during grasping movements.

8.1.1. Experiment

A observational study was performed where users had to interact with 3D-like UI elements. With this experiment we aimed to get initial insights on the posture of the hand when approaching and interacting with a virtual object via grasp gestures on an interactive surface.

Participants

Six right-handed participants took part in the study (two female, four male). All of them were members of our institute or university students. All of them told us that they had good to excellent experience with touch devices. The study took around 30 minutes per subject. To track the user interaction, we videotaped the subjects while they touch the interactive surface, using a video camera with 25 frames per second.

Task

The participants were required to complete three manipulation tasks: a button pressing task, a knob rotation task and a slider dragging task. Each task was repeated randomly for buttons, knobs and sliders of different sizes. After each step the hand had to be placed back in a resting position on the table before the next element was to be grabbed.



Figure 8.2.: Experimental setup for the study: table that serves as hand rest and a tilted transparent projection screen as interactive surface.

Procedure

After a short introduction, each participant was instructed on how to perform the tasks. The subjects were asked to interact with the virtual objects just as they would use real physical buttons, sliders or knobs. Then they were allowed to play around with a set of samples in order to get used to the interactions. After this training phase, the real test was performed, where the subjects interacted with a second set of UI elements with changing sizes. Afterwards, they were asked to give additional comments on the setup, the interaction technique and potential applications in a semi-structured interview.

Apparatus

For this study a hardware prototype with a transparent projection screen was built (see Figure 8.2). The transparent screen allowed the videotaping of the participants through the surface. altogether the prototype consisted of a transparent screen in a simple wooden case, a camera and projector.

Virtual UI objects (knobs, slider and buttons) were displayed on the projection screen. The participants were sitting in front of the table in an upright position. The *Reach to Grasp* interaction then took place between the table (hand rest) of the prototype and the interactive surface. This guaranteed that all participants had to travel the same way from the table to the surface. Thus, travel time (in frames) and distance could be measured. In addition, speed could also be derived from this.

8. *Reach to Grasp Interaction*

Research question

To adapt the parallax so that the user can interact with objects on the surface (zero parallax), the system needs to know which object the user is going to touch before she reaches the surface. This leads to our working hypothesis on pre-detection of gestures: We are able to recognize the gestures to grasp a specific 3D UI element a certain number of frames before the user explicitly touches the element on the interactive surface. This means that the user's hand pose right before hitting the interactive surface can be used to predict which interface element she is going to touch. We expect to observe that the hand posture changing depending on distance and which object the participant is going to touch.

We further assume that at least the following hand properties can be adopted for this purpose (see Figure 8.3):

- Number of fingers used
- Use of specific finger(s)
- Opening (boundary) between the fingers
- Moving direction of the grasping hand
- Moving speed of the grasping hand

Limitations

To a certain extent the study lacks a real user experience, due to the fact that a low-fidelity prototype was used. Nevertheless, we believe that this early-stage prototype was well suited to reasonably explain the interaction concepts to the participants.

8.1.2. Results

As this is an early study to get an overview of the relevant factors, the focus lies on qualitative feedback. Therefore the captured video was manually inspected and user comments from the questionnaire were analyzed. The first results are promising and all relevant insights are specified in detail in the following.

Reach to Grasp Phase

The first and most important result is that the last frames before impact can be used to predict the gestures to grasp a specific 3D object. The majority of distinguishable hand postures can already be determined at least ten frames before hitting the surface. This means that in the last phase of the hand's approach, its posture does not change any longer. One-finger pointing gestures (press a button, move a slider with a one-finger gesture) seemed to perform best. The detection of gestures with multiple fingers tended to perform a little less well, but if such a posture is unambiguously detected the properties (finger count, finger boundary, etc.) allow a good prediction of which object might be grabbed.



Figure 8.3.: Available candidates for hand properties: Number of fingers; use of specific finger(s); opening (boundary) between the fingers; moving direction of the grasping hand; moving speed of the grasping hand.

Hand Aperture

The number of fingers used is an important indicator to predict the intended object. As assumed, the number of fingers that are used to grab the different objects is important. There was less variation between the subjects regarding different grasp postures for one specific object. At least after some trials, the subjects tended to use a specific number of fingers depending on the object they were about to touch. Thus, there seemed to be a training effect. If finger count is combined with additional properties it should be possible to build a robust predictor for multi-touch gestures through hand posture tracking in the short-term interaction history of the *Reach to Grasp* phase.

User Dependence

As a third result, it can be stated that the task was highly user dependent. Further, to some extent the object size also influenced the grasp posture. It could be observed that there was little variance in one subject's postures for different objects. The subjects tended to use the same finger(s) for similar tasks (e.g. pressing a button with either the index finger or the middle finger). However, some subjects varied the number of fingers depending on the size of the objects. This also needs to be investigated in more detail in a further more extensive study.

8. *Reach to Grasp Interaction*

Subjective Feedback

After the subjects finished the tasks, they had the opportunity to give additional feedback in a semi-structured interview on the setup, the interaction technique and potential applications. The participants understood the general idea of the interaction concept and acknowledged the novelty of the interaction. This was well reflected by their comments on the prototype and interactions. For example, two participants stated that they manipulated the more complex knob hesitantly because they thought that this object needed more force to be manipulated. They wanted some kind of (haptic) feedback to better cope with that. Thus, affordance [76, 152] should also be taken into account in future research when investigating gestural and grasp interaction with virtual objects. In a graspable UI, one could not only support the discoverability of possible (physical) actions, but actively guide the user's actions, or more precisely, their grasp interactions. In addition, the participants gave important feedback on the design for further studies by proposing specific tasks and application scenarios. According to them, one important task could be menu control: "The use of such [graspable] 3D menu controls is closer to their physical counterpart than the 2D menu items". The participants proposed UIs for control rooms and tools for musicians as potential applications for this concept.

8.2. Grasp Corpus Study

To collect a corpus of grasping postures, we conducted a data acquisition study. In this study we investigated whether a stereoscopic rendered object could be detected in advance, while the user reaches to grasp it, based only on her hand posture. The main goal of the study was to determine the parameters that affect this detection. Therefore, the participants had to perform typical *Reach to Grasp* tasks using different virtual stereoscopic displayed objects as visual stimuli. This was similar to the initial observational study, but in this study the hand grasps and motions were recorded with multiple depth cameras to build a corpus for later in-depth offline analyses.

8.2.1. Experiment

Participants

22 participants (19 male, three female) naive to the experimental conditions, took part in this study. The subjects were between 22 and 56 years old ($M = 28$, $SD = 6.9$) and none has reported any visual or stereopsis disruptions. All subjects were members of our institute or university students and reported, in a 5-point Likert scale, good to excellent experience with touch devices. All participants were right-handed.

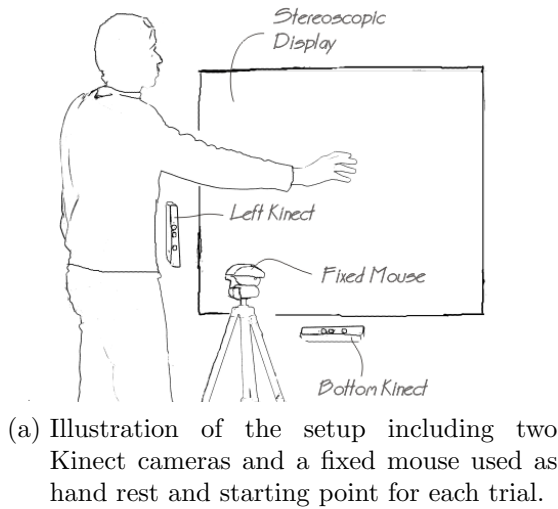


Figure 8.4.: Experiment setup and task.

Task

In this study subjects were asked to grasp virtual objects that are graspable counterparts to standard UI widgets (see Figure 8.4b). These widgets were designed to have approximately the same size and were meant to be interacted with via grasp gestures.

Subjects were positioned in front of the projection screen at a distance of approximately $3/4$ of their arm length, such that they could conveniently perform all grasp gestures during the study with their dominant hand. All trials had to be performed with the dominant hand. To guarantee a consistent initial start position at the beginning of each trial, the subject had to press the left button on the mouse mounted at a convenient distance (approx. 25 cm) on her right. As visual stimuli, ten different stereoscopically rendered virtual objects shown in Figure 8.5 were projected at five different object positions: $(-a, 0)$, $(a, 0)$, $(0, 0)$, $(0, -a)$, $(0, a)$ with a being the half arm length of the subject and $(0, 0)$ being adjusted to match the orthogonal projection of the participant's right shoulder on the surface. The back of all objects were aligned with the zero parallax plane (aligned with the projection screen's surface), and their positions were varied.

Procedure

After the left mouse button was pressed, the visual stimulus was displayed and the video recording of both depth sensors was activated. The participant had to grasp and manipulate each object in a way that felt natural. The visual stimuli were static and thus could not be manipulated, but the participants were asked to mimic an appropriate manipulation. Once the participant had reached the projection wall, the video recording was disabled and the visual stimulus was blended out two seconds afterwards to allow the subject to denote the intended manipulation.

8. Reach to Grasp Interaction

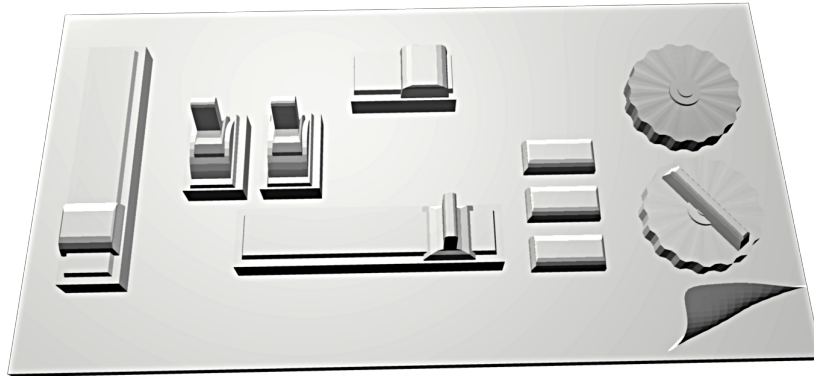


Figure 8.5.: Sample interface widgets that had to be grasped by subjects during the study.

Four trials had to be performed for each object at the five different positions resulting in a total of 200 trials per subject. Five additional trials (not included in the evaluation) had to be performed at the beginning to ensure that the subjects understood the task and received some initial training. After all trials were completed the subjects were asked to fill out a short questionnaire addressing their subjective experience with the interface, visualization issues, and fatigue during the performance of the study.

The entire study took about 30 minutes including training, questionnaire, breaks and debriefing. The subjects had to take mandatory two minute breaks at regular intervals to minimize errors due to fatigue or poor concentration. They additionally were allowed to take breaks at any time during the study.

Apparatus

The setup for the study is shown in Figure 8.4a. The study was performed using a prototype stereoscopic multi-touch projection wall. For the back projection, a projector with native resolution of 1400×1050 , using a frame-sequential stereoscopic projection at 120 Hz, was used. The projection uses only a portion of the touch-enabled screen with dimensions $136 \text{ cm} \times 102 \text{ cm}$, resulting in an effective pixel size of approximately 1mm (645 pixel per in^2). Although we could track the subjects' head positions, this was not needed since the subjects remained in the same position during the entire study. The position of the virtual camera and its viewing frustum were adjusted to match the subject's height.

Hand motions were recorded with two Microsoft Kinect depth sensors as RAW video streams with resolution 640×480 at 30 frames per second (fps). Both sensors were arranged (one at the left side of the projection and one below it) in such a way that the user's hand was in her FOV during the entire time of each trial. To indicate the start of each trial, the participants used a common computer mouse which was mounted on a camera tripod and also adjusted to each participant's height. The study was run on a PC with an Intel Core i7 processor with 8GB of RAM and an nVidia GeForce GTX470 graphics card.

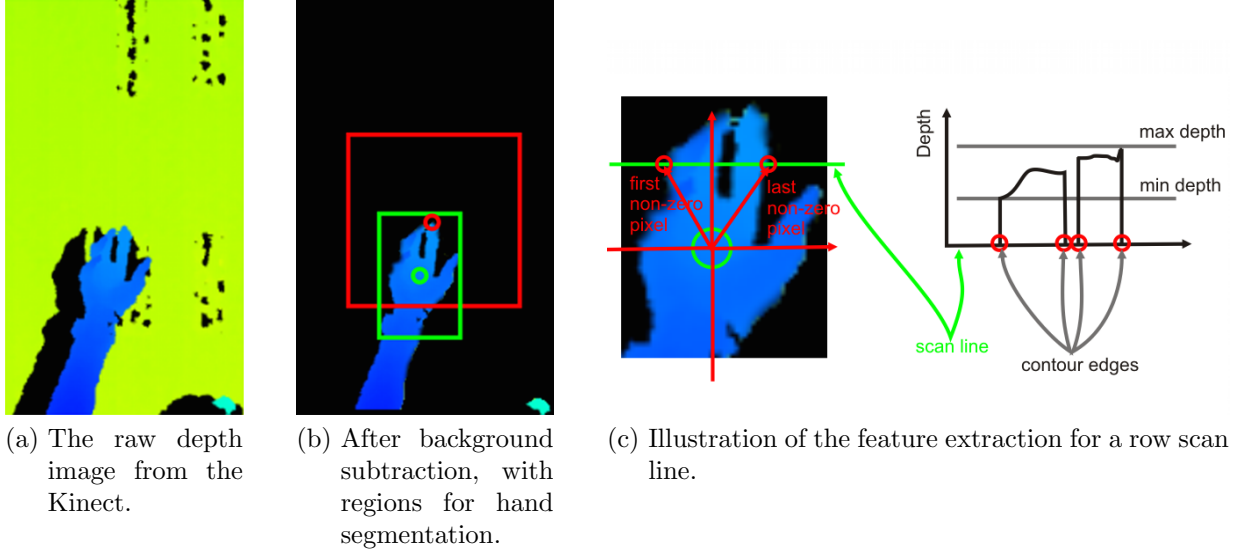


Figure 8.6.: Feature extraction pipeline.

8.2.2. Analysis

The collected RAW video streams are not applicable for direct evaluation. Thus, we first pre-processed the data set to extract an image-based set of representative features for each frame. Afterwards, we split the video streams into time segments, filtered the frame feature sets within each time segment to remove redundant information and evaluated the results with correlation-based algorithms.

Feature Extraction

For each frame in each video sequence, we first removed the background by clamping to zero all values above a threshold (i.e. too far away from the camera) and subtracting a static captured background image from the resulting frame as shown in Figure 8.6b. This pixel was used as a reference to determine a rectangular subregion of 200×200 pixels that contained the user's hand (the red rectangle in Figure 8.6b). We then found the weight center of the region as the mean of the pixel coordinates of all non-zero pixels within the region and built up a new subregion of 100×150 pixels centered in the weight center (marked with a green circle in Figure 8.6b). The hand contour and the distribution of the depth extrema within a depth image of the hand had been successfully used in multiple works as features for hand-gesture recognition [129, 121]. In our approach similar parameters, i.e. the number and the distribution of the depth minimums and maximums, were used, as well as the outer contour of the hand. In order to evaluate the frame data, we extracted from each region some representative parameters that seemed to contain meaningful data about the current hand posture. Nevertheless, we did not use the hand

8. Reach to Grasp Interaction

contour and topology directly, but extracted from the segmented subregions some unified representative parameters that were more appropriate for direct comparison.

The following parameters have been carefully considered to be most useful as feature vector (see Figure 8.6c): the number of depth minimums, as well as their mean, minimal and maximal values; the mean depth of the region; the number of non-zero pixels within the region; for each row, the unprojected positions of the first and the last non-zero pixel, relative to the unprojected coordinates of the region’s center; the number of contour edges; the number of non-zero pixels and the mean, maximal and minimal depths of the row; for each column, the same parameters as for the rows. This leads to a 2206-dimensional feature vector (6 global image features, 11 features per row and 11 features per column) which contains, for our consideration, the essential information of a frame. Such ad-hoc feature extraction may indeed contain a lot of redundant information. Unfortunately, this redundancy cannot be easily determined based on local features. We therefore performed additional filtering on the entire data set as described in the next subsection.

Feature Sets

Since the hand runs through the same phases while performing a *Reach to Grasp* task (or reaching task in general) the whole motion can be normalized by the time [167, 191]. The progress data should be temporarily scaled for each trial such that the trial begins at “time” 0 and ends at “time” 1. Such a normalization is usually done to enable direct comparison of the progress-relevant features among all subjects and conditions.

We normalized the trial performance times for each video sequence so that the mouse click (which indicated the beginning of the trial) is at “time” 0 and at “time” 1 the subject’s hand was 1 cm away from the virtual object to be grasped. Each frame and also each feature vector was labeled with its normalized time. We split the set of feature vectors into six groups based on their normalized time. In the first half of the motion, the grasp pre-shaping and the wrist transport were in a too-early stage, which made a prediction in this case a very challenging task. Indeed, in common settings, the wrist path is unpredictable until the transport phase reaches its peak velocity, usually at time 0.5 [134]; i.e., at this early stage, it would be even more difficult to determine if the participant was reaching towards the display or in some other direction. Thus the frames from the set $[0, 0.5]$ were excluded from further evaluation, because we were more interested in robust object prediction in a short interval before grasping that object.

To reduce information redundancy in the extracted feature vectors, features with constant values or very low variance within the datasets of each time segment were removed. Afterwards, the data sets were transformed with algorithms for principal component analysis (PCA) and the transformed feature vectors were constrained to the first n principal components, with n determined such, that at least 99% of the information was contained in the components.

normalized time	OT	OTP	OTU	OTPU
0.5-0.6	30.61	45.25	76.75	97.89
0.6-0.7	30.56	45.23	76.98	97.91
0.7-0.8	30.37	45.05	76.26	97.95
0.8-0.9	30.36	45.32	76.44	97.94
0.9-1.0	29.99	45.53	76.44	97.90

Table 8.1.: Mean prediction rates in percentages for the LEFT sensor.

normalized time	OT	OTP	OTU	OTPU
0.5-0.6	24.51	44.78	55.23	93.59
0.6-0.7	26.74	48.09	61.07	96.19
0.7-0.8	24.30	43.97	58.09	96.11
0.8-0.9	22.19	32.45	52.56	92.79
0.9-1.0	21.90	29.55	49.03	90.65

Table 8.2.: Mean prediction rates in percentages for the BOTTOM sensor.

8.2.3. Results

None of the participants assessed the study as being too long or the task as too difficult. Since all participants were able to perform the study without problems, we took all the data acquired into account.

The participants were asked to grasp in a natural way, with moderate but realistic speed from the resting position to the surface. From the video sequences we determined that the mean task performance time was $1584ms$ ($SD = 363.38ms$). This timeframe allows the recognition and tracking of grasps a feasible amount of time before touching the surface. As mentioned above, the performance times for each video sequence were normalized and the calculated feature set was split into six sets based on their relative time. From each training set, the features with constant values or very low variance were removed. Afterwards, the training sets were transformed with PCA. For the subsequent training of the classifier, only the first n principle components were used, with n determined such that at least 99% of the information was contained in the components.

Since we were interested in the influence of different parameters on the correlation between captured frames and the visual object, we used a very simple correlation-based classification algorithm, the *Naive Bayes* classifier. This classifier is based on maximization of the cross-correlation within the group of measurements (represented as multidimensional feature vectors) and minimization of the between-groups cross-correlation. The clustering variable of the *Naive Bayes* classifier was varied in order to test the desired parameters. The clustering variable is a nominal value that assigns each feature vector to a cluster [63]. *Naive Bayes* calculates (based on a training set) a clustering such that the cross-correlation

8. Reach to Grasp Interaction

of the feature vectors is maximal within each cluster and minimal between the clusters. Thus conclusions about the similarities of the feature vectors could be drawn based on the prediction rates of the classifier.

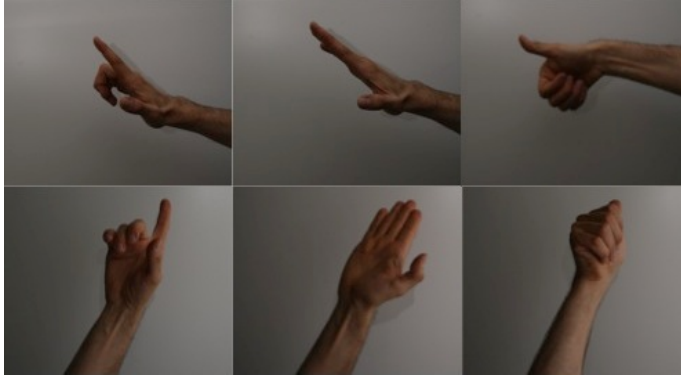
We tested four clustering variables: the object type alone (OT); object type and position (OTP); object type and user (OTU); and object type, position and user (OTPU). For each clustering variable and each training set, a classifier was trained with 80% of the feature vectors and its prediction rate was tested with the other 20% of the set. This process was repeated ten times and the calculated prediction rates were further evaluated with statistical methods. The results of this analysis are presented in the following section.

The achieved mean prediction rates for the left (LEFT) and the bottom (BOTTOM) sensors are shown in Tables 8.1 and 8.2, respectively. The data was analyzed with a factorial ANOVA in order to test the within-group effects of the time set, sensor position and clustering variable. The analysis revealed a significant main effect for the sensor position ($F = 16556$, $p < 0.001$) as well as for the time set ($F = 684.99$, $p < 0.001$) and clustering variable ($F = 169820$, $p < 0.001$). The subsequent post-hoc analysis with Tukey's test revealed significant difference for all the tested conditions and values (with $p < 0.01$).

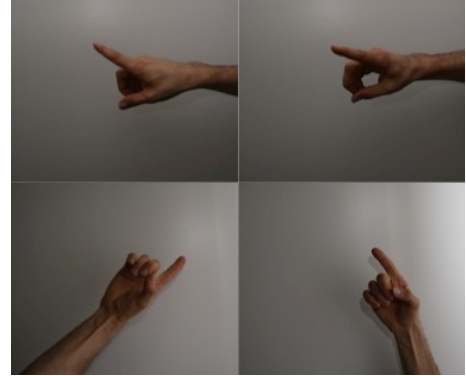
8.3. Discussion

The pre-study gave first insights on the interaction on and above interactive surfaces with virtual 3D sliders, knobs, etc. The participants understood the general idea of the interaction concept and acknowledged the novelty of the interaction. This was reflected both in their interaction with the interface and their comments on the interaction techniques. The results implied that grasp postures remain fixed and thus can be unambiguously determined at least ten frames before hitting the surface. These results enabled us to model the intention of the user while interacting right before finally touching the intended UI element. This information about the user's intention allows the adaptation of the user interface to improve interaction and user experience. With the knowledge of the intended UI element it can be, for example, shifted to zero parallax just before the user hits the interaction plane. Besides, an immersive interaction experience with virtual objects is provided by the passive haptic response of the physical border of the interactive surface. This was also identified by a participant who proposed graspable 3D menus. 3DUIs that rely on (touch-based) gestural interaction could be extended by such 3D menus and thus provide intuitive cues for manipulation and passive haptic feedback when the menus are projected close to zero-parallax.

As indicated by the pre-study, the results show that the hand posture reflects the object to be grasped. Thus the object type could be anticipated in advance based on features extracted from the captured hand posture. Given the best prediction rates residing in the time segment $[0.6, 0.7)$ and the mean task performance time of 1584 ms, this gives us about 500 ms in advance for this information to be used by the interface. Although the participants in our study performed the task slightly more slowly, than they might



(a) Subject dependent grasps: Illustration of three different types of pressing a virtual button as seen from the viewpoint of the depth camera.



(b) Position dependent grasps: Illustration of pressing the same button at different positions on the projection screen.

Figure 8.7.: Grasp variations.

in a real user interface, the 500 ms is a reasonable amount of time for a UI to adapt to the user's intention or to execute complicated background tasks, reducing the overall latency of the interface. One of the interesting results is that the prediction rates do not constantly increase with the hand approaching the visual object as initially expected, but have their peak values in the time cluster $[0.6, 0.7)$ and then fall. We assume the prediction rates might depend on the acceleration and deceleration of the hand. We currently cannot explain this effect with our data. This needs to be addressed in detail in future research by using additional tracking devices.

Not surprisingly, the object type on its own is not sufficient as a clustering variable. Indeed, the hand posture depends on the personal preference of the user, as shown in Figure 8.7a. This may have led to the significantly better prediction rates in the condition OTU. Nevertheless, it is currently unclear if there is a (perhaps broader) set of typical hand postures which could be mapped onto a single object to compensate for the personal differences. Surprisingly, the object position also has a significant effect on the prediction rates, although its effect is not as strong as the personal preferences (see Figure 8.7b). This might be due to the fact that our initial feature extraction does not fully compensate for different hand orientations. We expect that using more advanced feature extraction techniques will reduce or eliminate this effect. Indeed, more evolved feature sets which compensate for different hand orientations and sizes could be extracted from each frame as well as from the frame sequence. Such feature extraction would then make the recognition user-independent.

In general, the recognition of an object to be grasped depends on different parameters including the users' personal characteristics and habits, which may make a robust mapping of objects to grasp posture a challenging task. Nevertheless, our approach shows the feasibility of the task at hand and provides an easily reproducible procedure for establishing

8. *Reach to Grasp Interaction*

an initial corpus of training data. Based on the reported prediction rates achieved even with this very simple algorithm, we believe that a complex alternative method (e.g. [156]) fed with our training corpus may achieve remarkable, in many cases user-independent, prediction rates.

8.4. Design Considerations

The experiments have shown that the affordance of an object plays an important role. Because there are often multiple ways to grasp an object, UI elements should be carefully designed. If the usage of all elements is unambiguous, the recognition as well as the overall usability of the UI could be improved. Hence, it seems reasonable to design objects with unambiguous affordances that would reduce the variability of possible grasps to a single gesture. For this purpose it might be feasible to semantically enrich (graspable) objects with information about different grasps depending on their intended use. For example, there might be different grasps for sliders: a slight but precise manipulation of the slider's value with a subtle pinch gesture (precision grip) in contrast to a broader large value change of the slider by moving it to an extreme position with a powerful pinch (power grip) or wiping gesture. Moreover, in the context of adaptation, grasping is superior to simple pointing.

8.4.1. Optional Menus

In particular, on handheld stereoscopic multi-touch devices, saving screen space is crucial. A permanently displayed menu requires a lot of screen space. If the user's intention can be predicted early, only those menu items that will come into consideration could be displayed. These items then could be arranged in 3D in such a way that the user could reach and manipulate them easily.

8.4.2. Level of Detail Interaction

In computer graphics, accounting for level of detail rendering involves decreasing the complexity of the representation of a 3D object as the distance between viewer and object increases; other metrics may be applied such as object importance, eye-space speed or position. Such level of detail techniques increase the efficiency of rendering by decreasing the workload on graphics pipeline stages. However, the reduced visual quality of the object often goes unnoticed because of the small effect.

If the system knows which object the user wants to interact with, the same concepts could be applied for interaction with this object, and the object could be prepared for interaction. For instance, if the UI supports precise grabbing, precise collision detection is required, which is usually based on spatial hierarchies. By exploiting the knowledge of which object the user wants to interact with, complex collision detection algorithms could be reduced

to a single object in space. Other level of detail interaction concepts are conceivable, in particular for complex interactions such as object manipulation or deformation.

8.4.3. Adapting the 3D User Interface

It is now possible to design user interfaces that can be dynamically adapted based on predictions of the user's intention. Adaptation means that stereoscopically displayed 3D objects serve as virtually graspable counterparts of real objects, which respond to the user's grasping behavior. The adaptation can be either subtle or overt.

Subtle adaptation means that objects can be slightly adapted (e.g. change size or position) to fit the user's grasp shortly before she reaches the surface. Here, the parallax change plays the most important role, especially if the object is manipulated through direct multi-touch interaction (these parallax problems are discussed in [196]). If objects are displayed with negative parallax, they cannot be reached because they are located behind the surface, whereas interactions with objects displayed with positive parallax might lead to problems because the visual stereoscopic effect is disturbed when the fingers are between object and display. These problems can be compensated for just by shifting the objects to be displayed with zero parallax, i.e. to the surface. Then, corresponding haptic feedback can be provided. If the shift is small enough, this adaptation is not noticeable by users.

The key idea of this example is that virtual objects may provide haptic feedback, whereas in stereoscopic 3D multi-touch environments the touch surface itself loses importance. The touch surface serves only as a feedback device for virtual objects, and is "invisible" otherwise. This could be beneficial in immersive setups, where the user is immersed in a virtual environment. However, in certain situations, it may be more appropriate if the user perceives an interaction surface (e.g. when interacting with 2D objects such as photos). In such a setup the adaptation can be overt, and objects could snap to the 2D surface, when they have been predicted based on the grab gesture.

8.4.4. Improving Object Recognition

Fitts' law [69] is a well-known model of human movement that predicts the required time to rapidly move to a target area as a function of the distance to the target and the size of the target. This model can be applied for pointing to an object either by using a pointing device or by physically touching an object with a hand or a finger. Assuming the opposite situation, that we have knowledge about the distance between hand and object as well as the time until the hand reaches the object (e.g. touch surface), this model could be used to predict the size of the intended object.

One example for this is pointing. The process of predicting the intended object is rather simple. Based on the distance to the touch surface and the predicted time until the finger reaches the surface, one could estimate the size of the object and therefore predict the object. In the next step, the grasp posture will be determined and now the objects' semantics can help to reduce the search space. The interface can now be adapted so that the most likely object, for example a button, is moved into focus.

8. *Reach to Grasp Interaction*

Another example is the precision pinch where, again, in the beginning the motion of the hand is observed. In contrast to the example above, now the user forms a pinch gesture rather than a pointing gesture. Then a pinch posture will be detected and only the objects that are graspable and have an appropriate size are marked as possible candidates.

The design considerations discussed in this section can be seen as an initial step towards a new class of interfaces that allow an intuitive as well as graspable interaction with stereoscopic data.

8.5. Conclusion

In this part the concept of *interaction context* for touch-enabled interactive surfaces was introduced. After introducing the general concept several samples of it were discussed. Then as a dedicated example, the *Reach to Grasp* task was investigated in more depth. We thus conducted an initial study and a corpus acquisition study in which we analyzed hand postures above and on the interactive surface. Finally, we provided design considerations for *Reach to Grasp* interfaces and interaction.

In the initial study we gathered insights on the interaction with virtual 3D objects on and above interactive surfaces. The results of this study enabled us to model the intention of the user while interacting right before finally touching the intended UI element. This information about the user's intention allows the adaptation of the user interface to improve interaction and user experience. With the knowledge of the intended UI element, for example, that 3D object can be shifted to zero parallax just before the user hits the interaction plane. In addition, an immersive interaction experience with virtual objects is provided by the passive haptic response of the physical border of the interactive surface. In the second study, we collected a corpus of grasp postures for stereoscopic objects. The analysis of the gathered data shows that a recognition of the grasp posture during the *Reach to Grasp* phase is feasible within a certain amount of time (≈ 500 ms) before the user reaches the surface. This can be used to improve interaction and gives rise to novel user interfaces. These findings show that the objects the user wants to interact with can be predicted unambiguously before the user actually touches these objects.

We discussed design considerations and proposed a grasp recognition and adaptation approach. Following this, information about the grasp intention now allows the adaptation of the user interface to improve interaction. With such knowledge, the potential of novel interaction techniques and improvements in UIs can be tremendous. There are different potential application domains that can benefit from this UI and interaction concept. The most evident application example for our concept is 3D modeling. A 3DUI that enables direct touch interaction together with *Reach to Grasp* interaction enables direct manipulation for 3D modeling. Another good example for cluttered interfaces is a (virtual) mixer for DJs. Such a tool that emulates a real mixer console can be seen as a customizable expert interface. In such an interface that consists mainly of buttons, knobs and sliders the interaction concepts and virtual widgets of our studies can be directly applied. Brows-

ing and interaction in large image databases visualized in 3D might offer other interesting scenarios to investigate.

The findings of the experiments have also shown that the affordance of an object plays an important role. Because there are often multiple ways to grasp an object, a careful design of the UI items has to be taken into account. If the use of all elements is unambiguous, the recognition as well as the overall usability of the UI can be improved. Hence it seems reasonable to design objects with unambiguous affordances that reduce the range of possible grasps to a single gesture. For this purpose it might be feasible to semantically enrich (graspable) objects with information about different grasps depending on their intended use. For example, there might be different grasps for sliders: a slight but precise manipulation of the slider's value with a precision pinch in contrast to a broader large value change of the slider by moving it to an extreme position with a power pinch or wiping gesture. Moreover, in the context of adaptation, grasping is superior to simply pointing. It is now possible to design user interfaces that can be dynamically adapted based on predictions of the user's intention. Adaptation means that stereoscopically displayed 3D objects serve as virtually graspable counterparts of real objects which respond to the user's grasping behavior. Thus an immersive interaction experience can be realized by "touching" virtual objects together with haptic feedback through the physical border of the interactive surface.

Part IV.

Handheld 3D Interaction

9. Interactive Handheld Stereoscopic Devices

Quite a large body of research exists for mobile interaction in general, but little work has been done so far related to interactive stereoscopic mobile devices (see Chapter 2.3.2). With this class of handheld devices, new research challenges arise. This entails questions on how to design interaction techniques that are best suited for mobile stereoscopic displays. Nowadays, mobile devices are equipped with various sensors that allow additional input modalities, but combinations of input and output modalities need to be carefully chosen (see Chapter 2.2). Sensor-based input has a great potential for mobile interaction with stereoscopic data, for example, navigating a virtual scene by freely moving the handheld device in space. But in contrast to mobile interaction with 2D data, the possibilities and limitations of interactive handheld stereoscopic devices still need to be investigated in more depth.

This chapter is structured as follows. A set of mobile interaction techniques is presented that allows all basic 3DUI tasks for interaction with mobile stereoscopic devices. The interaction techniques are evaluated in a mobile 3D game on a standard mobile device that provides anaglyph stereoscopic output. After describing the mobile 3D game, the results of the study are presented. Contributions of this chapter have been also published in [51, 126].

9.1. Interaction with Handheld Stereoscopic Devices

The interaction techniques that are presented in the following are specifically designed for interaction with stereoscopically displayed data on handheld devices. The input relies only on the handheld's sensors. Thus no external tracking system is used to track how the user is holding, moving and touching the device. This approach makes the interactions applicable to real life scenarios where in general no external tracking technology is available. All common ways of holding a mobile device are suitable for the proposed interaction techniques (see Figure 9.1).

The mobile interaction techniques comprise all basic 3D tasks (selection, manipulation and travel). Objects selection and manipulation was realized by direct touch interaction. In order to touch and select an object the user needs to travel through the scene. The measurement of absolute movement based on the available accelerometer data is difficult due to errors based on the double integration of accelerometer readings [192]. Nevertheless, the accelerometer is able to reliably track discrete shaking gestures and thus accelerometer-

9. Interactive Handheld Stereoscopic Devices

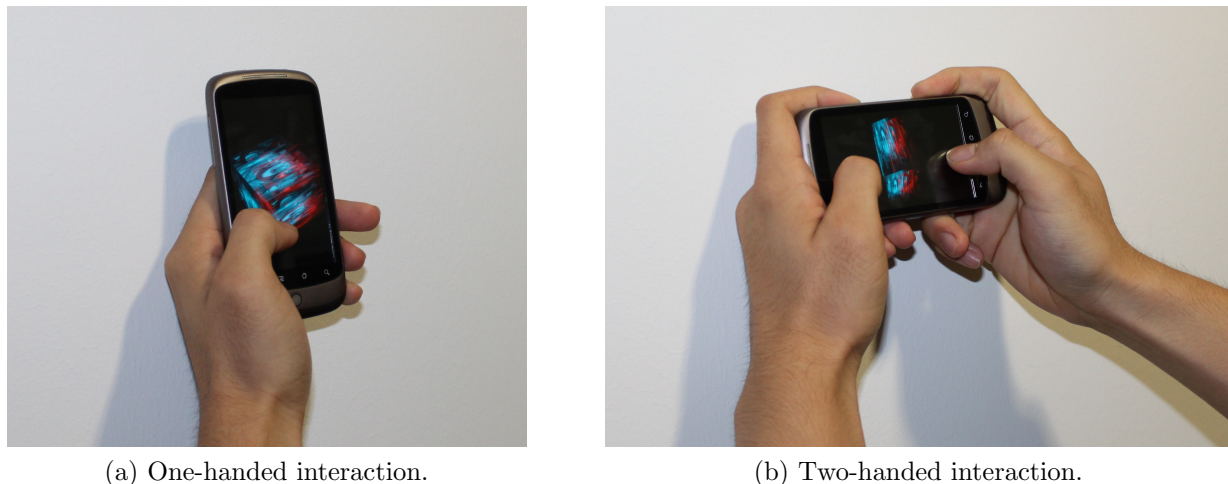


Figure 9.1.: Sensor-based interaction with stereoscopically displayed data on a mobile device.

based flipping was used instead of absolute movement of the device. Travel was performed through rotation and flipping of the mobile device. The movement of the mobile device in the real world also induces a change of view in the virtual world. This simply means that the camera in the 3D scene changes with respect to the movement of the mobile device. In the following mobile interactions for the canonical 3D tasks are presented.

9.1.1. Selection

Object selection is triggered by direct touch input. An object is selected by simply tapping it. The top most object will then be selected, i.e. through simple ray-casting. In combination with the travel techniques, this provides an intuitive yet effective way to select objects even if they are behind others (e.g. tilt and rotate the device to manipulate the scene and then touch to finally select the object).

9.1.2. Manipulation

Object manipulation is also realized by direct interaction. In addition, the combination of tilting and rotating the handheld with touch extends the interaction space and allows an easy and intuitive way to manipulate 3D objects. Direct touch manipulation and the passive haptic prop [90] property of the mobile enables a natural means of interaction. Interactions for the main direct manipulation tasks (translate, rotate, and scale) were designed and will be explained in the following.

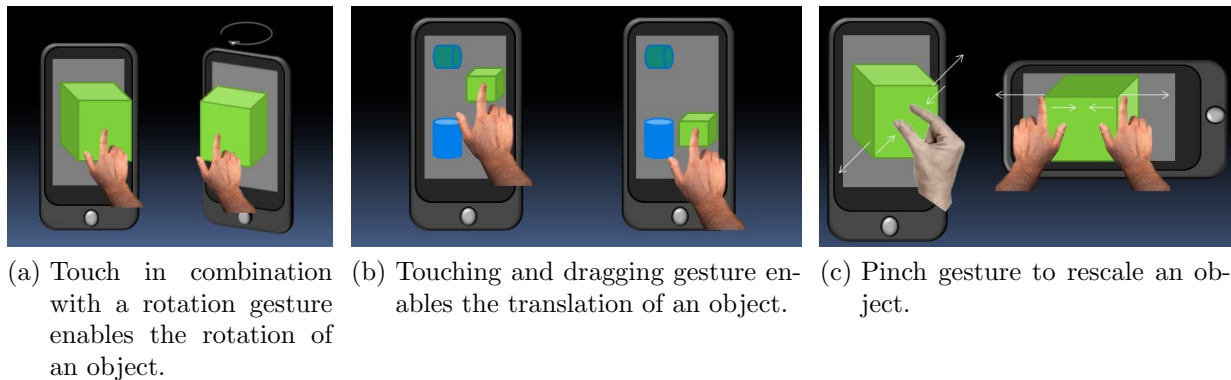


Figure 9.2.: Object manipulation techniques.

Object rotation

Rotating an object can be performed by touching and holding it followed by a rotation of the mobile device (see Figure 9.2a). This technique allows the rotation around an arbitrary axis with respect to the rotational movement of the device. This method extends the viewpoint rotation (travel) by additionally selecting and holding the object gesture to rotate the toggled object but not the whole scene.

Object translation

Moving an object can be done by touching and dragging it, which translates the object in 3D space with respect to the rotation and tilt of the mobile device. This technique allows the movement along an arbitrary plane that is easily comprehensible and reflected by the posture of the device. Figure 9.2b shows an example of moving an object with respect to the vertical plane defined by the position of the mobile device.

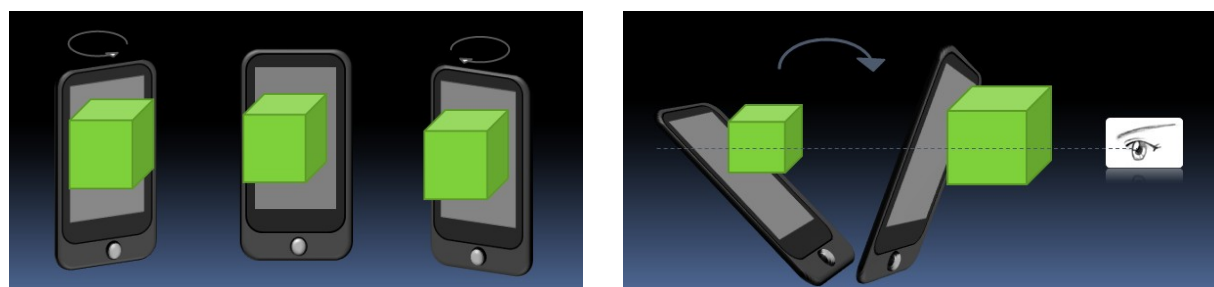
Object deformation

Object scaling is supported in order to deform objects. It is realized with the well established pinch gestures that change the size of a 3D object. Figure 9.2c shows an example of resizing an object with the pinch-to-zoom gesture.

9.1.3. Travel

Travel is realized through rotation and flipping gestures of the mobile device. The movement of the mobile device in the real world also induces a change of view in the virtual world. This simply means that the camera in the 3D scene changes with respect to the movement of the mobile device. The mobile device can be seen metaphorically as a tangible camera. Thus, the camera can be physically moved in every direction to change the field

9. Interactive Handheld Stereoscopic Devices



(a) Rotation gesture enables center rotation of the scene around an object.

(b) Flipping gesture allows (step-wise) zooming into and out of the scene.

Figure 9.3.: Travel techniques.

of view, its focus, etc. An interaction technique for the travel task (translate and rotate the virtual camera) will be introduced in the following.

Rotation

Rotating the mobile device can be either an orbiting around an object or a rotation in 3D space. Rotating the device corresponds to a rotation of the virtual camera in the scene. For a large 3D virtual space, rotation provides a viewpoint navigation method in the aspect of virtual camera rotation. For a single 3D object, it presents the 3D object as a fixed state, which means a stereoscopic object is fixed in front of the mobile device while performing device rotation gesture (see Figure 9.3a).

Translation

A flipping gesture is used to realize translation. Flipping gestures move the virtual camera in the scene forwards or backwards in a step-wise manner. A simple but precise navigation through the scene is realized with this technique. By flipping the mobile device quickly in some direction, the virtual camera is immersed along the corresponding axis (see Figure 9.3b). Flipping towards the user, for example, results in a movement of the objects closer to the viewpoint. Flipping the mobile device quickly away from the user has the inverse effect. This enables the user to quickly zoom into and out of the scene.

These interaction styles can be seen as a basic set of interactions for mobile 3D interaction. The framework that specifies this interaction technique was already introduced in Table 3.5 in Chapter 3. Depending on a specific use case or application, there might be the need to adapt the set to the requirements of the application. Based on these concepts, interactive demos and applications were developed that use the Android SDK¹ and OpenGL ES for the rendering of anaglyph stereoscopic 3D content. The sensors of the device are used to measure various dimensions of input: the touch sensor for direct touch input, the orientation sensor for rotational input and the accelerometer for shake gestures.

¹<https://developer.android.com/sdk>

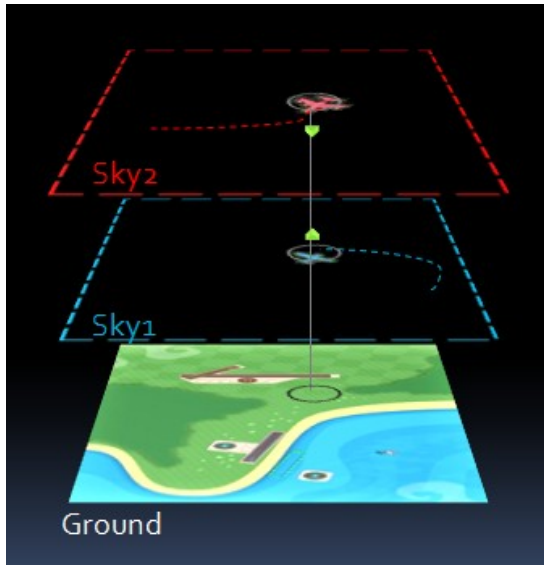
Red-cyan anaglyph pictures for stereoscopic 3D were rendered on the phone while users have been equipped with suitable anaglyph glasses. In the following an example application is presented that illustrates this adaptation process.

9.2. Flight Control 3D

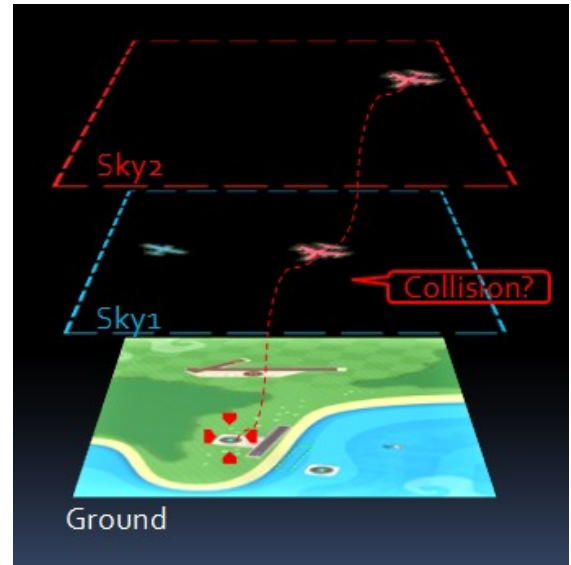
Based on the proposed interactions various application concepts can be developed. As an example application the concept of a mobile 3D game is presented. *Flight Control 3D* is a mobile game that combines the interaction styles proposed above with a 3D mobile game design. The presented game is inspired by *Flight Control*, which is a popular game for mobile device platforms. The goal of this game is to safely land planes and prevent collisions. New planes that appear periodically can be navigated by selection and sketching a path to define a flight route and to land at an appropriate airport. If a collision of two (or more) aircrafts happens, the game is over. So players need to plan and sketch the paths carefully to avoid collisions and to land as many aircraft as possible, which results in higher scores.

In contrast to the original game, in the *Flight Control 3D* scenario the planes have to be controlled in a 3D world. The integration of an additional spatial dimension leads to several design issues and a game experience that totally differs from the original. First, a simple 3D world has been developed that is based on a three layer concept: Ground, Sky1 and Sky2 layers (see Figure 9.4a). The airplanes move freely in the Sky1 and Sky2 layers, and only approach the Ground layer when landing. The airplanes cannot collide when flying in different layers. Thus, one strategy to avoid collision is to distribute planes on the different layers. On the one hand, the extension of the interaction space increases the spatial complexity of the game; On the other hand, more diverse game strategies can be pursued.

By tilting the mobile device, the view can continuously be changed from a birds-eye view to a side view. Depending on the view the user can define a path to determine the movement of an airplane. From above (birds-eye view), she can sketch the path and from the side (side view) a layer change can be performed. The user has to do two things: define the path of a plane and make it descend until it reaches the airport on the Ground layer. As in the 2D version, paths are defined by selecting and sliding. Layer changes are invoked by tilting the device and moving the plane to another level. Landing can be seen as a special level change and is only possible from the level next to the ground level, which is the Sky1 layer in Figure 9.4b. Layer changes are also needed in the landing procedure for planes in the Sky2 layer. Successfully landing will be rewarded just like in the original version.



(a) Layers can be switched by dragging the aircraft up and down.



(b) Airplanes do not collide when flying on different layers.

Figure 9.4.: Layered 3D world with Ground, Sky1 and Sky2 layers.

9.3. User Study

The mobile 3D game application was developed in order to evaluate the proposed mobile interactions. This approach was chosen to lower the participants' access barrier to 3D technology by providing a comprehensible 3D scene (the 3D world) and task (control air traffic) instead of requiring them to do an artificial 3D task (e.g. a docking task). The actual user study that is presented in the following covered general questions on the interaction techniques and the usability of the application.

9.3.1. Participants

15 participants were recruited for the study (12 male, 3 female). The average age of the participants was 25.6 ($SD = 2,03$). All participants had experience with smartphones and they all had played games on them to some extent. Six players out of the 15 had played the 2D version of *Flight Control* before and were familiar with the game concept. The others could be categorized as casual players. In terms of 3D experience, all of the participants had some experience with 3D movies, 3D TVs or other 3D devices.

9.3.2. Task

The participants were requested to play four levels of the game. They were asked to complete these four levels sequentially and try their best to safely bring down every airplane without any collision. In the first three levels, the participants needed to safely land a fixed

but increasing number of planes (Task 1: 1, Task 2: 3 and Task 3: 5). In the final level they had to land as many planes as possible within two minutes (Task 4).

The evaluated application was a game and the level concept of games has an inherent ordering effect in any case. Thus, the tasks were not counterbalanced.

9.3.3. Procedure

After the participants filled out a short questionnaire to gather personal details (age, gender etc.) and information about their level of experience with mobile devices and 3D graphics, the participants were introduced to the mobile 3D game. To familiarize the participant with the game, three quick tutorial lessons were performed. After the tutorial the actual game started. Similar to the tutorial, four levels had to be completed. The participants were asked to complete these four levels sequentially and challenged to try their best to safely bring down every airplane without any collision. All user interactions were logged on the device and the actual tests were videotaped for later analysis. After the actual usability study, the participants were asked to fill out a post-study questionnaire with open-ended questions to gather subjective feedback about the interactions, the game concept, etc.

9.3.4. Apparatus

As a study apparatus, the game was deployed on a Google Nexus One. Red-cyan anaglyph pictures for stereoscopic 3D were rendered on the phone, while users were equipped with suitable anaglyph glasses. All user interactions were logged on the device and the actual test was videotaped for later analyses.

9.3.5. Independent and Dependent Variables

Independent variables were number of flights (Task 1-3) and time (Task 4). Dependent variables were task completion time (Task 1-3) and error rate (Task 4).

9.4. Results

Different insights emerged from the gathered data. First, results regarding the game usability are introduced. Second, the results of further analyses with respect to the interaction techniques are presented.

9.4.1. Game Usability

The game concept was mainly evaluated through the questionnaire. However, some performance metrics will be reported as well. The feedback from the participants gives insights on the user experience of the game. In addition, general feedback is reported that was gathered via open-ended questions in the questionnaire. Finally, quantitative results (i.e. the

9. Interactive Handheld Stereoscopic Devices

performance and error rate) are presented. From the results of the open-ended questions, usability issues were derived and further analyzed by the logged data.

Game Performance

Fixed-plane task (Level 1-3) performance was measured with task completion time and success rate. Level 1 was completed with an average task completion time of 20.7 sec ($SD = 12.0$) and 100% success rate. Level 1 proved at least that all participants were able to perform the task. Average task completion time of 62.4 sec ($SD = 18.7$) and an 82% success rate were reached in Level 2. In Level 3 the participants performed with an average task completion time of 82.0 sec ($SD = 34.6$) and a success rate of 68%. This makes an overall average of 57.7 sec ($SD = 32.1$) with a success rate of 83.4%. In the free play condition (Level 4) time was fixed to two minutes, and thus performance was measured by the number of successful landings and success rate. On average, the participants were able to successfully land 9.9 planes within two minutes, with a success rate of 79.4%.

Questionnaires

The feedback from the questionnaire shows that the participants liked the game concept and enjoyed playing the game. Almost all participants agreed with the statement that they had a lot of fun playing the game. Seven of 15 agreed and six strongly agreed with the statement ($M = 4.4$, $SD = 0.61$). This result is also confirmed by comments from the open-ended questions: “I feel the game is so interesting!” (S9), “I really expect the game can be downloaded via Android Market.” (S1).

The difficulty of the game was judged less uniformly and could be labeled as moderate ($M = 2,87$, $SD = 1,02$). That result is also confirmed by the success rate. Both results illustrate that the level of difficulty is pretty suitable for most of the players.

General Feedback

Besides the general results on the game concept, specific feedback and new ideas were also provided by the participants through the open-ended questions in the post-study questionnaire. First, four participants suggested to adding full 6-DOF path drawing and rotation to enable greater freedom of control in the game: “360° movements” (S1), “maybe also use the second angle” (S2), “path displaying and distinguishing / more realistic to reject abnormal paths” (S8), “vertical path drawing” (S15). Due to the anaglyph visualization, a grayscale scene was chosen to provide a comprehensible 3D scene without confusing false-color effects. However, colors were missed in the games and were requested by four participants for different reasons: “colors would help a lot (crash, select, landing)” (S1), “if more colors can be added it will be more interesting” (S3), “Make it more colorful” (S9), “Better display” (S11).

Additional features were proposed that could add more variety and might increase the difficulty, such as including weather conditions (S5). For example, in rainy weather landing could be harder to perform, or in windy weather, the paths of the airplanes could be

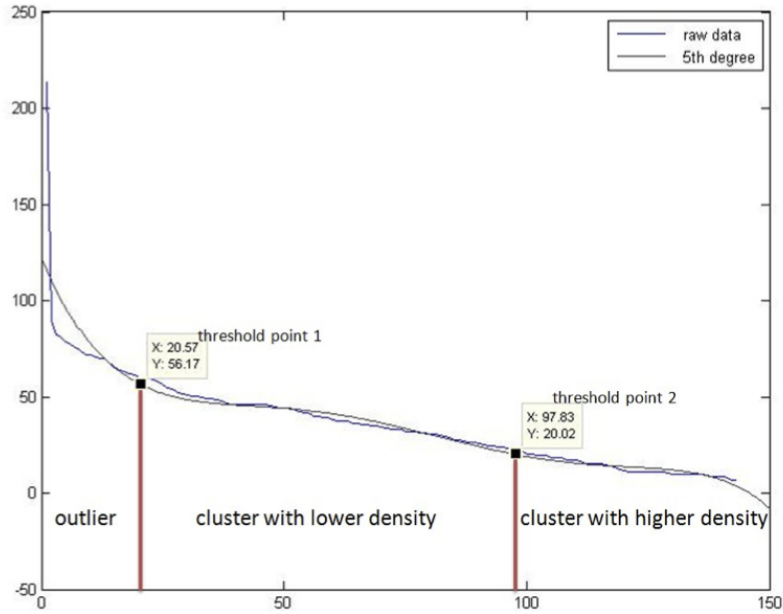


Figure 9.5.: The 4-dist graph: the 5th degree polynomial represents the trend in the data as a smoother curve. Thus the thresholds are obviously deducible. The first threshold delimits the outliers or the noise while the second differentiates the different clusters [126].

affected depending on the wind direction and speed. Another suggestion was to change the perspective of the user into a first-person view (S2). All suggested features and concepts seemed aimed to increase the fun of the game and the experience of 3D content.

In summary, it can be said that the basic game was playable and challenging with a moderate difficulty. While the participants' feature requests (additional difficulty levels, full 6-DOF control, etc.) mainly address an increase in game difficulty and fun, some technical issues were also revealed by the questionnaire. The selection and manipulation of 3D objects in the game are discussed in the following.

9.4.2. Interaction techniques

Selection

Regarding the selection of objects on stereoscopic mobile displays, the selection of objects (here airplanes) was investigated. The selections (successful and failed) were manually counted through video analysis and the following success rate was used as a metric for a successful selection:

$$p_s = \frac{\text{SuccessfulSelections}}{\text{TotalSelection}}$$

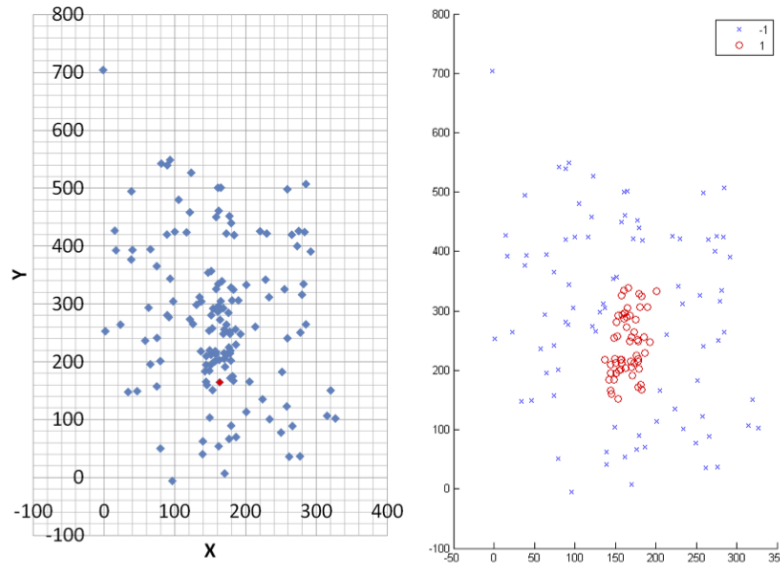


Figure 9.6.: The touches clustered with DBSCAN before (left) and after (right) clustering. The starting condition on the right shows all touches that initiate a manipulation (in blue) and the target position at the start of the runway (red). The clustering results in a high-density cluster of the failed landing attempts and the low-density cluster of other touches [126].

The mean of this success rate was 0.86 ($SD = 0.079$), which indicates fairly good results for selection. The subjective results from the post-study questionnaire show the same results. On a 5-point scale (1 = strongly disagree - 5 = strongly agree) five participants agreed that the selection was intuitive and easy to use, while ten participants strongly agreed with this statement. On the other hand, the questionnaire revealed an issue that might also have affected the selection of objects: “I noticed that it became difficult to touch a plane at about 45° . It seemed the game couldn’t tell if I wanted to draw a path or change flight level.” (S14). The latter issue will be briefly addressed in the discussion section below.

Manipulation

The success rate for object manipulation was investigated by observing the runway approach. The tracking of successful and failed landings was a critical issue that was also revealed by the questionnaire. In order to solve this problem, we analyzed all paths in layer Sky1 that did not result in a successful landing. Video analysis was not helpful to distinguish these two interactions. So all manipulations were split up into failed landing attempts and simple maneuvering (e.g. to avoid crashes). We then used a clustering approach to distinguish touches that intended a landing approach, but failed, from other touches. *DBSCAN* [65] was used for clustering. *DBSCAN* is a density based clustering algorithm for spatial data that is able to efficiently discover clusters of arbitrary shape. The algorithm is based on a so-called Eps-neighborhood. The neighborhood can be of

arbitrary shape and is defined by a distance function for two points p and q : $dist(p, q)$. The Eps-neighborhood of a point p is defined by $N_{Eps}(p) = \{p \in D \mid dist(p, q) \leq Eps\}$. The parameters *Eps* (given radius) and *MinPts* (minimum number of points) were determined as proposed by Ester et al.[65]. An explorative analysis showed that the k-dist graphs for $k > 4$ did not significantly differ from the 4-dist graph (see Figure 9.5). The threshold for outliers is the first gap in the sorted k-dist graph. The threshold can be interactively adjusted by the user. All points with a higher k-dist value than the threshold are considered noise, while all other points are assigned to clusters. The 4-dist value of the threshold point is used as the Eps value for DBSCAN. Figure 9.6 shows the clustering results. The red cluster represents all failed landing attempts. Thus, the success rate for landing operations is defined as

$$p_i = 1 - \frac{\#FailLanding}{\#TotalAttempts} = 1 - \frac{\#FailLanding}{\#FailLanding + \#SucceedLanding}$$

where $\#FailLanding$ was calculated with DBSCAN, and $\#SucceedLanding$ was logged by the system. The resulting mean success rate was 0.895 ($SD = 0.081$). While the data shows that all participants performed reasonably well, the feedback from the questionnaire was a little more ambiguous ($M = 4, 13$, $SD = 0, 99$).

9.5. Discussion

The results of the study indicate that recent mobile technology is well suited for 3D input and output not only from a technological viewpoint, but also from an interaction design perspective. Sensor-based interaction that enables the user to navigate in the virtual world by moving the mobile device in space offers a rich set of metaphors and interaction for 3D. Movement in the real world, such as viewpoint changes and traveling, can be more or less directly mapped to the corresponding interactions in the virtual world.

9.5.1. 3D Game Experiences

The participants liked the game concept and enjoyed playing the game. Although the game was not comparatively tested against the classic game, one can speculate that there might be an even better user experience than in the 2D version of the game. Most of the participants strongly agreed with the statement that they had a lot of fun playing the game, which was also confirmed by responses to the open-ended questions. The difficulty of the game was judged as moderate, which is also confirmed by the success rate. Both results illustrate that the level of difficulty is pretty suitable for most of the players. However, to satisfy more experienced players, additional modes or levels need to be created that would enable higher speed, more airplanes, etc. This trend was also reflected by the answers to the open-ended questions. Many participants suggested additional features that would make the game more difficult and varied.

Interaction Techniques

Selection of objects therefore can be effectively performed in stereoscopic mobile devices. This conclusion can be drawn from the quantitative and qualitative results of the study. However, a comment from the questionnaire and video analysis revealed a potential usability problem that might affect the toggle functionality. The orientation of the device toggles either the layer switch or path drawing (vertical vs. horizontal). One reason might be that immersion in the game has influenced the recognition of the tilt angle. The observations indicate that the participants misjudged the distance between objects displayed in different layers. Thus, there might be a disparity in the user's perception of where to touch to select an object when the device angle has been changed from the unambiguous horizontal position. This issue needs to be investigated in detail. A selection task needs to be designed in a highly cluttered 3D scene to further analyze this problem.

From the game perspective, interactions such as viewpoint navigation and object manipulation helped to realize the game tasks. For selection, the airplane is simply selected by tapping on the screen at the relevant position. Moreover, tilting the device correspondingly rotates the viewpoint of the camera, and choosing an appropriate angle helps to select an airplane which is occluded by other airplanes. Drawing paths and landing lead to problems in some cases because sometimes it is hard to identify the current layer of an airplane from above. And as already introduced above, only in the layer next to the Ground layer can the airplane land, so the user needs to make sure which layer the intended airplane belongs to. However, frequent tilting of the device seems to be a solution for this problem: the airplanes' layers are clearly perceptible in the side view of the virtual world. Moreover, tilting the device over a threshold angle may trigger the layer change mode. Therefore, tilting the device as well as selecting and sliding the airplane may perform a layer change operation.

Designing 3D Game Interactions

Some problems occur when adapting the interactions to the game. In the gaming scenario the flipping gesture (for viewpoint movement) has the disadvantage that flipping the devices is already associated with the rotating gesture. This can annoy the user if she does not want to rotate the device. For example, frequently watching the unexpected tilting of the stereoscopic scene will cause ergonomic problems because flipping is much more tiring than normal tilting. The conclusion that can therefore be drawn is that the design of gestural interactions for games needs to be done carefully. In general, requirements other than the input modalities play an important role and might restrict input to a few distinct gestures to make the game playable at all.

User-centered design for productivity applications versus games differs because it targets slightly different design goals. In particular, balancing difficulty is an important task in game design, which makes it harder to measure the usability of a game [153]. Our approach is somewhere in between these poles, since we used a game to investigate basic interaction

tasks. We believe that this approach helps us to gather insights on those tasks regarding effectiveness, difficulty and frustration that will inform all kinds of applications.

9.6. Conclusion

In this chapter we investigated interaction with handheld devices that support 3D input as well as output. Those stereoscopic mobile devices offer new opportunities but also technological and perceptual challenges. To investigate this, we explored different research directions. We first studied 3D interactions in VR on stereoscopic mobile devices.

The proposed interaction styles for 3D interaction on mobile devices using touch and motion sensors proved to work in a mobile 3D scenario. The interaction concepts were applied to a mobile (anaglyph-based) 3D game that was realized on the Android platform. The results of the study indicate that the participants liked the game concept and had a lot of fun playing the game. However, the difficulty of the game needs to be reviewed. Regarding the interaction techniques, the study revealed some usability flaws that indicate general issues regarding 3D perception and interaction. These issues need to be investigated in more depth. Nevertheless, sensor-based mobile 3D interaction, when carefully designed, provides an intuitive and joyful means of interaction. Mobile stereoscopic 3D is thus very well suited for mobile games such as the one presented here. Nevertheless, some open research questions regarding interactive mobile stereoscopic devices remain. Beyond mobile 3D interaction, the perception of stereoscopic data on small mobile screens plays a crucial role. The perception of stereoscopic data on mobile devices will thus be investigated in the following section in a handheld AR scenario.

10. Handheld Stereoscopic Augmented Reality

Handheld devices that are equipped with autostereoscopic displays, in principle enable users to perceive stereoscopic 3D without additional equipment such as shutter or anaglyph glasses. These mobile stereoscopic devices lead to new interaction design challenges (as discussed previously in Chapter 9), but on the other hand, perceptual issues need to be investigated as well. Stereoscopic displays are known to enhance depth discrimination in large displays like television, however it is not known if the positive effect is produced in the viewing conditions that are typical for handheld stereoscopic devices or not. Moreover, the mobile case involves a drastically smaller display size. The technology itself poses further limitations. The mobile autostereoscopic devices that are available on the market use the parallax barrier technique. This technology has constraints on small displays such as limited viewing range and field of view (cf. [13]). But again, little research exists regarding the perceptual limitations of this technology.

Depth interpretation is a common problem in AR applications and creating a perceptually correct augmentation is still a challenging task [116]. Even less research has investigated the use of autostereoscopic mobile devices for AR. We address this by proposing and studying the concept of stereoscopic handheld augmented reality. The main goal of this approach is to produce effective depth cues for AR by inserting stereoscopic 3D objects into the viewfinder. In order to address this, we conducted a psychophysical experiment that investigated depth discrimination in a commercial autostereoscopic mobile display. Virtual objects are overlaid on a real scene on a camera viewfinder and the participants have to distinguish which one is closest.

In the following the concept of stereoscopic handheld augmented reality is introduced. Then, a psychophysical experiment that investigates depth discrimination in stereoscopic handheld AR is presented. Finally, the results of the experiment will be discussed. Contributions of this chapter have been partially published in [109].

10.1. Handheld Stereoscopic Augmented Reality

The two common display technologies possible for mobile AR are HMDs and video see-through displays. In contrast to HMDs that provide mono- as well as stereoscopic projection, systems that rely on video see-through displays are based on monoscopic displays.

10. Handheld Stereoscopic Augmented Reality

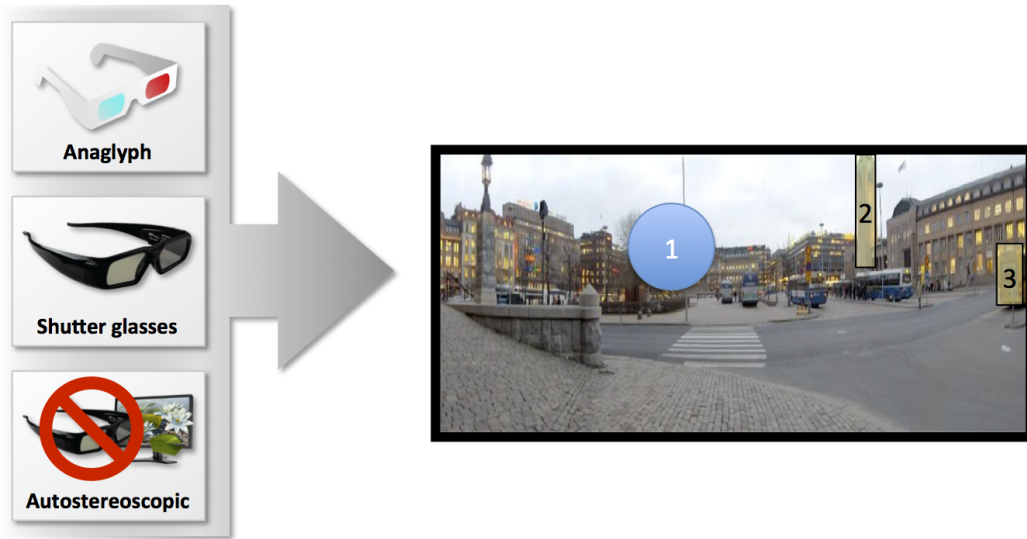


Figure 10.1.: Stereoscopic Handheld AR Concept: Object-referring augmentations are displayed in depth (1-3). This enables “free floating” augmentations that are easily comprehensible (1).

Depth perception in MR (mainly using HMDs) has been investigated in indoor and outdoor environments. It has been shown, that people often underestimate depth in indoor environments while they overestimate depth in outdoor environments (e.g. [130]). However, only a little literature on depth perception on magic lens displays exists (e.g. [60]). Besides published work (see Subsection 2.1.3), anecdotal evidence suggests that users have difficulties in assessing the distance of virtual objects on magic lens displays, i.e. people cannot accurately register how AR objects in distance relate to the real world. We assume that AR applications have not had a breakthrough because of this problem. A currently emerging class of handheld devices is equipped with autostereoscopic displays, which in principle would enhance users’ 3D perception without additional user instrumentation (i.e. shutter or anaglyph glasses). However, it is not known if an enhanced depth discrimination is reproducible for handheld viewing conditions.

A new class of handheld AR, which we refer to as stereoscopic handheld AR, could solve the problem of depth discrimination. The main idea is to use autostereoscopic mobile displays that enable additional depth cues, in particular stereoscopic depth cues. Our approach towards stereoscopic handheld AR will add stereoscopic depth cues to object-referring augmentations (see Figure 10.1). This might improve the overall impression of depth of objects in the magic lens view in addition to the existing monoscopic depth cues (e.g. augmentations 2 and 3 in Figure 10.1). But it further enables “free floating” augmentations (see augmentation 1 in Figure 10.1) and thus allows a better spatial guidance.

10.1.1. Interaction

The device properties and requirements for interaction techniques are similar to the one in the last preceding section. The mobile device can be freely moved in space, it is touch sensitive and it provides stereoscopic output. The main difference is the level of virtuality in the reality-virtuality continuum. While the mobile 3D interactions presented above can be clearly positioned at the end of the continuum in the virtual environment, the stereoscopic handheld AR approach is located in the mixed reality environment. However, compared to common handheld AR approaches with 2D output, the see-through view is captured and displayed in stereoscopic 3D. This can be seen as another level of augmentation. Thus we claim that the level of virtuality is higher than in the standard handheld AR case. We propose to position it in the continuum right next HMDs that provide the environment with augmented virtuality by augmenting a camera view in the virtual space. The handheld stereoscopic AR approach still allows context switches between the stereoscopic augmented view and reality. On that account it cannot be categorized as fully augmented virtuality.

As already mentioned, the interactions that are well suited for stereoscopic handheld AR are similar to the mobile 3D interactions (see Chapter 9). Navigation is basically the same, i.e. changing the position and orientation of the device enables a 3D navigation with 6-DOF (see Figure 10.2). The manipulation of virtual objects can be also made possible by directly touching and moving the device. Nevertheless, in AR additional interactions arise through the nature of MR. On the one hand, virtual objects that are coupled with (changing) real objects react to changes and update their position, pose and shape with respect to these objects. On the other hand, real objects can be interactively manipulated through their virtual counterparts, for example light switches, buttons or other controllable objects. We expect that cluttered physical environments can be intuitively and effectively manipulated via handheld stereoscopic AR. We envision that stereoscopy will be able to improve the affordance and depth discrimination of such environments. Due to their increasing complexity, smart environments provide various potential application scenarios. Smart homes, for example, can be controlled via handheld stereoscopic AR by augmenting a cluttered environment with interactive controls that are spatially correctly placed.

10.1.2. Proof-of-Concept Application

A first proof-of-concept stereoscopic handheld AR application shows that it works in principle (see Figure 10.2). Virtual 3D objects as well as the see-through camera view are both displayed in stereoscopic 3D. The handheld device is tracked with an Optitrack system and virtual objects are registered in the real-world space. Figure 10.2 shows a virtual cube that is projected on a real table. The virtual object can be interactively explored by moving the device around it like a magnifying lens. While the viewpoint of the device can be changed, the cube remains in a stable position at the corner of the table at all times. However, this early prototype has already revealed some issues that are critical for the handheld AR applications that we envision. Most prominently, problems with depth discrimination have been observed in our initial pilot studies.

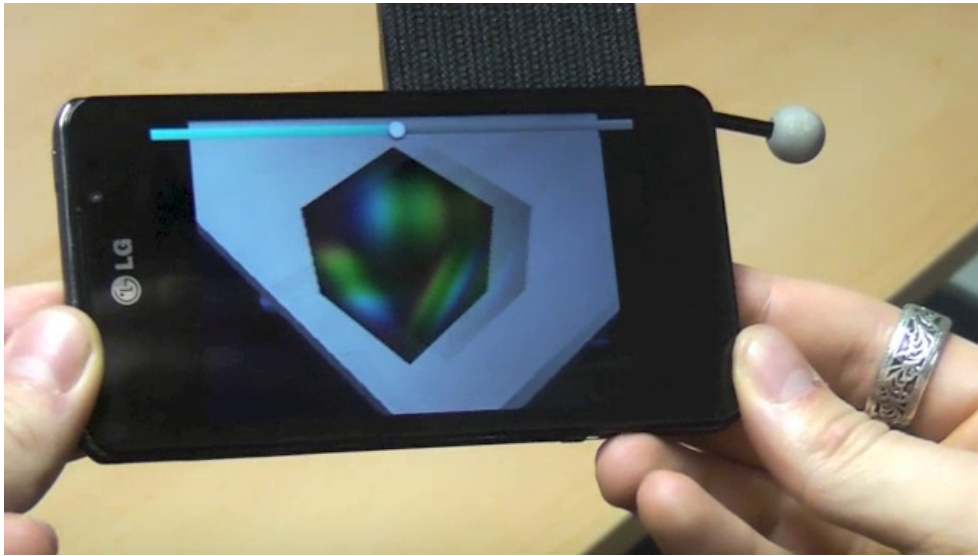


Figure 10.2.: Stereo Handheld AR Demo as proof-of-concept: A virtual cube is projected on a real table. The virtual object can be interactively explored by moving the device is moved around it like a magnifying lens. While the viewpoint of the device can be changed the cube remains in a stable position at the corner of the table at all time.

The goal of this approach is to produce effective depth cues for AR. Therefore, we need to exploit monoscopic and stereoscopic depth cues for AR by inserting stereoscopic 3D objects into the viewfinder. By manipulating mono- and stereoscopic cues and investigating them in studies, further insights will be generated about the depth discrimination of handheld AR. In the following we present an experiment that investigates binocular parallax and size as depth cues in handheld stereoscopic AR.

10.2. Experiment

Based on methodology from psychophysics, we designed an experiment with two discrimination tasks. Realizing the limitations of present-day autostereoscopic displays, we calibrated the viewing conditions to be as close to ideal as possible. Therefore, a chin rest was used to ensure that the participants had a consistently stereoscopic effect throughout the whole experiment. We noticed in a pilot study that, without such constraints, users would intuitively hold the device at angles and distances unfavorable to the autostereoscopic display. This is especially critical for users without previous experience with such devices. Our data is thus based on the best-case scenario.

Besides other mostly monoscopic depth cues (e.g. motion, occlusion, perspective, gradient, etc.) it is assumed that the additional binocular parallax depth cue might help with the discrimination of objects, especially in a densely cluttered environment. These questions are of particular interest for AR settings where virtual objects of different sizes

are used. Thus, we further investigate the object size as another depth cue. Therefore the following research questions are addressed:

- Does autostereoscopy improve users' depth discrimination ability, or do they rely more on monoscopic cues such as object size?
- Does the presentation of virtual objects in different parallaxes influence the depth discrimination ability?

The main goal of the study is to measure the effects of stereoscopy in a mobile context and thereby exclude the influence of any other depth cues. Our experiment considers both negative and positive parallax. Stereoscopy is expected to improve spatial perception and guidance of virtual objects in AR scenarios. However, in stereoscopic handheld AR, the see-through metaphor holds only for the positive parallax case, because the real environment in the camera view lies behind the device. Nonetheless, the negative parallax condition is also taken into consideration since it can even be used in an AR scenario (e.g. for controls).

10.2.1. Participants

Twelve participants (two female, ten male) aged between 21 and 34 ($M = 25.3$) were recruited for the experiment. All of them were informed about the aim of the study and the procedure. All participants were members of the university (75% students and 25% researchers). They were invited for two to three sessions

Every participant had normal or corrected-to-normal vision. Ten of them reported prior experience with stereoscopic effects (e.g. 3D cinema) and four reported prior experience with autostereoscopic devices. To ensure that participants were generally capable of perceiving 3D, we conducted a pre-test in which we sequentially showed several cubes on a white background, either in NEG-P or POS-P. Participants had to state whether they had the impression of cubes floating in front of the display (NEG-P) or behind the display (POS-P). The results showed that none of the participants had any severe problems with stereoscopic vision (success rate $\geq 85\%$, $M = 94.2\%$).

10.2.2. Task

We designed two discrimination tasks that investigated the depth discrimination capability with respect to object size (Task 1) and autostereoscopy and parallax (Task 2).

Task 1: Object Sizes

In the first task, we investigated the influence of the objects' sizes on depth discrimination capability. The camera image and the cubes were always shown in autostereoscopic mode. The design considered cube size (large/small) and parallax (NEG-P/POS-P) as the independent variables within subjects. We uniformly varied whether cubes were shown large or small and in NEG-P or POS-P, resulting in four different conditions which were presented five times each in a random order. The cubes' sizes were adjusted such that the

10. Handheld Stereoscopic Augmented Reality



Figure 10.3.: Autostereoscopic device showing augmented camera images with two relative depth differences (left: 10 *mm*, right: 104 *mm*)

small cubes in NEG-P had the same apparent size (measured in pixels) as the big cubes in POS-P and vice versa. Cubes in NEG-P were placed at a depth of 700 *mm*, those in POS-P at 1700 *mm* (referring to the cubes' front face, measured from the camera's position).

Task II: Autostereoscopy and Parallax

With the second task, we investigated the effect of autostereoscopy as well as the parallax on the depth discrimination capability. As shown in Figure 10.3, the participants had to decide which of the two cubes with varying depth was closer. With the help of an adaptive stair-casing procedure, we determined the required minimal depth distance between the two virtual cubes to be able to discriminate between them.

10.2.3. Procedure

In Task I we uniformly varied whether cubes were shown large/small and in NEG-P/POS-P. The cubes' sizes were adjusted so that the small cubes in NEG-P had the same apparent size (measured in pixels) as the big cubes in POS-P and vice versa. Cubes in NEG-P were placed at a depth of 700 *mm*, those in POS-P at 1700 *mm* (referring to the cubes' front face, measured from the camera's position). Participants were told that both cubes had been placed at the same depth and that they should decide whether they were shown in NEG-P or POS-P and report their choice verbally.

In Task II the first two independent variables (autostereoscopy and size) were counterbalanced via a balanced Latin square and the order for the possible cube depths was randomized. We designed this study based on an adaptive stair-casing procedure: The presented stimulus remained the same until a discrimination capability could be assumed or rejected with a certain confidence. If the participant could discriminate the stimulus, the next was presented with reduced intensity. Otherwise, it was presented with a higher intensity. The goal of this procedure was to find the minimal intensity at which the participant was able to discriminate. We decided to use a PEST procedure [189] for this, as it has the advantage of adjusting the change in intensity based on the prior performance to achieve a faster convergence towards the final intensity level. In our case, the stimulus

intensity mapped to the depth distance of the cubes' front faces to each other and additionally, after one completed PEST procedure (i.e. change in stimulus intensity $\leq 1\text{ mm}$), a new PEST procedure was started with a different object depth. For every stimulus presentation, one randomly chosen cube was displaced in depth accordingly. Between every two steps, a fixation crosshair (a black cross on a background similar to the wall in the camera image) was shown for 500 ms . The participants were instructed by the experimenter to decide which of the two shown cubes is closer to them. They had to report their choice by pressing one of the corresponding shoulder buttons on a Playstation 2 controller even if they were unsure about their decision (two-alternative forced-choice).

The experiment was divided into two to three sessions to ensure that eyestrain would not affect the results. A session lasted 48 minutes on average. The first session allowed for completion of Task I and the first half of Task II. The second session allowed for the completion of the remaining part of Task II.

10.2.4. Apparatus

We considered the two main autostereoscopic smartphones that were available on the market: an HTC Evo 3D and an LG Optimus 3D Max. Due to incompatibilities in HTC's 3D SDK, an LG Optimus 3D Max was used for the experiment. The smartphone's dimensions are $126.8 \times 67.4 \times 9.6\text{ mm}$ with a 4.3-inch screen, having a resolution of 480×800 pixels. For the experiment, the device was fixed in a frame as depicted in Figure 10.4. With this setup, we ensured a constant distance from the device to the scene of interest as well as a constant distance between viewer and device for a consistent 3D effect. To compensate for individual differences in head size (i.e. length between chin and eyes) and body height, the chin rest, the chair and both tables could be adjusted in height. At the beginning of each task it was ensured that the participants' eyes were at the right height during the task. In addition, the participants were asked to maintain a constant seated position during the task. For all situations we ensured that the relative difference in height between the two tables remained constant. The device was always mounted at the same height on the first table. Furthermore, it was ensured that the same illumination conditions were used for all participants.

We augmented the device's camera image showing a real-world table with two virtual cubes floating 200 mm above the table as shown in Figure 10.3. All cubes were presented with the same texture. No lighting effects were used, to avoid introducing additional visual cues. Again, in the experimental setup, the real and the virtual space had to be carefully integrated to ensure that no other influences affected the study. This also constrained the available space where objects could be placed at a reasonable size and without touching the display's borders. We therefore chose 700 mm and 1700 mm as the values for NEG-P and POS-P respectively.

10. Handheld Stereoscopic Augmented Reality

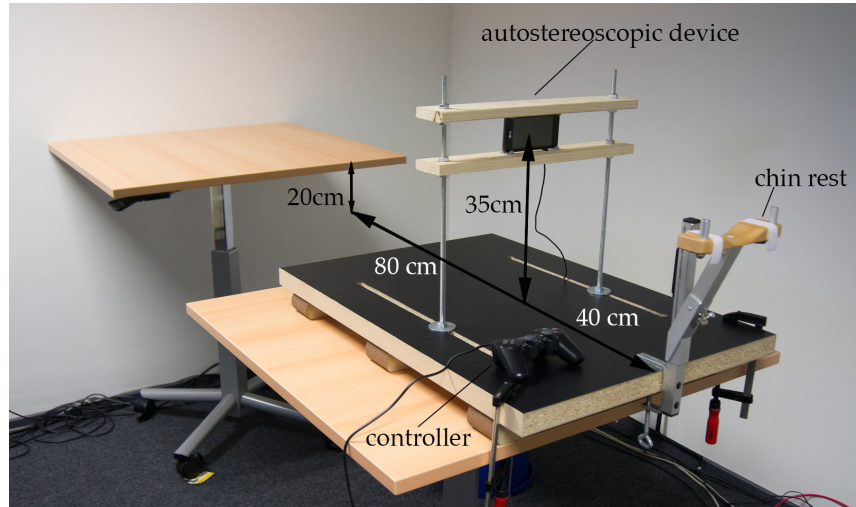


Figure 10.4.: Experimental setup.

10.2.5. Independent and Dependent Variables

Task I considered cube size (large/small) and parallax (NEG-P/POS-P) as the independent variables. Participants were told that both cubes had been placed at the same depth and that they should decide, as the dependent variable, whether they were shown in NEG-P or POS-P and report their choice verbally.

In Task II we varied the depth of the two virtual cubes and the participants had to decide which of the two cubes was closer. With the help of an adaptive stair-casing procedure, we determined the required minimal depth distance between the two virtual cubes for it to be possible to discriminate between them (dependent variable). We considered three independent variables within subjects, namely:

Autostereoscopy (*On*, *Off*): To measure the effect of the autostereoscopic cue we integrated two conditions. In the *On* condition we displayed the camera image as well as the augmented cubes with stereoscopic effects. In the *Off* condition no stereoscopic effects were enabled. Only one of the stereo camera images was used (randomly chosen) together with the virtual cubes.

Size (*Randomized*, *Fixed*): To test the influence of the size cue, we separately varied every cube's edge lengths in the *Randomized* condition, uniformly distributed between 75 mm and 100 mm for every stimulus presentation. In the *Fixed* condition both cubes had a constant size of $87.5 \times 87.5 \times 87.5 \text{ mm}^3$. Additionally, the sizes were adjusted perspectively correctly regarding displayed depth in both conditions.

Object depth (700, 800, 900, 1400, 1500, 1600, 1700 mm):

To check the influence of the cubes' depths (referring to the cubes' front face measured from the camera's position) we considered seven different conditions, the first three being in NEG-P, the latter four in POS-P.

Parallax	Size	Correct answers	Incorrect answers
NEG-P	Large	56	4
NEG-P	Small	33	27
POS-P	Large	10	50
POS-P	Small	52	8

Table 10.1.: Overview of conditions and answers given.

10.3. Results

In the following the experimental results of the two discrimination tasks are reported. The tasks investigated the depth discrimination capability with respect to object size (Task 1) and autostereoscopy and parallax (Task 2).

10.3.1. Object Size

In Task I the rate of correct answers varied between 50% and 75% ($M = 62.9\%$). Table 10.1 shows the distribution of answers in the different conditions. Participants mainly judged cubes to be in NEG-P if displayed large and to be in POS-P if displayed small. An interesting finding is that people were able to better discriminate (i.e. make fewer wrong categorizations) in NEG-P (error rate = 45%) than in POS-P (error rate = 83.3%) despite the “conflicting” size cue. The size variable and the depth that was assumed by the participants were strongly correlated (Pearson’s $r(478) = 0.56$, $p < 0.01$) whereas the real depth and the assumed depth were only slightly correlated (Pearson’s $r(478) = 0.26$, $p < 0.01$).

10.3.2. Autostereoscopy and Parallax

Figure 10.5 shows the results of Task II. The x -axis shows the seven object depths, and the y -axis illustrates the mean minimal depth distance between the cubes’ front faces to enable a discrimination. The four conditions, autostereoscopy *On/Off* and size *Fixed/Randomized*, are shown separately.

To investigate the effect of the different conditions on the minimal depth distance needed between the cubes’ front faces to enable a discrimination, a $2 \times 2 \times 7$ repeated-measure ANOVA was performed. Where Mauchly’s test indicated that the assumption of sphericity had been violated, Greenhouse-Geisser correction was applied. Results indicate significant main effects for object depth ($F(3.08, 33.88) = 30.07$, $p < .001$, $\eta_p^2 = .73$) and for size ($F(1, 11) = 289.42$, $p < .001$, $\eta_p^2 = .96$) with static size having a lower minimal distance ($N = 168$, $M = 25.06$, $SD = 4.06$) than random size ($N = 168$, $M = 102.40$, $SD = 4.66$). No significant main effect for stereo was found ($F(1, 11) = .10$, $p = .76$, $\eta_p^2 = .01$). However, there is also a significant interaction for size and depth ($F(3.45, 38.00) = 15.79$, $p < .001$, $\eta_p^2 = .59$).

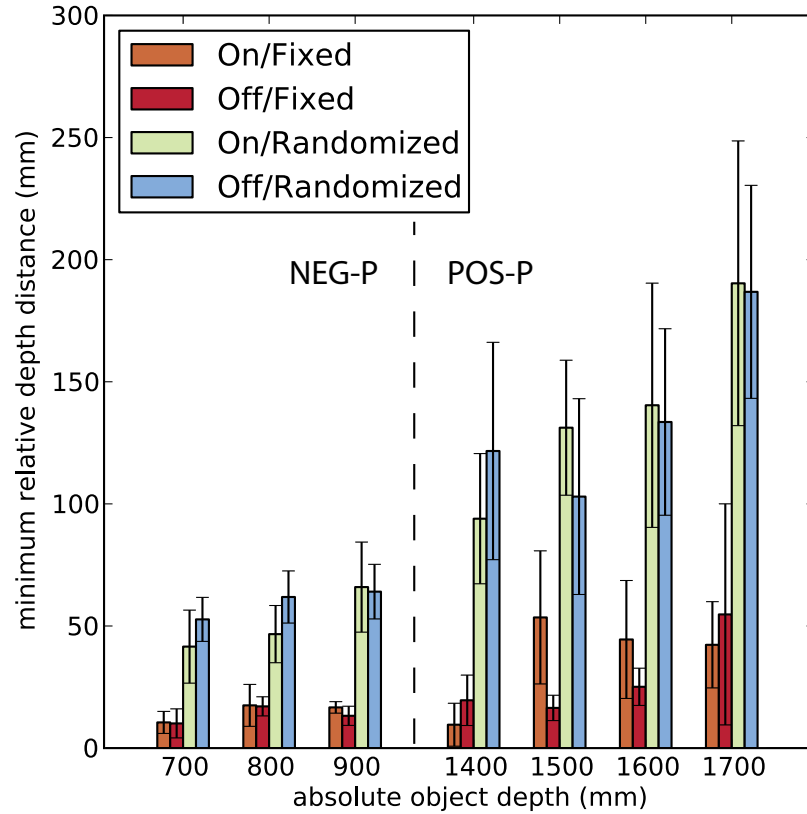


Figure 10.5.: Results of the Task II showing the mean relative depth distances required for discrimination of virtual objects at different depths for the independent variables autostereoscopy and size (error bars represent 95% confidence intervals)

We draw the following conclusions about the data from Task II. First, the minimal required depth increases with increasing absolute depth of the to-be-compared virtual objects. Second, no significant differences in the comparison of the autostereoscopic conditions (*On/Off*) could be found. In other words, autostereoscopy did not change performance in the task. Third, it was also easier for the participants to discriminate between objects in the *Fixed* size conditions. Lastly, it was easier for the participants to discriminate depths in NEG-P, independent of the size and autostereoscopic conditions. This can be seen by comparing the mean minimal relative depth distances between the NEG-P and the POS-P condition in Table 10.3.

10.4. Discussion

Human depth discrimination on a small mobile display using an off-the-shelf autostereoscopic device is poor. With our experiments we learned that in these conditions there is virtually no effect of autostereoscopy on users' ability to distinguish the depth of virtual objects imposed on a real scene.

Parallax	Size	ANOVA
NEG-P	Fixed	$F(1, 34) = 0.50, p = 0.48$
NEG-P	Random	$F(1, 34) = 2.57, p = 0.11$
POS-P	Fixed	$F(1, 46) = 1.23, p = 0.27$
POS-P	Random	$F(1, 46) = 0.03, p = 0.85$

Table 10.2.: Effect of the autostereoscopic conditions (*On/Off*) on the mean relative depth distance in the parallax/size conditions.

Size	Autostereoscopy	ANOVA
Fixed	On	$F(1, 46) = 13.62, p < 0.05$
Fixed	Off	$F(1, 46) = 5.46, p < 0.05$
Random	On	$F(1, 46) = 46.08, p < 0.05$
Random	Off	$F(1, 46) = 40.42, p < 0.05$

Table 10.3.: Effect of the parallax conditions (NEG-P/POS-P) on the mean relative depth distance in the size/autostereoscopic conditions.

Parallax	Autostereoscopy	ANOVA
NEG-P	On	$F(1, 34) = 64.22, p < 0.05$
NEG-P	Off	$F(1, 34) = 228.89, p < 0.05$
POS-P	On	$F(1, 46) = 73.16, p < 0.05$
POS-P	Off	$F(1, 46) = 84.28, p < 0.05$

Table 10.4.: Effect of the size conditions (*Fixed/Randomized*) on the mean relative depth distance in the parallax/autostereoscopic conditions.

10. Handheld Stereoscopic Augmented Reality

In Task II we would have expected a positive effect of autostereoscopy in the *Fixed* size condition, but this did not appear even in the viewing conditions that were calibrated to the user and the device. This can be explained by a lack of an effect with rivalry of cues in the overlaid objects (figure) and the background (camera viewfinder image). We did a pilot test previously with a virtual-only scene and observed that there can be a positive effect of autostereoscopy. Our results suggest that the positive effect disappears in the AR condition when the VR objects are seen superimposed on the viewfinder's image. For the present study, the effect to match the provided camera image to ensure that no perceptual mismatch occurred was carefully adjusted. We believe that it is hard for users to fuse the two representations based on the autostereoscopic cues, and they instead rely on the monoscopic cues. This hypothesis calls for further studies.

The study also sheds light on some underlying perceptual factors. The results of the first task show that people rely more on the object size cue than on the autostereoscopic cue. In Task II we found that the capability to discriminate depths in NEG-P is better than in POS-P. This finding is consistent with the first task in which fewer errors were produced in NEG-P.

This work provides a first contribution to designers of mobile AR applications intending to use autostereoscopic displays. First, present-day autostereoscopic displays are not superior to monoscopic displays with regards to mobile AR applications. In other words, the autostereoscopic cue should not be relied on as a primary cue for depth discrimination. Second, the size cue is dominant over the autostereoscopic cue, so placing large objects at the back should be done carefully, as it can lead to false depth interpretations. Third, the results also indicate that virtual objects of known size can be arranged closer to each other. Task II shows that users are better able to discriminate the depth of objects if they have a fixed size. Fourth, displaying objects in NEG-P works better than in POS-P. However, this is not often suitable for AR applications, as the real environment in the camera view lies behind the mobile device's screen. We used a fixed viewing distance and angle which we calibrated ourselves by trying to maximize the effect. It might be possible that a positive result could be found with some other setting, or if users were able to adjust the viewing angle themselves. Even if that were the case, however, we can conclude that if there is any effect, it is spatially constrained. Spatial constraining is counter-productive in mobile use where the viewing angle and distance change dynamically.

10.5. Conclusion

Handheld devices that are equipped with autostereoscopic displays have the potential to enable users to perceive and interact with stereoscopic data without additional user instrumentation. These stereoscopic devices offer new opportunities but also technological and perceptual challenges. In order to investigate this, we explored different research directions. In this chapter we explored handheld stereoscopic AR and, as a first step towards this scenario, investigated human depth perception in handheld stereoscopic AR in a constrained lab experiment.

The results of this study provide first insights to designers of mobile AR applications intending to use autostereoscopic displays. According to our results, present-day autostereoscopic displays are not superior to monoscopic displays with regards to mobile AR. In other words, the autostereoscopic cue should not be relied on as a primary cue for depth discrimination. Instead, the size cue appears to be dominant over the autostereoscopic cue. Hence, placing large objects at the back should be done carefully as it can lead to false depth perception. The results also indicate that virtual objects of known size can be arranged closer to each other. Users will also benefit from displaying objects in NEG-P rather than POS-P. Although human depth discrimination on current commercial mobile autostereoscopic displays is bad, we believe that stereoscopy can enhance the AR experience by guiding the user with additional cues. This needs to be investigated in future work.

11. Conclusion

11.1. Thesis Summary

This thesis investigated interaction with stereoscopic data on and above interactive surfaces. The review of the background and related work in a variety of fields clearly showed that the foundations in each distinctive field has been laid out in HCI (including mobile HCI, interactive tabletops and surfaces, pervasive displays, etc.), 3DUI (including AR and VR), psychology and many others. However, the challenges that arise when these parts emerge in the space where the flat digital world of surface computing meets the physical, spatially complex, 3D space in which we live have not been considered much before. We started our explorations of this fairly new field with a taxonomy that was iteratively developed and builds on related work to classify all interaction techniques that were investigated in the thesis.

Besides the taxonomy that focuses on input (interaction primitives, modalities and DOF), we aimed to get deeper insight into 3DUIs by investigating 3D interaction with stereoscopic surfaces for the canonical 3DUI tasks. In particular mobile and gestural interaction techniques were designed and evaluated for the selection and manipulation of 3D objects and the navigation in the virtual 3D space.

With these lessons learned we then extended the interaction space of 3DUI and 3D interaction by discussing research probes that go beyond classical approaches in HCI, AR and VR. The concept of *interaction context for multi-touch 3D interaction* was explored and applied to what we call the *Reach to Grasp interaction* for multi-touch interactive 3D surfaces. Finally, two approaches explored handheld devices that support 3D input as well as output. First, 3D interactions in VR were investigated on stereoscopic mobile devices. Second, 3D perception in handheld stereoscopic AR was studied.

11.2. Contributions

This dissertation work contributes usable and natural interaction techniques and UI concepts for interactive surfaces for the interaction with stereoscopic data. It further contributes insights into human factors of interactive stereoscopic devices.

We investigated interaction with interactive stereoscopic displays for canonical 3DUI tasks to gain insights about human aspects when touch and gestural input is combined with stereoscopic output, with special consideration of commodity devices. Our studies have shown that usable interactions can be designed for the interaction with stereoscopically displayed 3D data. The perception of objects displayed with different parallax seems to

11. Conclusion

depend on display size and/or 3D output technology. Indirect multi-touch selection could be performed better in the positive parallax space, while depth discrimination in handheld stereoscopic AR performed better in the negative parallax space. We further showed that stereoscopic 3D output outperforms monoscopically rendered 3D, especially for complex spatial interactions.

Studying research probes that extend the interaction space led to promising results even on commodity 3D devices. In our first probe, we extended the interaction space of 3D multi-touch interaction from the flat surface to the space above the device. By implicitly using the space above the interactive surface, the actual multi-touch interaction concepts were mainly preserved but enriched with additional context information. The context allows richer interactions, although fewer DOF of multi-touch input are available. The knowledge about what (i.e. which virtual object) the grasping hand is reaching for can be used even in non-stereoscopic environments for example in 2D map applications, (interactive) information visualizations, or even classic GUIs.

The extension of VR and AR to stereoscopic mobile devices has proved its general applicability and usability. However, a number of perceptual issues have been identified that occur particularly due to the small display size. As one important outcome of this work we have identified the issues that need to be carefully addressed when designing 3D interactions for handheld VR and AR. Our proposed interaction concepts for 3D interaction on mobile devices using touch and motion sensors proved to work in a mobile 3D scenario, and sensor-based mobile 3D interaction, when carefully designed, provides an intuitive and joyful means of interaction. The results of our study on handheld stereoscopic AR provide first insights for the design of mobile AR applications for autostereoscopic displays. Although human depth discrimination on state-of-the-art commercial mobile autostereoscopic displays is bad, we believe that stereoscopy can enhance the AR experience by guiding the user with additional cues.

11.3. Future Work

The research that has been conducted in the scope of this thesis provides insights into novel interactions for device classes ranging from mobile devices to large displays that will emerge in the coming years. While stereoscopic 3D has become very popular in recent years, the development of interactive stereoscopic devices in particular autostereoscopic technologies, is advancing rapidly. Based on that research many different future research directions can be followed.

The proposed taxonomy consists of different dimensions that can be filled with appropriate tasks, devices and input modalities, and IPs. It enables designers to easily and consistently integrate new modalities for 3DUI. Initial evaluations of the interaction concepts showed the viability of this approach. However there is also a need for further refinements as well as additional investigation on other input modalities. Furthermore we want to address the problems outlined in the evaluations. User adopted as well as user generated gestures also have to be investigated in detail. With the rise of smartwatches

the idea of wearables as remote input device might now have good potential and should be revisited with commodity wearable devices. The use of these interaction techniques for 3DUI can be also further explored.

Designing applications for interactive stereoscopic devices in the living room will be one big challenge. Our research on universal 3DUI tasks with commodity input devices makes a first step towards the question of how users can interact with 3D data in such non-expert 3D environments. While 3D output technology such as 3D television is already available, interaction with such devices has barely been considered so far, and interactive applications that are specifically designed for such stereoscopic displays are still missing.

Our research has further shown the applicability of interactive handheld stereoscopic devices and also outlined potential scenarios for these devices such as mobile 3D gaming or the navigation in complex data (e.g. geospatial or medical data, databases, etc.). A potential application idea that emerge from feedback in our studies is browsing rich content on stereoscopic displays such as navigating in a 3D movie (or music) database.

We believe that handheld stereoscopic AR is another promising design space for mobile stereoscopic devices, and we believe that stereoscopy can enhance the AR experience by guiding the user with additional cues. However, autostereoscopic displays need to be researched and developed that better support human depth discrimination. Currently the capabilities of off-the-shelf handheld stereoscopic devices do not improve depth discrimination in AR. In future work, the factors affecting depth discrimination in conditions that involve free user movement in the scene need to be studied. We hypothesize that, should there be any effect, it will be spatially constrained. Spatial constraining is counter-productive in mobile use where the viewing angle and distance change dynamically. Further experiments need to investigate other depth cues as well as scenarios that integrate and augment real and virtual objects.^f

Promising new sensors such as mobile depth sensors and projects such as the Google Tango project¹ might lead to improved approaches in handheld AR that would provide precise registration of augmentations in the real world. Stereoscopic handheld AR will also profit from these developments because they might reduce cognitive implausible effects in the stereoscopic display.

Besides handheld stereoscopic technologies, large 3D displays that go beyond classical VR setups, for example public displays, are now possible in practice as well. Since passers-by are expected to interact with such public stereoscopic displays these setups require intuitive means of interaction like those proposed in this thesis. The results of our 3D travel task study give implications for the design of intuitive 3D interaction techniques that might enable spatial interactions in public places. One potential scenario is 3D gaming with stereoscopic public displays and media facades. The physical interaction techniques are very suitable candidates for such a scenario. The general drawback of physical demand and effort might even increase the complexity and thus the gaming experience.

In general, a standardized docking task to evaluate 3D input devices and interaction techniques would be desirable, like the ISO9241-9 [99] standard for selection tasks. We

¹<https://www.google.com/atap/projecttango/>

11. Conclusion

will address this issue in further studies with a focus on metrics that integrate translation and rotation. In particular the problem of 3D rotation needs to be studied in more depth because this is a critical issue not only in 3D docking tasks but also in 3D navigation tasks. In order to study this aspect, additional interaction techniques and input devices need to be taken into consideration.

11.4. Concluding Remarks

This work addressed interaction with stereoscopic data on and above interactive surfaces, and we showed that touching the third dimension is indeed possible, even in the virtual world and even without turning everything into gold like King Midas. However, Midas reminds us that interactions, in particular interactions with 3D virtual objects, need to be designed carefully and in a human-centered way. Although virtual objects are not physical, virtual environments often provide a variety of cues from the real world, such as depth perception but also affordance. Such cues guide human perception and behavior and need to be carefully considered when designing interactions for the next generation of UIs.

Appendices

List of Figures

1.1. Interaction with stereoscopically displayed geo-spatial data on a multi-touch surface with anaglyph display (left). Mobile and gestural 3D interaction with a large stereoscopic display (right).	18
2.1. Monoscopic depth cues [21].	28
2.0. Monoscopic depth cues [21].	29
2.1. Stereoscopic depth cues: accommodation and convergence.	30
2.2. Accommodation-convergence conflict.	31
2.3. Parallax spaces: On stereoscopic displays objects may appear in front of (negative parallax), on top of (zero parallax), or behind (positive parallax) the screen.	32
2.4. The problem of touching stereoscopically objects [197].	33
2.5. Early examples of AR systems.	34
2.6. The reality-virtuality continuum by Milgram and Kishino [147].	35
2.7. The problem of depth perception in AR.	36
2.8. Stereo glasses: active shutter glasses (left), passive polarized glasses (middle), passive spectral (anaglyph) glasses (right).	37
2.9. Location multiplexing examples.	38
2.10. Projected 3D Displays.	39
2.11. HMD examples.	40
2.12. 3D technology that was used in this research: passive and active stereo glasses as well as autostereoscopic displays were used in combination with multi-touch and gestural tracking technologies.	42
2.13. 6-DOF Input Taxonomy by Zhai and Milgram [216]	43
2.14. Desktop input devices.	44
2.15. Early 3D mice.	45
2.16. Speech input examples.	46
2.17. Examples for multi-touch stereoscopic devices.	47
2.18. Examples of commodity 3D input devices.	48
2.19. 3D selection tasks.	50
2.20. Simple docking task examples.	52
2.21. Travel task examples.	54
2.22. 6-DOF multi-touch interaction examples.	57
2.23. Multi-touch 3D selection.	58
2.24. Prehensile movements of the human hand [151].	59
2.25. Hand postures in the <i>Reach to Grasp</i> phase.	60

List of Figures

2.26. Gaze-based interaction.	61
2.27. Handheld 3D interaction.	63
3.1. Multi-touch GIS Interaction.	66
3.2. Set of physical multi-touch gestures.	67
3.3. Physical foot gestures.	70
3.4. Main IPs for the interaction with stereoscopic data on and above the interactive surface: touch, mid-air and tangible IPs	72
4.1. Interaction with stereoscopic data on a multi-touch surface with anaglyph display (with the <i>Balloon/Fishnet Selection</i> technique).	80
4.2. The original version of the <i>Balloon Selection</i> technique [6]: users wear pinch gloves and perform gestures on a multi-touch tabletop. a) Instantiation of the balloon; b-c) Stretching the string and raising the balloon; d-e) Pinch to scale the balloon; f) Select an object.	81
4.3. Initialization of the <i>Balloon/Fishnet Selection</i> tool.	82
4.4. Controlling the <i>Balloon/Fishnet Selection</i> pointer.	82
4.5. Controlling the <i>Corkscrew Selection</i> tool.	83
5.1. Mobile and gestural interaction with stereoscopic 3D user interface in a docking task. The left hand (NDH) toggles the object selection with a grip gesture. The right hand (DH) controls the 6-DOF object manipulation. The input is captured with the Kinect and the orientation sensors of a mobile device.	92
5.2. Experiment setup conditions and tasks (c.f. [179]).	93
5.3. Task completion time for position and orientation in the monoscopic (blue) and stereoscopic (red) conditions [179].	97
5.4. Overall translation task precision in the monoscopic (blue) and stereoscopic (red) conditions of the eight target positions [179].	97
5.5. Translation task precision for the eight target positions by axis (x, y, z → blue, red, green) [179].	98
5.6. Rotation task precision regarding the four target rotations (A,B,C: simple; D: complex) [179].	99
5.7. NASA TLX results: The blue color indicates the results for the monoscopic condition, the red color for the stereoscopic condition. The gray color indicates the overall results independent of display mode [179].	100
6.1. Physical travel techniques.	105
6.2. Mobile gestures for virtual camera positioning.	106
6.3. Mobile gestures for virtual camera orientation.	107
6.4. Mobile tilt gestures for virtual camera positioning.	108
6.5. The user's perspective on the 3D scene consisting of grids of tetrahedrons and textured faces.	111
6.6. Boxplots of the execution time in seconds w.r.t. interaction technique. . . .	113

6.7.	The boxplots of execution time grouped by interaction technique.	114
6.8.	Participants' averages concerning the four interaction techniques from NASA TLX.	115
6.9.	Overall workload of all four interaction techniques from NASA TLX. . . .	115
7.1.	Interaction context in the user feedback loop: the UI is affected by both explicit and implicit user input.	122
7.2.	Interaction context in the user feedback loop: position, orientation and posture of the user's head, hands and fingers.	124
8.1.	Design concept of a multi-touch enabled stereoscopic surface equipped with additional depth sensors that can predict the user's intention during grasping movements.	126
8.2.	Experimental setup for the study: table that serves as hand rest and a tilted transparent projection screen as interactive surface.	127
8.3.	Available candidates for hand properties: Number of fingers; use of specific finger(s); opening (boundary) between the fingers; moving direction of the grasping hand; moving speed of the grasping hand.	129
8.4.	Experiment setup and task.	131
8.5.	Sample interface widgets that had to be grasped by subjects during the study.	132
8.6.	Feature extraction pipeline.	133
8.7.	Grasp variations.	137
9.1.	Sensor-based interaction with stereoscopically displayed data on a mobile device.	146
9.2.	Object manipulation techniques.	147
9.3.	Travel techniques.	148
9.4.	Layered 3D world with Ground, Sky1 and Sky2 layers.	150
9.5.	The 4-dist graph: the 5th degree polynomial represents the trend in the data as a smoother curve. Thus the thresholds are obviously deducible. The first threshold delimits the outliers or the noise while the second differentiates the different clusters [126].	153
9.6.	The touches clustered with DBSCAN before (left) and after (right) clustering. The starting condition on the right shows all touches that initiate a manipulation (in blue) and the target position at the start of the runway (red). The clustering results in a high-density cluster of the failed landing attempts and the low-density cluster of other touches [126].	154
10.1.	Stereoscopic Handheld AR Concept: Object-referring augmentations are displayed in depth (1-3). This enables "free floating" augmentations that are easily comprehensible (1).	160

List of Figures

10.2. Stereo Handheld AR Demo as proof-of-concept: A virtual cube is projected on a real table. The virtual object can be interactively explored by moving the device is moved around it like a magnifying lens. While the viewpoint of the device can be changed the cube remains in a stable position at the corner of the table at all time.	162
10.3. Autostereoscopic device showing augmented camera images with two relative depth differences (left: 10 <i>mm</i> , right: 104 <i>mm</i>)	164
10.4. Experimental setup.	166
10.5. Results of the Task II showing the mean relative depth distances required for discrimination of virtual objects at different depths for the independent variables autostereoscopy and size (error bars represent 95% confidence intervals)	168

Acronyms

3D three-dimensional.....	17
3DUI 3D user interface.....	3
ANOVA ANalysis Of VAriance.....	86
AR augmented reality.....	20
DH dominant hand.....	52
DI diffuse illumination.....	46
DOF degrees of freedom.....	34
FOR field of regard.....	39
FOV field of view.....	39
fps frames per second.....	132
GIS geographic information system.....	65
GPS global positioning system.....	34
GUI graphical user interface.....	50
HCI human-computer interaction.....	18
HMD head-mounted display.....	35

List of Figures

IP interaction primitive	65
MR mixed reality	34
NDH non-dominant hand	74
NUI Natural user interface	18
PCA principal component analysis	134
RST rotation, scale and translation.....	57
UI user interface	17
VE virtual environment	17
VR virtual reality	3
WIM world-in-miniature	46
WIMP windows, icons, menus and pointer	17

Bibliography

- [1] T. Augsten, K. Kaefer, R. Meusel, C. Fetzner, D. Kanitz, T. Stoff, T. Becker, C. Holz, and P. Baudisch. Multitoe: High-precision interaction with back-projected floors based on high-resolution multi-touch input. In *Proceedings of the 23Nd Annual ACM Symposium on User Interface Software and Technology*, UIST '10, pages 209–218, New York, NY, USA, 2010. ACM.
- [2] R. T. Azuma. A survey of augmented reality. *Presence: Teleoperators and Virtual Environments*, 6(4):355–385, aug 1997.
- [3] R. Balakrishnan, G. W. Fitzmaurice, and G. Kurtenbach. User interfaces for volumetric displays. *Computer*, 34(3):37–45, 2001.
- [4] R. Ballagas, J. Borchers, M. Rohs, and J. G. Sheridan. The smart phone: A ubiquitous input device. *IEEE Pervasive Computing*, 5(1):70–77, Jan. 2006.
- [5] M. Barz. Computational modeling and prediction of gaze estimation error for head-mounted eye trackers. Master’s thesis, Saarland University, Saarbrücken, Germany, Department of Computer Science, 2015.
- [6] H. Benko and S. Feiner. Balloon selection: A multi-finger technique for accurate low-fatigue 3d selection. In *IEEE Symposium on 3D User Interfaces*, 3DUI '07. IEEE, 2007.
- [7] A. Benzina, A. Dey, M. Tonniss, and G. Klinker. Empirical evaluation of mapping functions for navigation in virtual environments using phones with integrated sensors. *International Journal of Innovative Computing, Information and Control (IJICIC)*, 9(12):4693–4709, Dec. 2013.
- [8] A. Benzina, M. Toennis, G. Klinker, and M. Ashry. Phone-based motion control in vr: analysis of degrees of freedom. In *Proceedings of the 2011 annual conference extended abstracts on Human factors in computing systems*, CHI EA '11, pages 1519–1524, New York, NY, USA, 2011. ACM.
- [9] R. Biedert, G. Buscher, and A. Dengel. The eye book. *Informatik-Spektrum*, 2009.
- [10] E. A. Bier, M. C. Stone, K. Pier, W. Buxton, and T. D. DeRose. Toolglass and magic lenses: the see-through interface. In *SIGGRAPH '93: Proceedings of the 20th annual conference on Computer graphics and interactive techniques*, pages 73–80. ACM, 1993.

Bibliography

- [11] B. Blundell and A. Schwarz. *Volumetric Three-dimensional Display Systems*. John Wiley & Sons, Incorporated, 2000.
- [12] B. G. Blundell. *An Introduction to Computer Graphics and Creative 3-D Environments*. Springer Publishing Company, Incorporated, 1 edition, 2008.
- [13] A. Boev and A. Gotchev. Comparative study of autostereoscopic displays for mobile devices. In C. G. M. S. N. S. L. K. D. A. R. Creutzburg, editor, *Multimedia on Mobile Devices 2011; and Multimedia Content Access: Algorithms and Systems V*, volume 7881 of *SPIE Proceedings*. SPIE, 2011.
- [14] R. A. Bolt. “put-that-there”: Voice and gesture at the graphics interface. In *Proceedings of the 7th annual conference on Computer graphics and interactive techniques*, SIGGRAPH '80, pages 262–270, New York, NY, USA, 1980. ACM.
- [15] S. Boring, M. Jurmu, and A. Butz. Scroll, tilt or move it: using mobile phones to continuously control pointers on large public displays. In *OZCHI '09: Proceedings of the 21st Annual Conference of the Australian Computer-Human Interaction Special Interest Group: Design: Open 24/7*, pages 161–168. ACM, 2009.
- [16] J. Boritz and K. Booth. A study of interactive 6 dof docking in a computerised virtual environment. In *Virtual Reality Annual International Symposium, 1998. Proceedings., IEEE 1998*, pages 139–146, 1998.
- [17] J. Boritz and K. S. Booth. A study of interactive 3d point location in a computer simulated virtual environment. In *Proceedings of the ACM Symposium on Virtual Reality Software and Technology*, VRST '97, pages 181–187, New York, NY, USA, 1997. ACM.
- [18] D. A. Bowman and L. F. Hodges. An evaluation of techniques for grabbing and manipulating remote objects in immersive virtual environments. In *Proceedings of the 1997 Symposium on Interactive 3D Graphics*, I3D '97, pages 35–ff., New York, NY, USA, 1997. ACM.
- [19] D. A. Bowman, D. B. Johnson, and L. F. Hodges. Testbed evaluation of virtual environment interaction techniques. In *Proceedings of the ACM Symposium on Virtual Reality Software and Technology*, VRST '99, pages 26–33, New York, NY, USA, 1999. ACM.
- [20] D. A. Bowman, D. Koller, and L. F. Hodges. Travel in immersive virtual environments: An evaluation of viewpoint motion control techniques. In *Proceedings of the 1997 Virtual Reality Annual International Symposium*, VRAIS '97, pages 45–52, Washington, DC, USA, 1997. IEEE Computer Society.
- [21] D. A. Bowman, E. Kruijff, J. J. LaViola, and I. Poupyrev. *3D User Interfaces: Theory and Practice*. Addison Wesley Longman Publishing Co., Inc., Redwood City, CA, USA, 2004.

- [22] D. A. Bowman and C. A. Wingrave. Design and evaluation of menu systems for immersive virtual environments. In *Proceedings of the Virtual Reality 2001 Conference*, VR '01, pages 149–156, Washington, DC, USA, 2001. IEEE Computer Society.
- [23] A. Bränzel, C. Holz, D. Hoffmann, D. Schmidt, M. Knaust, P. Lühne, R. Meusel, S. Richter, and P. Baudisch. Gravityspace: Tracking users and their poses in a smart room using a pressure-sensing floor. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '13, pages 725–734, New York, NY, USA, 2013. ACM.
- [24] G. Bruder, F. Steinicke, and W. Sturzlinger. To touch or not to touch?: Comparing 2d touch and 3d mid-air interaction on stereoscopic tabletop surfaces. In *Proceedings of the 1st Symposium on Spatial User Interaction*, SUI '13, pages 9–16, New York, NY, USA, 2013. ACM.
- [25] A. Bulling and H. Gellersen. Toward mobile eye-based human-computer interaction. *IEEE Pervasive Computing*, 9(4):8–12, 2010.
- [26] W. Buxton, E. Fiume, R. Hill, A. Lee, and C. Woo. Continuous hand-gesture driven input. In *Proceedings of Graphics Interface '83*, pages 191–195, May 1983.
- [27] W. Buxton and B. Myers. A study in two-handed input. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '86, pages 321–326, New York, NY, USA, 1986. ACM.
- [28] T. Capin, A. Haro, V. Setlur, and S. Wilkinson. Camera-based virtual environment interaction on mobile devices. In A. Levi, E. Savas, H. Yenigün, S. Balcişoy, and Y. Saygin, editors, *Computer and Information Sciences - ISCIS 2006*, volume 4263 of *Lecture Notes in Computer Science*, pages 765–773. Springer Berlin Heidelberg, 2006.
- [29] S. K. Card, J. D. Mackinlay, and G. G. Robertson. A morphological analysis of the design space of input devices. *ACM Transaction on Information Systems*, 9(2):99–122, Apr. 1991.
- [30] M. Chen, S. J. Mountford, and A. Sellen. A study in interactive 3-d rotation using 2-d control devices. In *Proceedings of the 15th annual conference on Computer graphics and interactive techniques*, SIGGRAPH '88, pages 121–129, New York, NY, USA, 1988. ACM.
- [31] A. Chernov. A method for 3d reconstruction of a foot with kinect. Bachelor's thesis, Saarland University, Saarbrücken, Germany, Department of Computer Science, 2014.
- [32] S. Chieffi and M. Gentilucci. Coordination between the transport and the grasp components during prehension movements. *Experimental Brain Research*, 94:471–477, 1993. 10.1007/BF00230205.

Bibliography

- [33] K. C. Clarke. *Analytical and computer cartography*, volume 290. Prentice Hall Englewood Cliffs (NJ), 1990.
- [34] D. Coffey, F. Korsakov, and D. F. Keefe. Low cost vr meets low cost multi-touch. In *Proceedings of the 6th International Conference on Advances in Visual Computing - Volume Part II*, ISVC'10, pages 351–360, Berlin, Heidelberg, 2010. Springer-Verlag.
- [35] D. Coffey, N. Malbraaten, T. Le, I. Borazjani, F. Sotiropoulos, and D. F. Keefe. Slice wim: a multi-surface, multi-touch interface for overview+detail exploration of volume datasets in virtual reality. In *Symposium on Interactive 3D Graphics and Games*, I3D '11, pages 191–198, New York, NY, USA, 2011. ACM.
- [36] A. Cohé, F. Dècle, and M. Hachet. tbox: a 3d transformation widget designed for touch-screens. In *Proceedings of the 2011 annual conference on Human factors in computing systems*, CHI '11, pages 3005–3008, New York, NY, USA, 2011. ACM.
- [37] J. V. Cohn, J. Feasel, S. J. Poulton, B. McLeod, and F. P. Brooks, Jr. Comparing ve locomotion interfaces. In *Proceedings of the 2005 IEEE Conference 2005 on Virtual Reality*, VR '05, pages 123–130, Washington, DC, USA, 2005. IEEE Computer Society.
- [38] A. Colley, J. Hkkil, J. Schning, and M. Posti. Investigating mobile stereoscopic 3d touchscreen interaction. In *Proceedings of the 25th Australian Computer-Human Interaction Conference*, OzCHI '13, New York, NY, USA, 2013. ACM.
- [39] B. D. Conner, S. S. Snibbe, K. P. Herndon, D. C. Robbins, R. C. Zeleznik, and A. van Dam. Three-dimensional widgets. In *Proceedings of the 1992 Symposium on Interactive 3D Graphics*, I3D '92, pages 183–188, New York, NY, USA, 1992. ACM.
- [40] C. Cruz-Neira, D. J. Sandin, T. A. DeFanti, R. V. Kenyon, and J. C. Hart. The cave: Audio visual experience automatic virtual environment. *Communications of the ACM*, 35(6):64–72, June 1992.
- [41] R. Dachsel and A. Hübner. A survey and taxonomy of 3d menu techniques. In *Proceedings of the 12th Eurographics Conference on Virtual Environments*, EGVE'06, pages 89–99, Aire-la-Ville, Switzerland, Switzerland, 2006. Eurographics Association.
- [42] F. Daiber. Interaction with stereoscopic data on and above multi-touch surfaces. In *Proceedings of the ACM International Conference on Interactive Tabletops and Surfaces*, ITS '11, pages 2:1–2:1, New York, NY, USA, 2011. ACM.
- [43] F. Daiber. 3d interaction on and above the surface. In *Dagstuhl-Seminar Report, 12151. Schloß Dagstuhl*, Schloß Dagstuhl, Germany, 2012.
- [44] F. Daiber, B. R. De Araujo, F. Steinicke, and W. Stuerzlinger. Interactive Surfaces for Interaction with Stereoscopic 3D (ISIS3D): Tutorial and Workshop at ITS 2013.

- In *Proceedings of the 2013 ACM international conference on Interactive tabletops and surfaces*, ITS '13, pages 483–486, New York, NY, USA, 2013. ACM.
- [45] F. Daiber, E. Falk, and A. Krüger. Balloon Selection revisited - Multi-touch Selection Techniques for Stereoscopic Data. In *Proceedings of the International Conference on Advanced Visual Interfaces*, AVI '12, pages 441–444, New York, NY, USA, 2012. ACM.
 - [46] F. Daiber, S. Gehring, M. Löchtefeld, and A. Krüger. Touchposing - multi-modal interaction with geospatial data. In *Proceedings of the 11th International Conference on Mobile and Ubiquitous Multimedia*, MUM '12, New York, NY, USA, 2012. ACM.
 - [47] F. Daiber, F. Kosmalla, and A. Krüger. Boulder: Using augmented reality to support collaborative boulder training. In *CHI '13 Extended Abstracts on Human Factors in Computing Systems*, CHI EA '13, pages 949–954, New York, NY, USA, 2013. ACM.
 - [48] F. Daiber, F. Kosmalla, M. Löchtefeld, S. Gehring, and A. Krüger. Handheld augmented reality for collaborative boulder training. In *Proceedings of ACM CHI Workshop: HCI and Sports*, 2014.
 - [49] F. Daiber, A. Krekhov, M. Speicher, J. Krüger, and A. Krüger. A framework for prototyping and evaluation of sensor-based mobile interaction with stereoscopic 3d. In *Proceedings of ACM ITS Workshop on Interactive Surfaces for Interaction with Stereoscopic 3D (ISIS3D)*, pages 13–16, 2013.
 - [50] F. Daiber, A. Krüger, J. Schöning, and J. Müller. Context-sensitive display environments. In A. Krüger and T. Kuflik, editors, *Ubiquitous Display Environments*, Cognitive Technologies, pages 31–51. Springer Berlin Heidelberg, 2012.
 - [51] F. Daiber, L. Li, and A. Krüger. Designing gestures for mobile 3d gaming. In *Proceedings of the 11th International Conference on Mobile and Ubiquitous Multimedia*, MUM '12, New York, NY, USA, 2012. ACM.
 - [52] F. Daiber, J. Schöning, and A. Krüger. Whole body interaction with geospatial data. In A. B. B. F. . M. Christie, editor, *Smart Graphics. Spain*, volume 5531/2009, pages 81–92. Springer, 2009.
 - [53] F. Daiber, J. Schöning, and A. Krüger. Towards a framework for whole body interaction with geospatial data. In D. England, editor, *Whole Body Interaction*, Human-Computer Interaction Series, pages 197–207. Springer London, 2011.
 - [54] F. Daiber, M. Speicher, S. Gehring, M. Löchtefeld, and A. Krüger. Interacting with 3d content on stereoscopic displays. In *Proceedings of The International Symposium on Pervasive Displays*, PerDis '14, pages 32:32–32:37, New York, NY, USA, 2014. ACM.

Bibliography

- [55] F. Daiber, D. Valkov, F. Steinicke, A. Krüger, and K. H. Hinrichs. Towards Object Prediction based on Hand Postures for Reach to Grasp Interaction. In *CHI 2012 Workshop on Touching the 3rd Dimension of CHI: Touching and Designing 3D User Interfaces*, pages 99–106, 2012.
- [56] B. R. De Araùjo, G. Casiez, and J. A. Jorge. Mockup builder: Direct 3d modeling on and above the surface in a continuous interaction space. In *Proceedings of Graphics Interface 2012*, GI '12, pages 173–180, Toronto, Ont., Canada, Canada, 2012. Canadian Information Processing Society.
- [57] J. B. De la Rivière, N. Dittlo, E. Orvain, C. Kervégant, M. Courtois, and T. Da Luz. 3d touch: A multiview multitouch surface for 3d content visualization and viewpoint sharing. In *ACM International Conference on Interactive Tabletops and Surfaces*, ITS '10, pages 312–312, New York, NY, USA, 2010. ACM.
- [58] J.-B. de la Rivière, C. Kervégant, E. Orvain, and N. Dittlo. Cubtile: A multi-touch cubic interface. In *Proceedings of the 2008 ACM Symposium on Virtual Reality Software and Technology*, VRST '08, pages 69–72, New York, NY, USA, 2008. ACM.
- [59] F. Decle and M. Hachet. A study of direct versus planned 3d camera manipulation on touch-based mobile phones. In *Proceedings of the 11th International Conference on Human-Computer Interaction with Mobile Devices and Services*, MobileHCI '09, pages 32:1–32:5, New York, NY, USA, 2009. ACM.
- [60] A. Dey, A. Cunningham, and C. Sandor. Evaluating depth perception of photorealistic mixed reality visualizations for occluded objects in outdoor environments. In *Proceedings of the 17th ACM Symposium on Virtual Reality Software and Technology*, VRST '10, pages 211–218, New York, NY, USA, 2010. ACM.
- [61] A. Dey, G. Jarvis, C. Sandor, and G. Reitmayr. Tablet versus phone: Depth perception in handheld augmented reality. In *Proceedings of the 2012 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*, ISMAR '12, pages 187–196, Washington, DC, USA, 2012. IEEE Computer Society.
- [62] P. Dietz and D. Leigh. Diamondtouch: A multi-user touch technology. In *Proceedings of the 14th Annual ACM Symposium on User Interface Software and Technology*, UIST '01, pages 219–226, New York, NY, USA, 2001. ACM.
- [63] P. Domingos and M. Pazzani. On the optimality of the simple bayesian classifier under zero-one loss. *Machine Learning*, 29(2-3):103–130, 1997.
- [64] A. T. Duchowski. A breadth-first survey of eye-tracking applications. *Behavior Research Methods, Instruments, & Computers*, 34(4):455–470, November 2002.
- [65] M. Ester, H. P. Kriegel, J. Sander, and X. Xu. A density-based algorithm for discovering clusters in large spatial databases with noise. In E. Simoudis, J. Han, and

- U. Fayyad, editors, *Proceedings of the Second International Conference on Knowledge Discovery and Data Mining*, pages 226–231, Portland, Oregon, 1996. AAAI Press.
- [66] K. M. Fairchild, B. H. Lee, J. Loo, H. Ng, and L. Serra. The heaven and earth virtual reality: Designing applications for novice users. In *Proceedings of the 1993 IEEE Virtual Reality Annual International Symposium*, VRAIS '93, pages 47–53, Washington, DC, USA, 1993. IEEE Computer Society.
 - [67] E. Falk. Multi-touch selection techniques for stereoscopic 3d content. Master's thesis, Saarland University, Saarbrücken, Germany, Department of Computer Science, 2011.
 - [68] S. Feiner, B. MacIntyre, T. Hollerer, and A. Webster. A touring machine: Prototyping 3d mobile augmented reality systems for exploring the urban environment. In *Proceedings of the 1st IEEE International Symposium on Wearable Computers*, ISWC '97, pages 74–, Washington, DC, USA, 1997. IEEE Computer Society.
 - [69] P. M. Fitts. The information capacity of the human motor system in controlling the amplitude of movement. *Journal of experimental psychology*, 47(6):381–391, June 1954.
 - [70] P. Flotho. Persisten user identification with the kinect. Bachelor's thesis, Saarland University, Saarbrücken, Germany, Department of Computer Science, 2013.
 - [71] J. D. Foley. Interfaces for advanced computing. *Scientific American*, 257(4):126–135, Oct. 1987.
 - [72] A. Freund. Mobicube: A novel approach to 3d menus on mobile devices - a comparative study on 2d vs. 3d mobile menus. Bachelor's thesis, Saarland University, Saarbrücken, Germany, Department of Computer Science, 2015.
 - [73] B. Fröhlich and J. Plate. The cubic mouse: A new device for three-dimensional input. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '00, pages 526–531, New York, NY, USA, 2000. ACM.
 - [74] S. Gehring, F. Daiber, and C. Lander. Towards universal, direct remote interaction with distant public displays. In *AVI 2012: Workshop on infrastructure and design challenges of coupled display visual interfaces (PPD '12)*, 2012.
 - [75] S. Gehring, M. Löchtefeld, F. Daiber, M. Böhmer, and A. Krüger. Using intelligent natural user interfaces to support sales conversations. In *Proceedings of the 2012 ACM international conference on Intelligent User Interfaces*, IUI '12, pages 97–100, New York, NY, USA, 2012. ACM.
 - [76] J. J. Gibson. The theory of affordances. In R. E. Shaw and J. Bransford, editors, *Perceiving, Acting, and Knowing: Toward an Ecological Psychology*. Lawrence Erlbaum Associates, Hillsdale, NJ, 1977.

Bibliography

- [77] GISIG. *Guidelines for Best Practice in User Interface for GIS*. BEST-GIS. ES-PRIT/ESSI project. Geographical Information Systems International Group, 1998.
- [78] J. Grosjean and S. Coquillart. Command & control cube: A shortcut paradigm for virtual environments. In *Proceedings of the 7th Eurographics Conference on Virtual Environments & 5th Immersive Projection Technology*, EGVE'01, pages 1–12, Aire-la-Ville, Switzerland, Switzerland, 2001. Eurographics Association.
- [79] T. Grossman and R. Balakrishnan. Pointing at trivariate targets in 3d environments. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '04, pages 447–454, New York, NY, USA, 2004. ACM.
- [80] T. Grossman and D. Wigdor. Going deeper: a taxonomy of 3d on the tabletop. In *Proceedings of the Second Annual IEEE International Workshop on Horizontal Interactive Human-Computer Systems*, TABLETOP '07, pages 137–144. IEEE Computer Society, 2007.
- [81] T. Grossman, D. Wigdor, and R. Balakrishnan. Multi-finger gestural interaction with 3d volumetric displays. In *Proceedings of the 17th Annual ACM Symposium on User Interface Software and Technology*, UIST '04, pages 61–70, New York, NY, USA, 2004. ACM.
- [82] M. Hachet, B. Bossavit, A. Cohé, and J.-B. de la Rivière. Toucheo: Multitouch and stereo combined in a seamless workspace. In *Proceedings of the 24th Annual ACM Symposium on User Interface Software and Technology*, UIST '11, pages 587–592, New York, NY, USA, 2011. ACM.
- [83] M. Hachet, F. Decle, S. Knodel, and P. Guitton. Navidget for easy 3d camera positioning from 2d inputs. In *Proceedings of the 2008 IEEE Symposium on 3D User Interfaces*, 3DUI '08, pages 83–89, Washington, DC, USA, 2008. IEEE Computer Society.
- [84] J. Y. Han. Low-cost multi-touch sensing through frustrated total internal reflection. In *Proceedings of the 18th Annual ACM Symposium on User Interface Software and Technology*, UIST '05, pages 115–118, New York, NY, USA, 2005. ACM.
- [85] M. Hancock, S. Carpendale, and A. Cockburn. Shallow-depth 3d interaction: design and evaluation of one-, two- and three-touch techniques. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, CHI '07, pages 1147–1156, New York, NY, USA, 2007. ACM.
- [86] M. Hancock, O. Hilliges, C. Collins, D. Baur, and S. Carpendale. Exploring tangible and direct touch interfaces for manipulating 2d and 3d information on a digital table. In *Proceedings of the ACM International Conference on Interactive Tabletops and Surfaces*, ITS '09, pages 77–84, New York, NY, USA, 2009. ACM.

- [87] S. G. Hart and L. E. Stavenland. Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. In P. A. Hancock and N. Meshkati, editors, *Human Mental Workload*, chapter 7, pages 139–183. Elsevier, 1988.
- [88] O. Hilliges, S. Izadi, A. D. Wilson, S. Hodges, A. Garcia-Mendoza, and A. Butz. Interactions in the air: Adding further depth to interactive tabletops. In *Proceedings of the 22Nd Annual ACM Symposium on User Interface Software and Technology*, UIST '09, pages 139–148, New York, NY, USA, 2009. ACM.
- [89] O. Hilliges, D. Kim, S. Izadi, M. Weiss, and A. Wilson. Holodesk: Direct 3d interactions with a situated see-through display. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '12, pages 2421–2430, New York, NY, USA, 2012. ACM.
- [90] K. Hinckley, R. Pausch, J. C. Goble, and N. F. Kassell. Passive real-world interface props for neurosurgical visualization. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '94, pages 452–458, New York, NY, USA, 1994. ACM.
- [91] K. Hinckley, R. Pausch, J. C. Goble, and N. F. Kassell. A survey of design issues in spatial input. In *Proceedings of the 7th annual ACM symposium on User interface software and technology*, UIST '94, pages 213–222, New York, NY, USA, 1994. ACM.
- [92] K. Hinckley, R. Pausch, D. Proffitt, and N. F. Kassell. Two-handed virtual manipulation. *ACM Transactions on Computer-Human Interaction*, 5(3):260–302, 1998.
- [93] K. Hinckley, J. Pierce, M. Sinclair, and E. Horvitz. Sensing techniques for mobile interaction. In *Proceedings of the 13th annual ACM symposium on User interface software and technology*, UIST '00, pages 91–100, New York, NY, USA, 2000. ACM.
- [94] S. Hodges, S. Izadi, A. Butler, A. Rrustemi, and B. Buxton. Thinsight: Versatile multi-touch sensing for thin form-factor displays. In *Proceedings of the 20th Annual ACM Symposium on User Interface Software and Technology*, UIST '07, pages 259–268, New York, NY, USA, 2007. ACM.
- [95] D. Holman. Gazetop: Interaction techniques for gaze-aware tabletops. In *CHI '07 Extended Abstracts on Human Factors in Computing Systems*, CHI EA '07, pages 1657–1660, New York, NY, USA, 2007. ACM.
- [96] S. Hoppe, F. Daiber, and M. Löchtefeld. Eype - using eye-traces for eye-typing. In *CHI 2013 Workshop on Grand Challenges in Text Entry*, 2013.
- [97] J. Huhtala, M. Karukka, M. Salmimaa, and J. Häkkinä. Evaluating depth illusion as method of adding emphasis in autostereoscopic mobile displays. In *Proceedings of the 13th International Conference on Human Computer Interaction with Mobile Devices and Services*, MobileHCI '11, pages 357–360, New York, NY, USA, 2011. ACM.

Bibliography

- [98] D. Q. Huynh. Metrics for 3d rotations: Comparison and analysis. *Journal of Mathematical Imaging and Vision*, 35(2):155–164, Oct. 2009.
- [99] ISO. ISO 9241-9 ergonomic requirements for office work with visual display terminals (VDTs) - part 9: Requirements for non-keyboard input devices, 2000.
- [100] H. Isono, M. Yasuda, and H. Sasazawa. Autostereoscopic 3-d display using lcd-generated parallax barrier. *Electronics and Communications in Japan (Part II: Electronics)*, 76(7):77–84, 1993.
- [101] F. E. Ives. Parallax stereogram and process of making same., Apr. 14 1903.
- [102] D. Jackson, T. Bartindale, and P. Olivier. Fiberboard: Compact multi-touch display using channeled light. In *Proceedings of the ACM International Conference on Interactive Tabletops and Surfaces*, ITS '09, pages 25–28, New York, NY, USA, 2009. ACM.
- [103] R. J. K. Jacob. What you look at is what you get: Eye movement-based interaction techniques. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '90, pages 11–18, New York, NY, USA, 1990. ACM.
- [104] L. A. Jones and S. J. Lederman. *Human Hand Function*. Oxford University Press, USA, 2006.
- [105] S. Jordà, M. Kaltenbrunner, G. Geiger, and M. Alonso. The reactable: a tangible tabletop musical instrument and collaborative workbench. In *SIGGRAPH '06: ACM SIGGRAPH 2006 Sketches*, page 91, New York, NY, USA, 2006. ACM.
- [106] R. Jota, M. A. Nacenta, J. A. Jorge, S. Carpendale, and S. Greenberg. A comparison of ray pointing techniques for very large displays. In *Proceedings of Graphics Interface 2010*, GI '10, pages 269–276, Toronto, Ont., Canada, Canada, 2010. Canadian Information Processing Society.
- [107] F. Kerber. Openindoormap - smartphone-based capture of uninstrumented indoor environments. Master's thesis, Saarland University, Saarbrücken, Germany, Department of Computer Science, 2012.
- [108] F. Kerber, P. Lessel, F. Daiber, and A. Krüger. Shift 'n' touch: Combining wii balance board and cubtile. In *Proceedings of the 7th Nordic Conference on Human-Computer Interaction: Making Sense Through Design*, NordiCHI '12, pages 789–790, New York, NY, USA, 2012. ACM.
- [109] F. Kerber, P. Lessel, M. Mauderer, F. Daiber, A. Oulasvirta, and A. Krüger. Is autostereoscopy useful for handheld ar? In *Proceedings of the 12th International Conference on Mobile and Ubiquitous Multimedia*, MUM '13, pages 4:1–4:4, New York, NY, USA, 2013. ACM.

- [110] J. M. Knapp and J. M. Loomis. Limited field of view of head-mounted displays is not the cause of distance underestimation in virtual environments. *Presence: Teleoperators and Virtual Environments*, 13(5):572–577, Oct. 2004.
- [111] F. Kosmalla. Boulder: Design and evaluation of a mobile augmented reality system for collaborative boulder training. Bachelor’s thesis, Saarland University, Saarbrücken, Germany, Department of Computer Science, 2013.
- [112] F. Kosmalla. Climbsense - automatic climbing route recognition using wrist-worn inertia measurement units. Master’s thesis, Saarland University, Saarbrücken, Germany, Department of Computer Science, 2014.
- [113] F. Kosmalla, F. Daiber, and A. Krüger. Climbsense: Automatic climbing route recognition using wrist-worn inertia measurement units. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*, CHI ’15, pages 2033–2042, New York, NY, USA, 2015. ACM.
- [114] S. Kratz and M. Rohs. Extending the virtual trackball metaphor to rear touch input. In *Proceedings of the 2010 IEEE Symposium on 3D User Interfaces*, 3DUI ’10, pages 111–114, Washington, DC, USA, 2010. IEEE Computer Society.
- [115] M. W. Krueger, T. Gionfriddo, and K. Hinrichsen. Videoplace—an artificial reality. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI ’85, pages 35–40, New York, NY, USA, 1985. ACM.
- [116] E. Kruijff, J. Swan, and S. Feiner. Perceptual issues in augmented reality revisited. In *Proceedings of the 9th IEEE International Symposium on Mixed and Augmented Reality*, ISMAR ’10, pages 3–12. IEEE, 2010.
- [117] J. Kuffner. Effective sampling and distance metrics for 3d rigid body path planning. In *Proceedings of the International Conference on Robotics and Automation*, volume 4 of *ICRA ’04*, pages 3993–3998. IEEE, 2004.
- [118] M. Kumar and T. Winograd. Gaze-enhanced scrolling techniques. In *Proceedings of the 20th Annual ACM Symposium on User Interface Software and Technology*, UIST ’07, pages 213–216, New York, NY, USA, 2007. ACM.
- [119] A. Kunert, A. Kulik, A. Huckauf, and B. Fröhlich. A comparison of tracking- and controller-based input for complex bimanual interaction in virtual environments. In *Proceedings of the 13th Eurographics conference on Virtual Environments*, EGVE’07, pages 43–52, Aire-la-Ville, Switzerland, Switzerland, 2007. Eurographics Association.
- [120] G. Kurtenbach and W. Buxton. User learning and performance with marking menus. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI ’94, pages 258–264, New York, NY, USA, 1994. ACM.

Bibliography

- [121] H. Lahamy and D. Litchi. Real-time hand gesture recognition using range cameras. In *Proceedings of the 2010 Canadian Geomatics Conference and Symposium of Commission I*, 2010.
- [122] M. E. Latoschik, M. Fröhlich, B. Jung, and I. Wachsmuth. Utilize speech and gestures to realize natural interaction in a virtual environment. In *Proceedings of the 24th annual Conference of the IEEE Industrial Electronics Society*, volume 4 of *IECON '98*, pages 2028–2033, 1998.
- [123] J. J. LaViola, Jr. A discussion of cybersickness in virtual environments. *ACM SIGCHI Bulletin*, 32(1):47–56, Jan. 2000.
- [124] J. J. LaViola Jr. Bringing VR and spatial 3d interaction to the masses through video games. *IEEE Computer Graphics and Applications*, 28(5):10–15, 2008.
- [125] B. Leibe, T. Starner, W. Ribarsky, Z. Wartell, D. Krum, J. Weeks, B. Singletary, and L. Hedges. Toward spontaneous interaction with the perceptive workbench. *IEEE Computer Graphics and Applications*, 20(6):54–65, 2000.
- [126] L. Li. Interaction with stereoscopic data displayed on mobile devices. Master’s thesis, Saarland University, Saarbrücken, Germany, Department of Computer Science, 2011.
- [127] H.-N. Liang, C. Williams, M. Semegen, W. Stuerzlinger, and P. Irani. An Investigation of Suitable Interactions for 3D Manipulation of Distant Objects through a Mobile Device. *International Journal of Innovative Computing, Information and Control (IJICIC)*, 9(12):4737–4752, Dec. 2013.
- [128] G. Lippmann. Épreuves réversibles donnant la sensation du relief. *Journal de Physique Théorique et Appliquée*, 7(1):821–825, 1908.
- [129] X. Liu and K. Fujimura. Hand gesture recognition using depth data. In *Proceedings of the Sixth IEEE international conference on Automatic face and gesture recognition*, FGR’ 04, pages 529–534. IEEE Computer Society, 2004.
- [130] M. Livingston, Z. Ai, J. Swan, and H. Smallman. Indoor vs. outdoor depth perception for mobile augmented reality. In *Proceedings of the Virtual Reality Conference*, VR ’09, pages 55–62. IEEE, March 2009.
- [131] M. Löchtefeld, S. Gehring, J. Schöning, F. Daiber, and A. Krüger. Tracking pointing gestures to support sales conversations. In *Adjunct Proceedings of the 28th International Conference on Human Factors in Computing Systems. Workshop on Performative Interaction in Public Spaces*. ACM, 2011.
- [132] P. Lubos, C. Garber, A. Hoffert, I. Reis, and F. Steinicke. The interactive spatial surface - blended interaction on a stereoscopic multi-touch surface. In A. Butz, M. Koch, and J. Schlichter, editors, *Mensch & Computer 2014 - Workshopband*, pages 343–346, Berlin, 2014. De Gruyter Oldenbourg.

- [133] M. Lucente. Interactive three-dimensional holographic displays: Seeing the future in depth. *ACM SIGGRAPH Computer Graphics*, 31(2):63–67, May 1997.
- [134] C. L. MacKenzie, R. G. Marteniuk, C. Lugas, D. Liske, and B. Eickmeier. Three-dimensional movement trajectories in fitts’ task: Implications for control’. *The Quarterly Journal of Experimental Psychology*, Section A, 39: 4:629–647, 1987.
- [135] J. D. Mackinlay, S. K. Card, and G. G. Robertson. Rapid controlled movement through a virtual 3d workspace. In *Proceedings of the 17th Annual Conference on Computer Graphics and Interactive Techniques*, SIGGRAPH ’90, pages 171–176, New York, NY, USA, 1990. ACM.
- [136] D. P. Mapes and J. M. Moshell. A two handed interface for object manipulation in virtual environments. *Presence: Teleoperators and Virtual Environments*, 4(4):403–416, 1995.
- [137] A. Martinet, G. Casiez, and L. Grisoni. 3d positioning techniques for multi-touch displays. In *Proceedings of the 16th ACM Symposium on Virtual Reality Software and Technology*, VRST ’09, pages 227–228, New York, NY, USA, 2009. ACM.
- [138] A. Martinet, G. Casiez, and L. Grisoni. Integrality and separability of multitouch interaction techniques in 3d manipulation tasks. *IEEE Transactions on Visualization and Computer Graphics*, 18(3):369–380, 2012.
- [139] M. R. Masliah and P. Milgram. Measuring the allocation of control in a 6 degree-of-freedom docking experiment. In *Proceedings of the SIGCHI conference on Human Factors in Computing Systems*, CHI ’00, pages 25–32, New York, NY, USA, 2000. ACM.
- [140] M. Mauderer. Combining touch and gaze for distant selection. Master’s thesis, Saarland University, Saarbrücken, Germany, Department of Computer Science, 2012.
- [141] M. Mauderer, F. Daiber, and A. Krüger. Combining Touch and Gaze for Distant Selection in a Tabletop Setting. In *CHI 2013 Workshop on Gaze Interaction in the Post-WIMP World*, 2013.
- [142] J. Maycock, B. Blaesing, T. Bockemühl, H. J. Ritter, and T. Schack. Motor synergies and object representations in virtual and real grasping. In *Proceedings of the 1st International Conference on Applied Bionics and Biomechanics*, ICABB ’10. IEEE Computer Society, 2010.
- [143] J. Maycock, K. Essig, R. Haschke, T. Schack, and H. J. Ritter. Towards an understanding of grasping using a multi-sensing approach. In *ICRA ’11: IEEE International Conference on Robotics and Automation - Workshop on Autonomous Grasping*. IEEE Computer Society, 2011.

Bibliography

- [144] R. Messing and F. H. Durgin. Distance perception and the visual horizon in head-mounted displays. *ACM Transactions on Applied Perception*, 2(3):234–250, July 2005.
- [145] M. Mikkola, A. Boev, and A. Gotchev. Relative importance of depth cues on portable autostereoscopic display. In *Proceedings of the 3rd Workshop on Mobile Video Delivery*, MoViD '10, pages 63–68, New York, NY, USA, 2010. ACM.
- [146] P. Milgram and F. Kishino. A taxonomy of mixed reality visual displays. *IEICE Transactions on Information and Systems*, E77-D(12):1321–1329, Dec. 1994.
- [147] P. Milgram, H. Takemura, A. Utsumi, and F. Kishino. Augmented reality: A class of displays on the Reality-Virtuality continuum. In *Proceedings of the SPIE Conference on Telemanipulator and Telepresence Technologies*, volume 2351, pages 282–292, Boston, Massachusetts, USA, Nov. 1995.
- [148] M. R. Mine. Virtual environment interaction techniques. Technical report, UNC Chapel Hill, 1995.
- [149] M. R. Mine, F. P. Brooks, Jr., and C. H. Sequin. Moving objects in space: exploiting proprioception in virtual-environment interaction. In *Proceedings of the 24th annual conference on Computer graphics and interactive techniques*, SIGGRAPH '97, pages 19–26, New York, NY, USA, 1997. ACM Press/Addison-Wesley Publishing Co.
- [150] B. A. Myers. A taxonomy of window manager user interfaces. *IEEE Computer Graphics and Applications*, 8(5):65–84, 1988.
- [151] J. R. Napier. The prehensile movements of the human hand. *The Journal of Bone and Joint Surgery*, 38-B(4):902–913, 1956.
- [152] D. A. Norman. *The Design of Everyday Things*. Basic Books, 2002.
- [153] R. J. Pagulayan, K. Keeker, T. Fuller, D. Wixon, and R. L. Romero. User-centered design in games. In A. Sears and J. A. Jacko, editors, *The Human-Computer Interaction Handbook: Fundamentals, Evolving Technologies and Emerging Applications*, pages 741–759. CRC Press, Boca Raton, FL, 2nd edition, 2008.
- [154] T. Pakkanen and R. Raisamo. Appropriateness of foot interaction for non-accurate spatial tasks. In *CHI '04 Extended Abstracts on Human Factors in Computing Systems*, CHI EA '04, pages 1123–1126, New York, NY, USA, 2004. ACM.
- [155] S. Pastoor and M. Wöpking. 3-d displays: A review of current technologies. *Displays*, 17(2):100–110, 1997.
- [156] V. I. Pavlovic, R. Sharma, and T. S. Huang. Visual interpretation of hand gestures for human-computer interaction: A review. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 19:677–695, July 1997.

- [157] G. Pearson and M. Weiser. Of moles and men: the design of foot controls for workstations. *ACM SIGCHI Bulletin*, 17(4):333–339, 1986.
- [158] G. Pearson and M. Weiser. Exploratory evaluation of a planar foot-operated cursor-positioning device. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '88, pages 13–18, New York, NY, USA, 1988. ACM.
- [159] J. S. Pierce, A. S. Forsberg, M. J. Conway, S. Hong, R. C. Zeleznik, and M. R. Mine. Image plane interaction techniques in 3d immersive environments. In *Proceedings of the 1997 symposium on Interactive 3D graphics*, I3D '97, New York, NY, USA, 1997. ACM.
- [160] I. Poupyrev, M. Billinghamurst, S. Weghorst, and T. Ichikawa. The go-go interaction technique: non-linear mapping for direct manipulation in VR. In *Proceedings of the 9th annual ACM symposium on User interface software and technology*, UIST '96, pages 79–80, New York, NY, USA, 1996. ACM.
- [161] J. L. Reisman, P. L. Davidson, and J. Y. Han. A screen-space formulation for 2d and 3d direct manipulation. In *Proceedings of the 22Nd Annual ACM Symposium on User Interface Software and Technology*, UIST '09, pages 69–78, New York, NY, USA, 2009. ACM.
- [162] J. Rekimoto. Tilting operations for small screen interfaces. In *Proceedings of the 9th annual ACM symposium on User interface software and technology*, UIST '96, pages 167–168, New York, NY, USA, 1996. ACM.
- [163] J. Rekimoto. Matrix: A realtime object identification and registration method for augmented reality. In *Proceedings of the Third Asian Pacific Computer and Human Interaction*, APCHI '98, pages 63–, Washington, DC, USA, 1998. IEEE Computer Society.
- [164] J. Rekimoto. Smartskin: an infrastructure for freehand manipulation on interactive surfaces. In *CHI '02: Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 113–120, New York, NY, USA, 2002. ACM.
- [165] J. Rekimoto and N. Matsushita. Perceptual surfaces: Towards a human and object sensitive interactive displays. In *Workshop on Perceptual User Interfaces*, PUI '97, 1997.
- [166] M. Santello, M. Flanders, and J. F. Soechting. Postural hand synergies for tool use. *Journal of Neuroscience*, 18(23):10105–10115, 1998.
- [167] M. Santello, M. Flanders, and J. F. Soechting. Patterns of hand motion during grasping and the influence of sensory guidance. *Journal of Neuroscience*, 22(4):1426–1435, 2002.

Bibliography

- [168] M. Santello and J. F. Soechting. Matching object size by controlling finger span and hand shape. *Somatosensory and Motor Research*, 14(3):203–212, 1997.
- [169] S. Scerbo and D. Bowman. Design issues when using commodity gaming devices for virtual object manipulation. In *Proceedings of the International Conference on the Foundations of Digital Games*, FDG '12, pages 294–295, New York, NY, USA, 2012. ACM.
- [170] J. Schöning, F. Daiber, and A. Krüger. Advanced navigation techniques for spatial information using whole body motion. 2009.
- [171] J. Schöning, F. Daiber, M. Rohs, and A. Krüger. Using hands and feet to navigate and manipulate spatial data. In *CHI '09: CHI '09 extended abstracts on Human factors in computing systems*, New York, NY, USA, 2009. ACM.
- [172] J. Schöning, B. Hecht, M. Raubal, A. Krüger, M. Marsh, and M. Rohs. Improving interaction with virtual globes through spatial thinking: Helping users ask "why?". In *Proceedings of the 13th International Conference on Intelligent User Interfaces*, IUI '08, pages 129–138, New York, NY, USA, 2008. ACM.
- [173] J. Schöning, J. Hook, N. Motamedi, P. Olivier, F. Echtler, P. Brandl, L. Muller, F. Daiber, O. Hilliges, M. Löchtefeld, T. Roth, D. Schmidt, and U. von Zadow. Building interactive multi-touch surfaces. *Journal of Graphics Tools*, 10:1–23, 2009.
- [174] J. Schöning, F. Steinicke, D. Valkov, A. Krüger, and K. H. Hinrichs. Bimanual interaction with interscopic multi-touch surfaces. In *Human-Computer Interaction INTERACT 2009. IFIP Conference on Human-Computer Interaction (INTERACT-2009), August 24-28, Uppsala, Sweden*, pages 40–53. Springer, 2009.
- [175] C. Shaw and M. Green. Two-handed polygonal surface design. In *Proceedings of the 7th Annual ACM Symposium on User Interface Software and Technology*, UIST '94, pages 205–212, New York, NY, USA, 1994. ACM.
- [176] L. E. Sibert and R. J. K. Jacob. Evaluation of eye gaze interaction. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '00, pages 281–288, New York, NY, USA, 2000. ACM.
- [177] J. D. Smith and T. C. N. Graham. Use of eye movements for video game control. In *Proceedings of the 2006 ACM SIGCHI International Conference on Advances in Computer Entertainment Technology*, ACE '06, New York, NY, USA, 2006. ACM.
- [178] G. Sörös, F. Daiber, and T. Weller. Cyclo: A personal bike coach through the glass. In *SIGGRAPH Asia 2013 Symposium on Mobile Graphics and Interactive Applications*, SA '13, pages 99:1–99:4, New York, NY, USA, 2013. ACM.

- [179] M. Speicher. Exploring 3d interaction techniques for stereoscopic content using consumer tracking devices. Master's thesis, Saarland University, Saarbrücken, Germany, Department of Computer Science, 2014.
- [180] O. Stefani and J. Rauschenbach. 3d input devices and interaction concepts for optical tracking in immersive environments. In *Proceedings of the Workshop on Virtual Environments 2003*, EGVE '03, pages 317–318, New York, NY, USA, 2003. ACM.
- [181] F. Steinicke, H. Benko, F. Daiber, D. Keefe, and J.-B. de la Rivière. Touching the 3rd Dimension (T3D). In *CHI '11 Extended Abstracts on Human Factors in Computing Systems*, CHI EA '11, pages 161–164, New York, NY, USA, 2011. ACM.
- [182] F. Steinicke, K. H. Hinrichs, J. Schöning, and A. Krüger. Multi-touching 3d data: Towards direct interaction in stereoscopic display environments coupled with mobile devices. In *Advanced Visual Interfaces (AVI) Workshop on Designing Multi-Touch Interaction Techniques for Coupled Public and Private Displays*, pages 46–49, 2008.
- [183] S. Stellmach and R. Dachsel. Look & touch: Gaze-supported target acquisition. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '12, pages 2981–2990, New York, NY, USA, 2012. ACM.
- [184] R. Stoakley, M. J. Conway, and R. Pausch. Virtual reality on a wim: interactive worlds in miniature. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '95, pages 265–272, New York, NY, USA, 1995. ACM Press/Addison-Wesley Publishing Co.
- [185] S. Strothoff, D. Valkov, and K. Hinrichs. Triangle Cursor: Interactions With Objects Above the Tabetop. In *ITS '11: Proceedings of the ACM International Conference on Interactive Tabletops and Surfaces*, pages 111–119. ACM, 2011.
- [186] I. E. Sutherland. A head-mounted three dimensional display. In *Proceedings of the December 9-11, 1968, Fall Joint Computer Conference, Part I*, AFIPS '68 (Fall, part I), pages 757–764, New York, NY, USA, 1968. ACM.
- [187] J. E. Swan, 2nd, A. Jones, E. Kolstad, M. A. Livingston, and H. S. Smallman. Egocentric depth judgments in optical, see-through augmented reality. *IEEE Transactions on Visualization and Computer Graphics*, 13(3):429–442, 2007.
- [188] V. Tanriverdi and R. J. K. Jacob. Interacting with eye movements in virtual environments. In *CHI '00: Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 265–272, New York, NY, USA, 2000. ACM.
- [189] M. M. Taylor and C. D. Creelman. PEST: Efficient estimates on probability functions. *Journal of the Acoustical Society of America*, 41:782–787, 1967.

Bibliography

- [190] R. J. Teather and W. Stuerzlinger. Pointing at 3d targets in a stereo head-tracked virtual environment. In *Proceedings of the 2011 IEEE Symposium on 3D User Interfaces*, 3DUI '11, pages 87–94, Washington, DC, USA, 2011. IEEE Computer Society.
- [191] P. H. Thakur, A. J. Bastian, and S. S. Hsiao. Multidigit movement synergies of the human hand in an unconstrained haptic exploration task. *Journal of Neuroscience*, 28(6):1271–1281, Feb 2008.
- [192] Y. Thong, M. Woolfson, J. Crowe, B. Hayes-Gill, and D. Jones. Numerical double integration of acceleration measurements in noise. *Measurement*, 36(1):73–92, July 2004.
- [193] J. Turner, J. Alexander, A. Bulling, D. Schmidt, and H. Gellersen. Eye pull, eye push: moving objects between large screens and personal devices with gaze and touch. In P. Kotz, G. Marsden, G. Lindgaard, J. Wesson, and M. Winckler, editors, *Human-Computer Interaction INTERACT 2013*, Lecture Notes in Computer Science, pages 170–186. Springer, 2013.
- [194] J. Turner, A. Bulling, J. Alexander, and H. Gellersen. Cross-device gaze-supported point-to-point content transfer. In *Proc. ETRA*, pages 19–26, 2014.
- [195] D. Valkov, F. Steinicke, G. Bruder, and K. Hinrichs. A multi-touch enabled human-transporter metaphor for virtual 3d traveling. In *Proceedings of the 2010 IEEE Symposium on 3D User Interfaces*, 3DUI '10, pages 79–82, Washington, DC, USA, 2010. IEEE Computer Society.
- [196] D. Valkov, F. Steinicke, G. Bruder, and K. Hinrichs. 2d touching of 3d stereoscopic objects. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '11, pages 1353–1362, New York, NY, USA, 2011. ACM.
- [197] D. Valkov, F. Steinicke, G. Bruder, K. H. Hinrichs, J. Schöning, F. Daiber, and A. Krüger. Touching floating objects in projection-based virtual reality environments. In *Joint Virtual Reality Conference*. Eurographics, 2010.
- [198] U. von Zadow, F. Daiber, J. Schöning, and A. Krüger. Globaldata: Multi-user interaction with geographic information systems on interactive surfaces. In *ACM International Conference on Interactive Tabletops and Surfaces*, ITS '10, pages 318–318, New York, NY, USA, 2010. ACM.
- [199] D. Wagner and D. Schmalstieg. First steps towards handheld augmented reality. In *Proceedings of the 7th IEEE International Symposium on Wearable Computers*, ISWC '03, pages 127–135, Washington, DC, USA, 2003. IEEE Computer Society.
- [200] C. Ware. Using hand position for virtual object placement. *The Visual Computer*, 6(5):245–253, 1990.

- [201] C. Ware and D. R. Jessome. Using the bat: A six-dimensional mouse for object placement. *IEEE Computer Graphics and Applications*, 8(6):65–70, Nov. 1988.
- [202] C. Ware and S. Osborne. Exploration and virtual camera control in virtual three dimensional environments. In *Proceedings of the 1990 symposium on Interactive 3D graphics*, I3D '90, pages 175–183, New York, NY, USA, 1990. ACM.
- [203] R. Wasinger, C. Stahl, A. Krüger, R. Wasinger, C. Stahl, and A. Krger. M3i in a pedestrian navigation & exploration system. In L. Chittaro, editor, *Human-Computer Interaction with Mobile Devices and Services*, volume 2795 of *Lecture Notes in Computer Science*, pages 481–485. Springer Berlin Heidelberg, 2003.
- [204] M. Weiser. The computer for the 21st century. In R. M. Baecker, J. Grudin, W. A. S. Buxton, and S. Greenberg, editors, *Human-computer Interaction*, pages 933–940. Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, 1995.
- [205] P. Wellner. Interacting with paper on the digitaldesk. *Communications of the ACM*, 36(7):87–96, July 1993.
- [206] C. Wheatstone. Contributions to the physiology of vision. part the first. on some remarkable, and hitherto unobserved, phenomena of binocular vision. *Philosophical transactions of the Royal Society of London*, pages 371–394, 1838.
- [207] D. Wigdor and D. Wixon. *Brave NUI World: Designing Natural User Interfaces for Touch and Gesture*. Morgan Kaufmann, 1 edition, Apr. 2011.
- [208] A. D. Wilson. Simulating grasping behavior on an imaging interactive surface. In *Proceedings of the ACM International Conference on Interactive Tabletops and Surfaces*, ITS '09, pages 125–132, New York, NY, USA, 2009. ACM.
- [209] A. D. Wilson, S. Izadi, O. Hilliges, A. Garcia-Mendoza, and D. Kirk. Bringing physics to the surface. In *Proceedings of the 21st Annual ACM Symposium on User Interface Software and Technology*, UIST '08, pages 67–76, New York, NY, USA, 2008. ACM.
- [210] C. Wingrave, B. Williamson, P. D. Varcholik, J. Rose, A. Miller, E. Charbonneau, J. Bott, and J. LaViola. The wiimote and beyond: Spatially convenient devices for 3d user interfaces. *IEEE Computer Graphics and Applications*, 30(2):71–85, 2010.
- [211] J. O. Wobbrock, M. R. Morris, and A. D. Wilson. User-defined gestures for surface computing. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '09, pages 1083–1092, New York, NY, USA, 2009. ACM.
- [212] M. Wu and R. Balakrishnan. Multi-finger and whole hand gestural interaction techniques for multi-user tabletop displays. pages 193–202, 2003.

Bibliography

- [213] M. Wu, C. Shen, K. Ryall, C. Forlines, and R. Balakrishnan. Gesture registration, relaxation, and reuse for multi-point direct-touch surfaces. In *Proceedings of the First IEEE International Workshop on Horizontal Interactive Human-Computer Systems, TABLETOP '06*, pages 185–192, Washington, DC, USA, 2006. IEEE Computer Society.
- [214] R. C. Zeleznik, J. J. LaViola Jr, D. A. Feliz, and D. F. Keefe. Pop through button devices for ve navigation and interaction. In *Proceedings of the IEEE Virtual Reality Conference 2002*, VR '02, pages 127–, Washington, DC, USA, 2002. IEEE Computer Society.
- [215] S. Zhai, W. Buxton, and P. Milgram. The “silk cursor” investigating transparency for 3d target acquisition. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '94, pages 459–464, New York, NY, USA, 1994. ACM.
- [216] S. Zhai and P. Milgram. Input techniques for hci in 3d environments. In *Conference Companion on Human Factors in Computing Systems*, CHI '94, pages 85–86, New York, NY, USA, 1994. ACM.
- [217] S. Zhai and P. Milgram. Quantifying coordination in multiple dof movement and its application to evaluating 6 dof input devices. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '98, pages 320–327, New York, NY, USA, 1998. ACM Press/Addison-Wesley Publishing Co.
- [218] D. Zilch, G. Bruder, and F. Steinicke. Comparison of 2d and 3d gui widgets for stereoscopic multitouch setups. *Journal of Virtual Reality and Broadcasting (JVRB)*, 2014.