Jens Grubert

Mobile Augmented Reality for Information Surfaces

DOCTORAL THESIS

to achieve the university degree of
Doktor der technischen Wissenschaften

submitted to

Graz University of Technology

Supervisor

Prof. Dr. techn. Dieter Schmalstieg
Institute for Computer Graphics and Vision, Graz University of Technology

Prof. Dr. rer. nat. Matthias Kranz
Institute for Embedded Systems, University of Passau

Graz, Austria, May 2015.

TU Graz
Graz University of Technology
Senat

Deutsche Fassung:
Beschluss der Curricula-Kommission für Bachelor-, Master- und Diplomstudien vom 10.11.2008
Genehmigung des Senates am 1.12.2008

EIDESSTATTLICHE ERKLÄRUNG

Ich erkläre an Eides statt, dass ich die vorliegende Arbeit selbstständig verfasst, andere als die angegebenen Quellen/Hilfsmittel nicht benutzt, und die den benutzten Quellen wörtlich und inhaltlich entnommenen Stellen als solche kenntlich gemacht habe.

Graz, am …………………………… ………………………………………………..
(Unterschrift)

Englische Fassung:

STATUTORY DECLARATION

I declare that I have authored this thesis independently, that I have not used other than the declared sources / resources, and that I have explicitly marked all material which has been quoted either literally or by content from the used sources.

…………………………… ………………………………………………..
date (signature)

To Leopold and Carina

Everything is best for something and worst for something else.

Bill Buxton

Abstract

The increasing amount of publicly accessible situated information and the increasing computational power of personal devices such as smartphones and tablets have created an ideal basis for the uptake of Augmented Reality among mobile users. Specifically, information surfaces provide a large potential for augmentations, as they already provide an utilitarian value to mobile users. Also, information surfaces are relatively easy to augment, due to the fact, that prior, to their production, they already exist as digital assets. Still, until today Augmented Reality has not become a mainstream user interface for interacting with situated information in mobile contexts. The goal of this thesis is to investigate under which circumstances Augmented Reality has the potential to increase the user experience when mobile users interact with information surfaces. This is done through a series of studies and prototypes exploring the applicability of Augmented Reality for a range of information surfaces in mobile contexts.We specifically consider large printed posters, security documents large public electronic displays and small personal displays. Based on the findings of these studies, we investigate strategies how to better utilize the potentials of Augmented Reality for these media.

We frame our work with surveys on the role of context for Augmented Reality and on insights about the usage of first generation Augmented Reality Browsers.

Specifically, this thesis

- extends the current understanding of context factors for mobile Augmented Reality.
- delivers new insights on the adoption and appropriation of Augmented Reality applications in mobile contexts.
- investigates the utility of Augmented Reality for interacting with security relevant information surfaces

- presents the potential of combining Augmented Reality with alternative user interfaces for interacting with situated information on a single handheld device.
- proposes means to facilitate the deployment of Augmented Reality content at public displays.
- demonstrates the potential of Augmented Reality interaction across multiple personal displays.

This thesis is intended to serve researchers and practitioners as a practical guide and as an inspiration how to incorporate Augmented Reality into current and next generation mobile applications for interacting with print and electronic information surfaces in mobile contexts.

Kurzfassung

Die steigende Zahl öffentlich zugänglicher verorteter Informationsquellen und die steigende Rechenleistung persönlicher mobiler Anzeigegeräte wie Smartphones oder Tablets bilden ideale Grundlagen für die Nutzung von Augmented Reality (Erweiterte Realität) durch mobile NutzerInnen. Insbesondere zeigen Informationsächen ein großes Potential für Augmentierungen, da diese schon einen intrinsischen utilitaristischen Wert für mobile NutzerInnen besitzen und mit relativ geringem technischem Aufwand zu augmentieren sind. Dennoch ist Augmented Reality bis heute keine weit verbreitete Benutzungsschnittstelle zur Interaktion mit verorteten Informationen in mobilen Kontexten. Das Ziel dieser Dissertation besteht darin, zu untersuchen unter welchen Umständen Augmented Reality das Potential aufweist die User Experience (das Nutzungserlebnis) bei der Interaktion mit Informationsflächen zu steigern. Dies wird durch eine Reihe von Studien und Prototypen getan, mit deren Hilfe die Anwendbarkeit von Augmented Reality für verschiedenen Typen von Informationsflächen untersucht wird. Im Besonderen werden Poster, Sicherheitsdokumente, große öffentliche und kleine private elektronische Anzeigegeräte untersucht. Auf der Grundlage der durchgeführten Studien, wird exploriert, wie die Potentiale von Augmented Reality für diese Informationsflächen besser nutzbar gemacht werden können. Insbesondere zielt diese Dissertation darauf ab:

- das aktuelle Verständnis von Kontextfaktoren für Augmented Reality zu erweitern.
- neue Einsichten in die Annahme von mobilen Augmented Reality Anwendungen zu generieren.
- den Nutzen von Augmented Reality zur Interaktion mit Sicherheitsdokumenten zu untersuchen.

- das Potential für die Interaktion mit Informationsflächen zu zeigen, welches die Kombination von Augmented Reality mit alternativen Benutzungsschnittstellen auf persönlichen Anzeigegeräten wie Smartphones besitzt.
- Mittel zur Bereitstellung von Augmented Reality Inhalten für öffentliche Anzeigegeräte vorzuschlagen.
- das Potential aufzuzeigen, welches Augmented Reality für die Interaktion mit mehreren tragbaren Anzeigegeräten besitzt.

Diese Dissertation zielt darauf ab Wissenschaftlern und Praktikern als Hilfestellung und Inspiration dafür zu dienen, wie Augmented Reality in mobile Anwendungen zur Interaktion mit Informationsflächen eingebettet werden kann.

Acknowledgments

This thesis would not have been possible without the support of several people. First, I thank my supervisor Prof. Dieter Schmalstieg. He gave me the freedom to conduct the research I wanted to pursue and at the same time offered me advice in framing my research. He also facilitated the means at the Institute for Computer Graphics and Vision (ICG) in which high quality research became possible in the first place. I thank Gerhard Reitmayr, who supervised me for the first two years of this thesis. He supported me in probing several research directions and convinced me that intrinsic motivation is the main driver for great research. I also thank Prof. Matthias Kranz, for accepting my request to be my second supervisor, showing interest in my work and delivering timely and constructive feedback for my dissertation.

Also, I want to thank my collaborators who shared their expertise and passion when I was approaching them with new research ideas or who invited me to be part of their exciting work. Among them are Hartmut Seichter, Raphael Grasset, Ann Morrison, Aaron Quigley, Andreas Hartl, Alessandro Mulloni and Lyndon Nixon.

Furthermore, I want to thank the current and past members of the ICG. Specifically, I thank Tobias Langlotz through whom the initial contact to the ICG was established and who offered me his advice throughout the past years. I also say thank you to the administrative staff, specifically, Christina Fuchs, who made it so much easier to concentrate on research.

Further, thanks to my students who trusted me and my research ideas and supported me with their motivated work. Specifically, I thank Christof Oberhofer and Matthias Heinisch without whom some publications would not have been possible.

My gratitude goes to my family who supported me in my endeavors for so many years. My father, mother, grandfather and grandmother gave me the opportunity to pursue excellent education, even when it meant of not seeing each other frequently.

I dedicate this thesis to my wife Carina and our son Leopold. Carina showed incredible patience and support. She supported me in my desire to conduct research and brought me back to the ground when needed. Finally, our son Leopold, who was born at the time of writing of this thesis, taught me the importance of focusing my mind sharply when needed as well as the joy of caring.

Contents

1 Introduction 1

1.1 Contributions . . . 3

1.1.1 Survey on context-aware AR . . . 5

1.1.2 User survey on first generation AR browsers . . . 5

1.1.3 Social context in AR gaming at posters in public space . . . 6

1.1.4 The utility of AR for information browsing at printed maps . . . 6

1.1.5 Hybrid AR interfaces for poster interaction in mobile contexts . . . 7

1.1.6 The utility of AR for verifying security documents . . . 8

1.1.7 Facilitating AR interaction with public electronic displays . . . 8

1.1.8 Mobile interaction with multiple displays on and around the body . 9

1.2 Results . . . 9

1.3 Publications and collaboration statement . . . 11

2 Related Work 15

2.1 Towards mobile AR . . . 15

2.2 Mobile interaction with information surfaces . . . 17

2.3 Hybrid user interfaces . . . 20

2.4 User studies of spatially-aware mobile interfaces . . . 21

2.5 Context-awareness . . . 23

2.6 Summary . . . 25

3 Towards Context-aware Mobile AR 27

3.1 Context-aware AR survey . . . 27

3.1.1 A Taxonomy for context-aware AR . . . 28

3.1.1.1 Context sources . . . 29

3.1.1.2 Context targets . . . 33

3.1.1.3 Controller 33
3.1.2 Survey on existing approaches for context-aware AR 34
3.1.2.1 Human factors 34
3.1.2.2 Environmental factors 35
3.1.2.3 System factors 38
3.1.2.4 Others 39
3.1.3 Discussion 39
3.1.3.1 Summary of existing approaches 39
3.1.3.2 Opportunities for future research 40
3.2 AR browser survey 41
3.2.1 Online survey 42
3.2.1.1 Method 42
3.2.1.2 Results 43
3.2.1.3 Discussion 59
3.2.2 Mobile distribution platform analysis 60
3.2.2.1 Method 60
3.2.2.2 Results 60
3.2.2.3 Discussion 63
3.3 Information access at event posters 64
3.3.1 Survey 64
3.3.2 Discussion 65
3.4 Summary 65

4 Interaction with Posters in Public Space 67
4.1 ML and SP interfaces for games in a public space 68
4.1.1 Game design and implementation 69
4.1.2 Study design 70
4.1.2.1 Participants 73
4.1.2.2 Hypotheses 73
4.1.2.3 Data collection 73
4.1.3 Findings 74
4.1.3.1 ML was used most of the time 74
4.1.3.2 Reasons for using ML 76
4.1.3.3 Reasons for using SP 76
4.1.3.4 Reactions from the public 77
4.1.3.5 Detachment from the environment 79
4.1.3.6 No Significant differences in performance between lab and public group 81
4.1.3.7 Using the interfaces outside of the study setting 81
4.1.4 Discussion 81
4.2 Repeated evaluation of ML and SP interfaces in public space 83

4.2.1 Findings 84
4.2.1.1 SP was used most of the time 84
4.2.1.2 Preference 85
4.2.1.3 Tracking errors 85
4.2.1.4 Interactions with passers-by 86
4.2.1.5 Reasons for using ML and SP 87
4.2.2 Discussion 87
4.3 The utility of ML interfaces on handheld devices for touristic map navigation 89
4.3.1 Semi-controlled field experiment 91
4.3.1.1 Study design and task 91
4.3.1.2 Apparatus and location 92
4.3.1.3 Data collection 92
4.3.1.4 Hypotheses 93
4.3.1.5 Procedure 93
4.3.1.6 Participants 94
4.3.1.7 Findings 94
4.3.1.8 Discussion 99
4.3.2 Lab study 100
4.3.2.1 Findings 101
4.3.2.2 Discussion 106
4.3.2.3 Limitations 107
4.3.3 Discussion 107
4.4 Exploring the design of hybrid interfaces for augmented posters in public spaces 108
4.4.1 Hybrid interfaces for augmented print media 109
4.4.1.1 Frame of reference 110
4.4.1.2 Navigation of information space 110
4.4.1.3 Transition between interaction spaces 112
4.4.2 Case studies 113
4.4.2.1 Hybrid interface: AR + zoomable view 113
4.4.2.2 Case study I: event poster 114
4.4.2.3 Case study II: game poster 117
4.4.3 Discussion and recommendations 117
4.5 Summary 117

5 Interaction with Security Documents 119
5.1 Introduction to document inspection 120
5.2 Technical background 120
5.2.1 Capturing view-dependent elements with mobiles 121
5.2.2 Retrieving reference views on a mobile phone 122
5.2.3 Automatic verification 123

5.3 Feasability of interactive user guidance for hologram verification 123
5.3.1 User interface concept 124
5.3.2 Implementation 125
5.3.2.1 SVBRDF capture 125
5.3.3 Evaluation 127
5.3.3.1 Study design and apparatus 127
5.3.3.2 Task and procedure 128
5.3.3.3 Participants 129
5.3.3.4 Data collection 129
5.3.3.5 Results 129
5.3.4 Discussion 133
5.4 Towards efficient AR user interfaces for hologram verification 134
5.4.1 Revisting user guidance for hologram verification 134
5.4.2 Alignment interface 136
5.4.3 Constrained navigation interface 137
5.4.4 Hybrid interface 139
5.4.5 Evaluation 140
5.4.5.1 Study design and tasks 140
5.4.5.2 Apparatus and data collection 141
5.4.5.3 Procedure 142
5.4.5.4 Participants 143
5.4.5.5 Hypotheses 143
5.4.5.6 Findings 144
5.4.6 Discussion 145
5.5 Summary 146

6 Interaction with Public and Personal Electronic Displays 149
6.1 Facilitating AR at public displays 150
6.1.1 User perspective ML approach 150
6.1.2 Implementation 151
6.1.2.1 Tracking screen content 152
6.1.2.2 Head tracking 152
6.1.2.3 Rendering 152
6.1.2.4 Sample application 153
6.1.3 Performance 153
6.1.4 Limitations and improvements 153
6.1.5 Conclusion 154
6.2 Multi-fidelity interaction with displays on and around the body 154
6.2.1 Interaction by dynamic alignment 155
6.2.1.1 Design factors 156
6.2.1.2 Alignment modes 157

6.2.1.3 Navigation 157
6.2.1.4 Focus representation and manipulation 158
6.2.1.5 Widgets and applications 159
6.2.2 Implementation 160
6.2.2.1 Software 160
6.2.2.2 Devices 161
6.2.2.3 Tracking 162
6.2.3 User study 162
6.2.3.1 Experimental design 162
6.2.3.2 Apparatus and data collection 163
6.2.3.3 Procedure 163
6.2.3.4 Participants 163
6.2.3.5 Hypotheses 163
6.2.4 Experiment 1: Locator task on map 164
6.2.4.1 Task completion time and errors 164
6.2.4.2 Subjective workload and user experience 165
6.2.5 Experiment 2: 1D target acquisition 165
6.2.5.1 Task completion time and errors 166
6.2.5.2 Subjective workload and user experience 167
6.2.5.3 Qualitative feedback 167
6.2.6 Discussion 168
6.3 Summary 170

7 Conclusion 171
7.1 Summary of the thesis results 171
7.2 Limitations 172
7.3 Directions for future work 173
7.4 Summary 175

A List of Acronyms 177

B Overview Context-Aware AR Systems 179

Bibliography 183

List of Figures

1.1 Overview of the thesis chapters. The chapters C2, C3, C7 provide the frame for the case studies presented in C4-C6. 4

3.1 Context sources (left) and targets (right) relevant for Augmented Reality (AR) interaction. The numbers in the circles indicate the amount of papers in the associated category, papers can be present in multiple categories. . . 32
3.2 Participant's background. 44
3.4 Frequency usage of Mobile Services. 45
3.5 Usage Time. 46
3.6 Usage Frequency and Duration. 47
3.7 Spineplots AR background w.r.t. usage frequency and active usage duration. 48
3.8 Usage scenarios. 49
3.9 Rating of performance of current ARB for application domains. 50
3.10 Rating of potential of ARB for application domains. 50
3.11 Type of consumed media. 51
3.12 Registration quality rating (blue) and issue frequency (orange). PA: Position Accuracy. PS: Position Stability. 51
3.13 UI and content quality and issue frequency. 52
3.14 Device related quality rating (blue) and issue frequency (orange). Bat: Battery. Net: Network. SS: Screen Size. SQ: Screen Quality. H: Device Handiness. W: Device Weight. 52
3.15 Movement patterns. S: standing. S+R: standing combined with rotation. MS+R: small (1-5 m) movements combined with rotation. ML+R: larger movements (> 5 m) combined with rotation. MML+R: multiple large movements (> 5 m) combined with rotation. 54
3.16 General Social Issues. 56

3.17 Times *AR* browsers were not used in several situations. 57

3.19 Difference of user ratings om both platforms based on Layar as example case (5 stars are very good, while 1 star is very poor). 61

3.20 Result of clustering the total 1135 comments of the Apple App Store by focusing on negative connotations. 62

4.1 A large target within selection distance (indicated by orange ring) in the Magic Lens (ML) view (left). User pinching to zoom in to a small target in the Static Peephole (SP) view (right) 70

4.2 A participant playing the game in front of the poster at the public transit place in Graz, Austria 71

4.3 Tracking errors indicated by black circle in the middle of the screen (left). Overview of one configuration of colored target boxes in the performance phase (right). 72

4.4 Participant playing the game in the laboratory. 72

4.5 Relative usage duration for the *ML* (blue) and *SP* (green) interface in the public and lab condition. 74

4.6 Absolute level completion times for the public and lab group. 75

4.7 Relative usage duration for the *ML* interface over individual levels in the public setting. 75

4.8 Participant using solely his arms to move back and forth (top row), bending knees to hit a target at the lower half of the poster (middle row), holding the phone above the head to reach targets at the top of the poster (bottom row). 77

4.9 Various ways to hold the phone in the *ML* condition: Switching from portrait to landscape mode (top row), holding the phone across the short or long edge (middle row), using gloves to cope with the cold (bottom row). . 78

4.10 Passersby not noticing the participants interacting with *ML* (left) and *SP* interfaces (right). 78

4.11 Passersby glimpsing (top row), watching from a distance (middle row) and approaching a participant (bottom row). 79

4.12 Ratings for selected questions concerning concentration on system and task and distraction by environment (5-point Likert scale, 1: totally disagree, 5: totally agree) 80

4.13 Number of participants who would use the interfaces at various locations (pt: public transportation. 82

4.14 Number of participants who would use the interfaces in front of various audiences. 82

4.15 The location of the repeated study was a public transportation stop in Vienna, Austria. 84

4.16 Relative usage durations for the *ML* (blue) and *SP* (green) interface for PUG, LAB and PUV . . . 85
4.17 Participant looking at two drunken men sitting on a nearby bench, who are chatting and watching the scene. . . . 86
4.18 Passer-by intruding the personal space of a participant, who is also watched by a woman sitting on a nearby bench. . . . 86
4.19 Schematic top down view of the space in the PUG condition with Hall's reaction bubbles indicating the intimate (0.5 m), personal (1.2m) and social space (3.6m) of the participants. . . . 88
4.20 Schematic top down view of the space in the PUG condition with Hall's reaction bubbles indicating the intimate (0.5 m), personal (1.2m) and social space (3.6m) of the participants. . . . 88
4.21 *ML* interface used in the semi-controlled field study (column 1), participant interacting with the interface during the semi-controlled field study (column 2), *ML* interface used in the lab study (column 3) and participant interacting with a large map used in the lab study (column 4). . . . 89
4.22 Schematic top down view of the study location with Hall's reaction bubbles indicating the intimate (0.5 m), personal (1.2 m) and social space (3.6 m) of the participants (left) and photograph of the location with one participant and three passers-by (right). . . . 93
4.23 Position in the x-y plane. *ML* camera centers: top row, left. *SP* camera center: top row, right. *ML* projected camera centers in poster plane: bottom row. Dots with unique colors represent individual participants. The LOWESS (locally weighted scatterplot smoothing) curve is shown in blue. Every 20^{th} camera sample is shown. . . . 96
4.24 Distance to poster in relation to the horizontal (left) and vertical (middle) positions and horizontal and vertical rotations (right). Dots with unique colors represent individual participants. Every 20^{th} camera sample is shown. 97
4.25 The small (left), medium (middle) and large workspace (right) used in the laboratory study. . . . 101
4.26 Camera position in the x-y plane of the small (row 1), medium (row 2) and large (row 3) workspace for *ML* (left) and *SP* (right), camera positions of *ML* projected on the workspace plane (middle). Dots with unique colors represent individual participants. Every 20^{th} position sample is shown. . . . 102
4.27 Camera distance in relation to the horizontal camera positions (columns 1-2) and vertical camera positions (columns 3-4) for the small (row 1), medium (row 2) and large workspace (row 3). Dots with unique colors represent individual participants. Every 20^{th} position sample is shown. . . . 104
4.28 Map visibility in the device screen relative to Task Completion Time (TCT) over all trials on the large map. Typical camera motion paths of a single participant overlaid in red (*ML*: left, *SP*: right). . . . 104

4.29 The representation of a physical print medium can be preserved by turning it into a digital surface. 111

4.30 To watch the augmented video users are forced to keep the physical magazine in view. 112

4.31 Transition from *AR* into zoomable view by pointing the phone away from a poster. 114

4.32 Posters with depicted digital content. left: event poster with 2D media items like widgets (1), image collection (2), trigger regions for showing / hiding content (3) and videos (not visible). right: game poster with 3D (1) and 2D (2) animations. 115

5.1 With the LED light source in a fixed configuration to the camera, there are only three Degrees of Freedom (DOF) in the input to the Spatially Varying Bidirectional Reflectance Distribution Function (SVBRDF) function. 122

5.2 First iteration of our interactive system for verification of view-dependent elements: It performs *SVBRDF* capture using the built-in LED on the mobile device (top-left). The user gets an overview of relevant views for verification, which are color-coded w.r.t. the decision of the user (right, note the number attached to each view). The system allows the user to accurately match given referene views and to compare the changes of holographic or similar security elements with the corresponding reference appearances (bottom). . 124

5.3 Geometry of the proposed alignment approach. Matching the current view with a given reference view takes place by aligning the viewing ray direction, position (base sphere on the device screen with the ray base circle, ray top circle with the ray bottom circle) and orientation (virtual horizon on top of ray with the virtual horizon on the device screen). 125

5.4 Exemplary alignment sequence: Not aligned (top left). Aligning direction using iron sights (top right). Adjusting distance (bottom left). Aligning rotation using the virtual horizon (bottom right). 126

5.5 Exemplary view used in the Digital Manual (DM). Overall image indicating the viewpoint (left). Zoomed image of the hologram patch (right). 127

5.6 Image showing table setup used during the study (left). Specimen banknote with window showing hologram to be checked by participants of the study (right). 128

5.7 Alignment errors for different views of the hologram captured in the user study. Translation (left). Rotation (right). Axis color-coded: x: red, y: green, z: blue 130

5.8 Matching registered patches: reference, warped image, registered image (left). Normalized Cross Correlation (NCC) scores with registered images for different views (right). 131

5.9 *TCTs* (s) for the *AR* and *DM* interfaces (left) and agreement to 'I think the hologram is real' (right). 131

5.10 Weighted NASA TLX dimensions for demands imposed on subject and for task interaction (MD: Mental Demand, PD: Physical Demand, TD: Temporal Demand, per: Performance, Eff: Effort, Fru: Frustration. 132

5.11 AttrakDiff scores for Pragmatic Quality (PQ) (*PQ*), Hedonic Identity (Hedonic Quality - Identity (HQ-I)), and Hedonic Stimulation (Hedonic Quality - Stimulation (HQ-S)) on a 5-item bipolar scale. 133

5.12 intrinsic motivation inventory scores for Interest/Enjoyment (IE) and Value/Usefulness (VU). 133

5.13 User interfaces for hologram verification: Constrained navigation (top-left), alignment (top-right) and hybrid user interfaces (bottom-left) are designed, implemented and evaluated within a user study. They allow to reliably capture image data suitable for automatic verification. Results are presented to the user in a summary (bottom-right). 135

5.14 Geometry of the revised alignment approach. Matching takes place by alignment of target rotation and pointing with the indicator at the element. Finally the viewing direction is refined using the direction rubber band at an acceptable viewing distance. 137

5.15 Exemplary alignment sequence: Not aligned (top left). Aligning target rotation (top right). Pointing at target (bottom left). Aligning viewing direction along hemisphere arc (bottom right). 138

5.16 Geometry of the proposed constrained navigation approach for sampling the hologram. The user is guided to point at the element and a cursor is controlled by the 2D orientation on an augmented pie, divided into slices and tracks. 138

5.17 We guide the user to point at the element using an animated rubber band (top-left). Focus adjustment showing the layout of the orientation map and green distance bounds (top-right). Constrained navigation UI with pie slices (bottom-left). Augmentation directly onto the document/element (bottom-right). 139

5.18 *AR* UIs with guidance for interesting subspaces. Either pie-slices (*AR*-CON, left) or circular regions (*AR*-HYB, right) are indicated for sampling by the user. 140

5.19 Samples used in our study. We evaluated all user interfaces with two original (no. 1, 4 - top row) and two fake (no. 2, 3 - bottom row) holograms, where each was placed on a different document template. Reference information recorded with the robot setup is used by the system for matching, while the other images are exemplary recordings during verification by the user. . . . 142

6.1 HeadLens requires only access to a remote screencast and is otherwise self-contained, making it suitable for multiple users 150

6.2 (left) Simultaneous 3D head tracking and Natural Feature Tracking (NFT) enable user perspective magic lenses for situated display content. (right) Device perspective rendering usually found in handheld Augmented Reality devices. 151

6.3 Overview of the HeadLens algorithm. (1) A content source such as a PC sends a video signal to a situated display. (2) The situated display shows the corresponding image. (3) The handheld device decodes a QR code to determine the screencast channel. (4) A screencast hardware or software multicasts the video signal to a wireless network. (5) The handheld device tracks the location of the situated display with the back-facing camera (6) and the location of the user's face with the front-facing camera. (7) Finally, the handheld device displays user-perspective *AR* content. 152

6.4 MultiFi widgets crossing device boundaries based on proxemics dimensions (left), e.g., middle: *ring menu* on a smartwatch with Head-Mounted Display (HMD) or right: *soft keyboard* with full-screen input area on a handheld device and *HMD*. 155

6.5 The extended screen space metaphor for showing a high resolution inlay of a map on smartwatch inside a low resolution representation on a *HMD*. . . 156

6.6 In body-aligned mode (left) devices are spatially registered in a shared information space relative to the user's body. In device-aligned mode (middle) the screen space of the touchscreen is extended. In side-by-side mode (right) devices have separated information spaces and do not require a spatial relationship. 158

6.7 Spatial pointing via a handheld triggers a low fidelity widget on the *HMD* to appear in high fidelity on the handheld. 159

6.8 Arm clipboard with extended screen space for low fidelity widgets (top). Spatial pointing enables switching to high fidelity on a handheld (bottom). 161

6.9 *TCT* (s) for the locator task. 164

6.10 *PQ* and *HQ-S* measures (normalized range -2..2) for the locator task (left) and the select task (right). 165

6.11 The selection task for SWRef. 166

6.12 *TCT* (s) for the select task. SWSide: side on which smartwatch was worn, SWOpSide: opposite side. 167

7.1 How to support ad-hoc around device interaction (left), continuous interaction with handhelds on and above large electronic information surfaces (middle) and cross-media interaction (right)? 174

B.1 Context targets relevant for *AR* interaction and the associated papers. . . . 180

B.2 Context sources relevant for *AR* interaction and the associated papers. . . . 181

List of Tables

3.1 Kendall's τ rank correlation between current usage rating and usage potentials. 50

3.2 Kendall's τ rank correlation between ratings of issue quality (low to high) and frequency of issues (never to very often). Interquartile range was 2 for all ratings and issue frequencies (IF Mdn: Issue frequency median). 53

3.3 Significant differences in feature quality ratings for frequent (f) vs. non-frequent (nf) users according to Mann-Whitney U test. Interquartile range was 2 for all ratings. 53

3.4 Differences in ratings in feature quality ratings for long-term (lt) vs. short-term (st) users according to Mann-Whitney U test. Interquartil range was 2 for all ratings. 54

3.5 Contingency table for standing combined with rotations (S+R) grouped by usage frequency. 54

3.6 Contingency table for multiple large movements ($> 5\ m$) combined with rotations (MML+R) grouped by usage frequency. 55

3.7 Contingency table for multiple large movements ($> 5\ m$) combined with rotations (MML+R) grouped by active usage duration. 55

3.8 Contingency table for multiple large movements ($> 5\ m$) combined with rotations (MML+R) grouped by *AR* background. 55

3.9 Contingency table for larger ($> 5\ m$) movements combined with rotations (ML+R) grouped by *AR* background. 55

3.10 Contingency table for small (1-5 m) movements combined with rotations (MS+R) grouped by *AR* background. 55

4.1 Questionnaire items that were rated significantly higher for the *ML* over the *SP* interface in the public group. 76

4.2 Questionnaire items that were rated significantly different between the public location (P) and lab (L) group. 80
4.3 *TCTs* (s) over 4 levels in performance phase. 81
4.4 Selection errors over 4 levels in performance phase. 81
4.5 TCTs (in seconds, mean and σ) for individual levels in *ML* condition and p-values from post-hoc pairwise comparison (using one-tailed t-tests, bold values indicate sign. differences). 95
4.6 TCTs (in seconds, mean and σ) for individual levels in *SP* condition and p-values from post-hoc pairwise comparison (using one-tailed t-tests, bold values indicate sign. differences) 95
4.7 Number of significant ($p < 0.05$) pairwise differences (from 136 possible) with medium (Cohen's d = [0.5, 0.8]) and large (Cohen's d >= 0.8) effect sizes for motion in x, y and z plane of the poster. 96
4.8 Text sizes (pt) and candidate button (hut) sizes at the moment of candidate selection. 97
4.9 Results of two-tailed Wilcoxon signed-rank tests did not indicate significant effects of interface on the depicted UX dimensions and mean and standard deviation of the ratings (scaled to -2..2). 98
4.10 Reactions of passers-by. 98
4.11 Objective measurements in the laboratory study. Reported are mean and standard deviation (σ). 102
4.12 Mean distances (in cm) between camera and the workspace and standard deviations (σ). 103
4.13 Test statistics and effect sizes (Cohen's d) of interface on the positions of the camera based on all position samples. For px and py the values based on camera positions projected on the workspace plane are listed in parentheses. 103
4.14 Number of significant ($p < 0.05$) pairwise differences (from 171 possible) with medium (Cohen's d = [0.5, 0.8]) and large (Cohen's d >= 0.8) effect sizes for motion in x, y and z plane of the small, medium and large workspaces. 105
4.15 TLX and *PQ* dimensions. Reported are mean and standard deviation (σ). . 105

5.1 Group differences for UX Qualities *PQ*, *HQ-I* and *HQ-S* between the *AR* and *DM* interface condition. 133

1 Introduction

Contents

1.1 Contributions 3

1.2 Results 9

1.3 Publications and collaboration statement 11

The idea to merge electronic and physical information has fascinated humans for over a century. An early description of enhancing physical entities with virtual information is the "Character Marker" which consists of "a pair of spectacles" which makes "electrical vibrations" visible on the forehead of of people "with a letter indicating his or her character" [16]. Situating this information into physical entities or directly into 3D space has the potential to facilitate the understanding of the physical world around us in complex contexts [70]. While this physical world is inherently 3D, *planar surfaces* are a dominant way to simplify the interaction with information in physical space. These planar surfaces are ubiquitous, often serve communicative purposes and range from small scale personal (e.g., smartwatches, badges, books) to large scale public surfaces (e.g., posters, electronic billboards).

A powerful interface metaphor for interacting with digital information situated in physical objects is *Augmented Reality* (Augmented Reality (AR)). The core idea of *AR* is to augment or overlay computer-generated information onto the physical world - i.e., making situated information accessible right at the physical object or space it relates to.

One of the first working *AR* systems was realized in the military domain in the 1960s by Sutherland [234]. In the early 1990s, *AR* was introduced in the industrial domain by Caudell and Mizell who coined the term Augmented Reality [44]. They used head-worn displays to guide workers in the manual manufacturing processes. In 1993, Fritzmaurice introduced the Chameleon Lens, a handheld device to augment physical information spaces such as posters [70]. Towards the end of the 1990s, several mobile *AR* systems for outdoor use emerged as wearable (i.e., backpack) variants of desktop systems, notably the Touring Machine [67], MARS [120] and Tinmith [184]. However, most of these early systems mainly focused on demonstrating the technological feasibility of *AR* and relied on relatively complex and expensive equipment. In addition, only few (e.g., [194]) systems reported user studies indicating utilitarian or hedonic benefits for users over existing interfaces.

In the mid-2000s, first consumer-oriented handheld systems (such as Personal Digital Assistants (PDAs)) emerged that were capable of running basic *AR* systems. Most of these systems offered physical buttons, joysticks or resistive touch screens with stylus as primary input modalities. With these devices available and new telecommunication standards such as UMTS being available, several researchers started to explore the use of mobile personal devices for interacting with information surfaces in mobile contexts, such as printed maps (e.g., [219]) or public displays (e.g., [12, 198]). First user studies on *AR* usage on maps were conducted (e.g., [202, 209]), and a first taxonomy for mobile interaction with situated displays emerged [11]. However, the computational capacities of the devices limited the performance of *AR* systems (e.g., tracking with 10 Hz and operational ranges of 6-21 cm [202]).

At the end of the 2000s, smartphones became an affordable platform for mobile *AR*. They had sufficient computing power both for 3D tracking and 3D computer graphics as well as multitouch-ready touch screens and a growing number of sensors. This contributed to opening up *AR* for consumer markets [217]. The release of freely available *AR* software development kits (SDKs) such as Qualcomm Vuforia[1] or *AR* browsers such as Metaio Junaio[2] enabled developers and content producers to create consumer-oriented *AR* experiences without the need for in-depth technical knowledge about the underlying technology. Hence, the number of mobile *AR* apps available in mobile application stores steadily increased over the last years, with a focus on gaming and marketing applications [140].

Hedonic user experience aspects (specifically the “wow” effect connected to arousal) might be one factor why *AR* is popular in marketing, as arousal is connected with increasing the attention of users [127]. Companies such as Layar or Blippar concentrate specifically on making situated information accessible via printed *information surfaces* such as posters and magazines for marketing purposes. Layar claims over 40 million app downloads as of July 2014[3], and Blippar claims up to 75 seconds attention span achieved with *AR* campaigns[4] (compared with an average of 30 seconds TV ads). Indeed, information surfaces lend themselves particularly well for augmentation with today’s commercially available *AR* solutions. Compared to complex physical 3D objects, planar information surfaces can be recognized and tracked well with commercial solutions, allowing both for a fast retrieval of associated situated media (such as videos or 3D models) and precise registration of this media on the surfaces.

Today, mobile *AR* apps are available to millions of users. Information surfaces can be (technically) easily augmented and *AR* could potentially increase utilitarian and hedonic aspects of the user experience. Still, there is lack of scientific evidence that *AR* does actually benefit consumers in interacting with those information surfaces. Even though there is a growing body of work on the evaluation of *AR*, most user studies so far focused on user task performance [60, 62] or low-level perceptual tasks [138] under laboratory

[1] https://www.qualcomm.com/products/vuforia, last retrieved 20.04.2015.

[2] http://www.metaio.com/junaio/, last retrieved 20.04.2015.

[3] https://www.layar.com/news/blog/tags/stats/, last retrieved 20.04.2015.

[4] https://blog.blippar.com/en/blog/176-understanding-roi-in-ar, last retrieved 20.04.2015.

conditions. There is clearly a lack investigating utilitarian and hedonic user experience factors in real-world contexts. It is primarily these aspects, hedonics and utility, which drive consumer attitudes and hence the potential adoption of *AR* [15].

1.1 Contributions

This thesis contributes to the fields of Mobile Human-Computer Interaction, Augmented Reality and Pervasive Computing by investigating how mobile *AR* user interfaces affect hedonic and utilitarian user experience aspects of interaction with information surfaces. Within this thesis we understand information surfaces as two-dimensional subspaces of the physical three-dimensional space that serve communicative purposes, i.e., they are intended to provide meaningful information to humans. While information surfaces can come in many shapes (such as curved monitors or deformable money bills), within this thesis we specifically concentrate on planar surfaces. Examples include printed posters, flat electronic displays such as public digital signage systems or personal displays such as smartphones and smartwatches. As information surfaces are designed for communicative purposes they can address both utilitarian and hedonic needs of users. They are also often artifacts for which digital information is readily available (in case of electronic displays) or even exists before they are made (for printed surfaces). We believe that *AR* has the potential for widespread adoption if it can provide further utilitarian or hedonic value to information surfaces.

To this end, this thesis provides insights on factors which influence *AR* interaction with information surfaces, studies context factors that are relevant for *AR* and evaluates concepts and prototypes of *AR* user interfaces targeted at increasing the utility of information surfaces. In the following, the main contributions of this thesis are summarized (see Figure 1.1).

- **A survey on context-aware *AR* systems** providing a comprehensive overview of how existing *AR* systems adapt to varying contexts, including a taxonomy of context sources and targets for *AR* and identification of opportunities for future research on context-aware *AR* systems (Chapter 3). [92]

- **A user survey investigating first generation *AR* browsers** identifying the main factors which drive the usage of consumer-oriented *AR* applications. The study shows that, currently, *AR* browsers are mostly used by early adopters for curiosity reasons, but that users see few benefits going beyond novelty effects [86, 140] (Chapter 3).

- **A series of semi-controlled field studies investigating the social context in *AR* gaming at posters in public space**. They highlight the importance social factors can have on the usage of *AR* in public space [87, 89] (Chapter 4).

- **A combination of semi-controlled field and laboratory studies on the utility of *AR* in goal-driven information browsing tasks** that indicate the importance to consider physical attributes, such as size, of information surfaces when employing *AR* user interfaces for information browsing tasks [88] (Chapter 4).

- **A concept and implementation for combining *AR* with complementary interface elements into hybrid user interfaces** on an individual handheld device for interacting with planar information surfaces in mobile contexts, which is the result of a user-centered iterative design process [83] (Chapter 4).

- **A series of studies investigating the utility of *AR* for verifying printed security documents** that highlight the subtleties of creating useful *AR* interfaces for document verification [102, 103] (Chapter 5).

- **A pipeline for enabling mobile *AR* interaction with public electronic displays** that lowers deployment costs of *AR* on public displays in order to facilitate the uptake of mobile *AR* in urban contexts [90] (Chapter 6).

- **Concepts, prototypes and user studies on the extension of mobile interaction beyond individual personal displays**. Specifically, a concept on seamless interaction with multiple displays on and around the body [85] (Chapter 6).

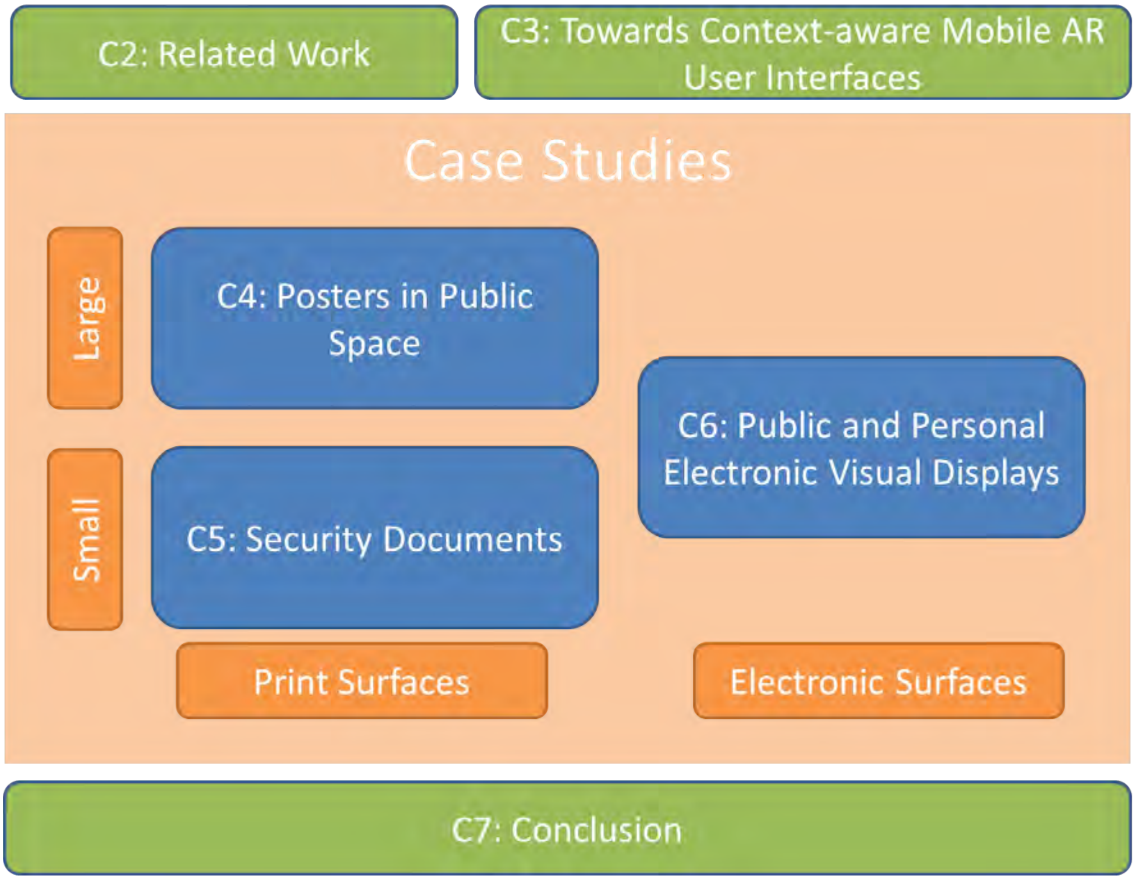

C2: Related Work
C3: Towards Context-aware Mobile AR User Interfaces
Case Studies
Large
Small
C4: Posters in Public Space
C5: Security Documents
C6: Public and Personal Electronic Visual Displays
Print Surfaces
Electronic Surfaces
C7: Conclusion

Figure 1.1: Overview of the thesis chapters. The chapters C2, C3, C7 provide the frame for the case studies presented in C4-C6.

1.1.1 Survey on context-aware AR

As information surfaces can be found in varying mobile contexts, it is important to consider the contextual factors that can influence *AR* interaction in the real world. Compared to other context-aware ubiquitous and mobile systems, the particularities of context-aware *AR* are often connected to the tight spatial link between the interactive system and the physical environment. This can have implications on visualization and interaction techniques for *AR* applications. Hence, it is worthwhile to study the role of context specifically for *AR* and to highlight distinct characteristics that are unique to *AR*. In our work [92], we contribute by providing:

- a taxonomy for context-aware *AR* systems.
- a comprehensive overview of how existing *AR* systems adapt to varying contexts.
- opportunities for future research on context-aware or adaptive *AR* systems.

We show that context-awareness is an important aspect for future *AR* applications, but is still widely underexplored. Specifically, while tracking is a relatively well-investigated area in context-aware *AR*, other fields (e.g., social factors, affective and perceptual factors, digital factors, configurations of input and output devices) are underexplored with a small number of seminal works. Furthermore, we identified that most of the existing works focus on integrating context sources into their system but do not demonstrate which context target (e.g., the system input or output) is adapted.

1.1.2 User survey on first generation AR browsers

AR browsers are mobile applications that provide access to information that is situated in the physical world and is accessible via web technologies. First generation *AR* browsers provided mainly access to geo-referenced Points of Interest (POIs), while newer generations allow interaction with planar surfaces (such as posters) or even physical 3D objects. Since their first appearance in 2008 (Wikitude[5]) on smartphones, *AR* browsers have become commercially successful. With over several commercial providers and over 40 million downloads in mobile app stores they are the most downloaded *AR* application type for consumers.

First generation *AR* browsers did not explicitly target information surfaces but focused on providing user interfaces to location-based data. Still, it is worthwhile to study them as they have been one of the first *AR* applications widely available to consumers in mobile contexts and hence can provide valuable insight on hedonic and utilitarian aspects of interacting with mobile *AR* systems.

So far, motivations for using *AR* browsers and usage patterns have been widely underexplored. We therefore conducted one of the first studies investigating the adoption and motivations of *AR* browser users. The study combined an online survey of *AR* browser users with an analysis of app market data.

The results of the study are presented in our technical report [86] and implications are discussed in a follow up article [140]. We found that while the usage of *AR* browsers is

[5] http://www.wikitude.com, last retrieved 20.04.2015.

often driven by their novelty factor, a substantial number of long-term users exist. The analysis of quantitative and qualitative data showed that poor and sparse content, poor user interface design and insufficient system performance are the major elements inhibiting the prolonged usage of this technology by early adopters.

1.1.3 Social context in AR gaming at posters in public space

With the advancement of 3D tracking technology for mobile application development, specifically computer-vision based tracking, new application scenarios are enabled. Those scenarios can deliver accurate 3D registration specifically on well textured planar surfaces, such as posters, and go beyond the limited interaction possible with sensor-based registration methods. Specifically, spatial interaction within arm's reach is enabled through Natural Feature Tracking (NFT) of nearby objects. Indeed, current generation *AR* browsers support *NFT* of physical objects, and companies like Layar[6] specifically focus on consumer experiences around print media.

One popular commercial use case is to support casual gaming at public posters (for example the Darksiders II game poster created by Blippar [7]). Similar, as for the first generation *AR* browsers, usage patterns for *AR* applications involving planar surfaces such as print media has not been studied in depth. Therefore, we started to probe this interaction space by investigating the usage patterns of *AR* gaming at posters in public contexts. Specifically, we were interested in the effects of the social context on the user behavior and hence conducted repeated evaluations in two public spaces with varying spatial and social characteristics and in a laboratory setting.

In public contexts, the visibility of interactions between users and computers can have a major effect both on the audience and in turn on the user herself [190]. Handheld *AR* allows rich spatial interactions without revealing the effects of those interactions to the audience. The gestures and postures involved in handheld interaction show resemblance with the acts of picture taking or deictic gestures, which is an established attention getter in human communication [242]. This may draw unwanted attention to the user.

Hence, we also contrasted the use of *AR*, with Static Peephole (SP) interaction, a socially accepted interface typically relying on less visible gestures [189, 197].

We conducted a series of semi-controlled field studies and a laboratory study [87, 89]. We found that, for a public space, where a noticeable social distance between participants and audience (as reported by participants) occurred, the *AR* interface was used significantly less and preferred less compared to a laboratory and another public condition, with different spatial and social aspects.

1.1.4 The utility of AR for information browsing at printed maps

Besides gaming, information browsing at planar surfaces such as printed maps is a popular application area for handheld *AR* that was envisioned by researchers for more than 10 years [215] and recently became also available in consumer contexts[8]. Early research investigated

[6] http://www.layar.com, last retrieved 20.04.2015.

[7] https://blippar.com/en/blipp, last retrieved 20.04.2015.

[8] http://www.tunnelvisionapp.com/, last retrieved 20.04.2015.

the applicability of *AR* for important information browsing tasks such as locator tasks, i.e., finding a target object with desired attributes among distractor objects [202]. However, the employed tracking technology in previous studies suffered from severe limitations such as a small operational range between handheld device and map (6-21 cm) and a low update rate of the tracker of only 10 Hz. We found that users adopt their behavior to the capabilities of the available tracking technologies for *AR* interaction [168]. Due to recent advances in computer-vision-based tracking (30 Hz update rate, large operational range of up to 200 cm in our studies [88]), it is advisable to re-investigate the potentials of *AR* for information browsing at public maps.

To this end, we investigated both performance and user experience aspects for *AR* browsing at printed maps [88]. In contrast to previous studies a semi-controlled field experiment in a ski resort indicated significant longer Task Completion Times (TCTs) for an *AR* interface compared to a *SP* interface. A follow-up controlled laboratory study investigated the impact of the workspace size on the performance and usability of both interfaces. We show that for small workspaces *SP* outperforms *AR*, confirming indications of previous studies. As workspace size increases, performance gets leveled out. Also, subjective measurements indicate less cognitive demand and better usability for *AR*. Our results indicate that *AR* might be a beneficial tool for interaction with public posters going beyond hedonic user experience aspects and adding utilitarian value to mobile interactive experiences.

1.1.5 Hybrid AR interfaces for poster interaction in mobile contexts

Our previous investigations indicate that *AR* interfaces on individual personal handheld displays can be of value for interacting with print media in public spaces. Still, there are circumstances when this interaction is inhibited and alternative interfaces might be more suitable.

Based on our previous observations, we created a hybrid interface for information access at public posters in a user-centered design process. Within this thesis, we understand as a hybrid user interface the combination of *AR* with alternative user interfaces modules in a single application. Our hybrid user interface combines the advantages of *AR* and *SP* interaction [83]. The design was informed by a user survey about information access at public posters. The survey results showed the opportunistic nature of information access at public posters and highlighted the need for enabling a continuous user experience even when users (have to) leave the poster. Our design process resulted in three design recommendations that were applied when we implemented and evaluated two prototypes. Based on our findings we propose following recommendations for designing hybrid *AR* interfaces for poster interaction [83]:

1. Allow users to explore information while away from the augmented surface. To this end, preserve the frame of reference of the physical surface.

2. If you employ complex 3D scenes think carefully what kind of interactions you want to support in an alternative view. Favor ease of navigation over complete navigability of the scene.

3. Minimize cognitive effort when transitioning between interaction spaces.

1.1.6 The utility of AR for verifying security documents

Besides large planar surfaces, we also investigated the utility of *AR* for small surfaces, i.e., documents that can be handheld by users. of Security elements of paper documents such as passports, visas and banknotes are frequently checked by inspection. In particular, view-dependent elements such as holograms are interesting, but the expertise of individuals performing the task varies greatly. *AR* systems can provide information on standard mobile devices for decisions on validity. We developed a series of handheld *AR* interfaces [102, 103] to support the interactive verification of view-dependent elements. Specifically, we:

- indicated the feasibility of checking view-dependent elements. with a mobile *AR* system using information from a real-time tracking system running on a consumer smartphone through a comparative user study
- iteratively designed and implemented follow-up prototypes with the aim of reducing *TCT* following three different interaction paradigms: precise alignment, constrained navigation and a hybrid approach.
- found that users preferred a user interface which did not exhibit the fastest *TCT* but gave users more freedom to move the device in 3D space.

Specifically, our last observation that users might prefer a user interface that allows for more freedom of movement over a faster but more constraining interface might be of interest for further studies on close range spatial maneuvering with handheld *AR* devices.

1.1.7 Facilitating AR interaction with public electronic displays

So far, our explorations concentrated on print media as an instance of planar surfaces. Within this thesis, we also investigated *AR* user interfaces for electronic displays.

Letting one or multiple users interact with situated displays through handheld devices is compelling for public-private display interaction or tabletop collaboration. With the proliferation of large format screens and handheld devices, "second screen" apps for handheld devices, providing background information for live TV programs, are becoming increasingly popular. Spatial interaction between handheld and situated displays should be the obvious next step. We believe that the major obstacle preventing spatial interaction between mobile and situated displays is the need for additional infrastructure. Previous attempts at showing perspectively correct overlays from the user's point of view have required stationary outside-in 3D tracking, often in combination with projectors. Such proof-of-concept implementations do not allow mobile operation outside the lab.

In our work, we addressed several limitations for interaction between mobile devices and situated displays [90]. First, our prototype provides Magic Lens (ML) (*ML*) interaction between situated displays and mobile devices with geometrically correct rendering from the user's point of view. Second, it only requires access to a screencast of the situated display, which can be easily provided through common streaming platforms and is otherwise self-contained. Our system performs all computations on the mobile device. Hence, it easily scales to multiple users.

1.1.8 Mobile interaction with multiple displays on and around the body

The rising trend in consumer-oriented displays on and around the body such as smartwatches, head-mounted displays (HMDs) and handheld displays has opened up new design possibilities for mobile interaction. In our work, we introduce MultiFi, a platform for designing and implementing user interface widgets across multiple displays with different fidelities for input and output [85]. MultiFi aims to reduce seams when interacting with individual devices and combines the individual strengths of each display into a joint interactive system for mobile interaction. Specifically, we:

- explore the design space of multiple displays on and around the body and identify key concepts for seamless interactions across devices.
- introduce a set of cross-display interaction techniques and applications such as mid-air pointing with haptic feedback or full screen virtual keyboards.
- present empirical evidence that combined interaction techniques can outperform individual devices such as smartwatches or head-mounted displays for browsing and selection tasks.

Through our findings we hope to spur future research for *AR* going beyond individual (often handheld) displays.

1.2 Results

The results of this thesis contribute to the fields of Augmented Reality, Mobile Human-Computer Interaction and Pervasive Computing by identifying benefits and challenges of mobile *AR* user interfaces for interaction with planar surfaces in consumer-oriented application contexts. In particular, the thesis provides following results:

- **Hedonic qualities of *AR* user interfaces should be carefully balanced with utilitarian qualities**. Based on our investigations of first generation *AR* browsers we could show that novelty was a main driver for using *AR* browsers. Hence, the hedonic value of those interfaces is high at the beginning of product use [22]. But as novelty wears of, so does the hedonic value of simple *AR* user interfaces. As attitude towards products is both influenced by hedonic and utilitarian values [15, 119], *AR* user interfaces that are primarily stimulating hedonic dimensions of the user experience without offering utility tend to be not used after the novelty effect wears of.
- ***AR* systems should consider context sources beyond mere location and time.** There is a large space of context sources which has not been considered in

depth for *AR* systems, but which provides rich possibilities for optimizing the use of *AR* in dynamic situations. Similarly, more context targets in *AR* systems should be considered when trying to adapt *AR* to varying situations. Our survey on context-aware *AR* systems [92] can be seen as a guideline which context sources and targets could be explored in the future.

- **Specifically, the social context of interaction** should be considered when deploying *AR* interfaces in public space. Similar to recent findings regarding interactive installations [2], we identified that the usage patterns with mobile *AR* games are influenced by the social properties of a public space. Specifically, inappropriate social contexts could inhibit the use of rich spatial interactions with *AR*.

- **Physical properties of media artifacts can influence the utility of *AR* interfaces**. Specifically, we showed that for **information browsing tasks**, the size of the information space can be crucial for users to experience benefits of *AR* interfaces when compared to traditional touch-controlled map interfaces. For example, in our studies, users did not see benefits of *AR* for common poster sizes of DIN A0 but found *AR* more useful as the workspace size increased.

- ***AR* can be beneficial for micro-tasks when interacting with security documents**. For small physical media such as security documents, we showed that *AR* can be used for the verification of security features, but does not necessarily provide a competitive performance compared to established verification workflows. However, *AR* can be beneficial for specific subtasks, such as the detailed verification of individual document elements after an initial rough verification step.

- Consequently, as the utility of *AR* depends on the nature of the task and the dynamic context, ***AR* should be integrated with complementary means of interaction into hybrid user interfaces** to allow users reach their goals in dynamic usage situations. For planar surfaces, we propose to combine *AR* with ***SP*** interfaces when no complex spatial navigation or manipulation of the scene is required.

- **Deployment costs of *AR* user interfaces for public displays should be kept to a minimum** to facilitate the provisioning of rich context sources in public spaces. Specifically, *AR* systems should augment public electronic displays in a self-contained way, without the need for costly server infrastructure. Within this thesis, a prototype is presented, which demonstrates that low-cost deployment of digital displays suitable for *AR* interaction is possible.

- **Interaction across multiple wearable displays can outperform interaction with individual displays**. We found that interacting with multiple wearable displays such as HMDs and smartwatches can be more efficient than interaction with a single wearable display only. However, this increase in efficiency can come at the cost of a higher workload.

1.3 Publications and collaboration statement

This thesis encompasses publications that are based on collaborations between researchers from various institutions. In the following, an overview of publications, that this thesis is based on, and the people who were involved in the creation of them, is given.

The following publications summarize studies about usage patterns and motivations for using current generation *AR* browsers. They influenced the creation of prototypical *AR* user interfaces in this theses.

- **Jens Grubert**, Tobias Langlotz, Raphael Grasset (2011). *Augmented reality browser survey*, Technical Report 1101, Institute for Computer Graphics and Vision, University of Technology Graz, 2011 [86].
  The author of this thesis developed and analysed the questionnaire for the online survey. The reflections on the motivations and usage patterns of AR browsers in the technical report reflect mainly his viewpoints. ***Raphael Grasset*** *contributed to the questionnaire as well to the reflections and design considerations, whereas* ***Tobias Langlotz*** *contributed by analysing data from mobile distribution platforms.*

- Tobias Langlotz, **Jens Grubert**, Raphael Grasset (2013). *Augmented reality browsers: essential products or only gadgets?*. Communications of the ACM, 56(11) (pp. 34–36) [140].
  The author of this thesis contributed by co-creating the structure and argumentation of the article, specifically, by reflecting on current usage patterns and the role of web-based technologies for the success of future AR browser generations. ***Raphael Grasset*** *contributed to the reflections on the article content whereas* ***Tobias Langlotz*** *contributed by developing and structuring the views on general AR browser technology.*

One main insight of the preciding publications was that, currently, early adopters used first generation *AR* browsers in mobile contexts mainly for curiosity reasons and that they did not see many advantages beyond novelty. This outcome, together with the observation that interaction with planar surfaces can take place in various mobile contexts, triggered research to investigate which *context factors* can be potentially relevant for mobile *AR* interaction. Hence, as background for the presented prototypes an overview of *AR* systems which are context-aware is presented in

- **Jens Grubert**, Stefanie Zollmann and Tobias Langlotz (2015). *Context-aware augmented reality: trends and opportunities*, submitted to Transactions of Visualization and Computer Gaphics [92].
  The author of this thesis was the principal investigator. He developed the employed taxonomies and provided the related work on context-awareness. **Stefanie Zollmann** and **Tobias Langlotz** together with the author conducted the literature reviews and provided further valuable input on the taxonomy.

The previous publication identified research opportunities for investigating the role of *AR* in changing contexts. Amongst others, it indicated that to date there is a lack of research on understanding the influence of social factors on *AR* interaction. Consequently, given these research opportunities, the following publications investigated the potential of *AR* in mobile contexts for two major applications of print media: casual gaming and goal-driven information browsing.

The following publications reflect on the usage of *AR* interfaces for gaming at printed posters in public contexts. They highlight the importance of social factors which can influence the user experience of *AR* interfaces in public spaces. Specifically, they compare *AR* as an interface with visible actions, but hidden effects [190], with a private and established zoomable map interface:

- **Jens Grubert**, Ann Morrison, Helmut Munz, and Gerhard Reitmayr (2012). *Playing it real: ML and SP interfaces for games in a public space.* In Proceedings of the 14th International Conference on Human-Computer Interaction with Mobile Devices and Services (pp. 231–240). ACM [87].
The author of this thesis was the principal investigator and responsible for planning, conducting, and evaluating the study, including the implementation of the employed prototype. ***Ann Morrison*** *gave valuable reflections on the study design and together with* ***Gerhard Reitmayr*** *helped editing the paper.* ***Helmut Munz*** *contributed by providing 3D assets for the prototype.*

- **Jens Grubert**, Dieter Schmalstieg (2013). *Playing it real again: a repeated evaluation of ML and SP interfaces in public space.* In Proceedings of the 15th International Conference on Human-computer Interaction with Mobile Devices and Services (pp. 99–102). ACM [89].
The author of this thesis was the principal investigator and responsible for planning, conducting, and evaluating the study, including the implementation of the employed prototype.

Besides gaming scenarios, we also investigated goal-driven information browsing tasks at public posters. Hedonic aspects are a crucial part of the user experience (specifically, in gaming and advertising), but they should not be the sole focus of *AR* systems and should complement utility aspects, which *AR* can potentially offer to mobile users. Hence, the following article focuses on the utility of *AR* in dependence of spatial properties of planar surfaces. Specifically, the article highlights the effect that the physical size of a poster can have on the utility of *AR*, when compared with established zoomable map interfaces:

- **Jens Grubert**, Hartmut Seichter, Michel Pahud, Raphael Grasset and Dieter Schmalstieg (2015). *The utility of ML interfaces on handheld devices for touristic map navigation.* In Pervasive and Mobile Computing. Vol. 18 (pp. 88-–103). Elsevier [88]. *The author of this thesis was the principal investigator leading the design, implementation and evaluation of the studies.* ***Hartmut Seichter*** *contributed by implementing technical components for the laboratory study and together with the other authors gave valuable input to the design of the laboratory study and structuring of the article.*

The preceding publications highlighted that *AR* interfaces can be beneficial for interacting with poster-sized print media under specific circumstances. Due to the dynamic nature of mobile contexts these specific circumstances can not always be met (e.g., through dynamic behavior of spectators, large distance to poster, mobility of the user). Hence, in the following publication a hybrid design of *AR* and zoomable map interface is proposed which eases the transition between *AR* and other user interfaces when interacting in mobile contexts:

- **Jens Grubert**, Raphael Grasset and Gerhard Reitmayr (2012). *Exploring the design of hybrid interfaces for augmented posters in public spaces.* In Proceedings of the 7th Nordic Conference on Human-Computer Interaction: Making Sense Through Design (pp. 238–246). ACM [83]. *The author of this thesis was the principal investigator and leading the design and evaluation of the study, the implementation of the prototype and case studies as well as the conceptualization of the design space.* ***Raphael Grasset*** *gave valuable feedback on structuring the design space.*

Besides large planar surfaces, we also investigated the utility of *AR* for small surfaces. Specifically, we investigated how *AR* can support laymen in the verification of security documents in the following publications:

- Andreas Hartl, **Jens Grubert**, Dieter Schmalstieg and Gerhard Reitmayr (2013). *Mobile interactive hologram verification.* In Proceedings of the IEEE International Symposium on Mixed and Augmented Reality 2013 (pp. 75-82). IEEE [103]. *The hologram detection and tracking system was implemented by* ***Andreas Hartl*** *and* ***Gerhard Reitmayr****. The author of this thesis was contributing to the design of the user interface (guidance system) and was responsible for planning, conducting and evaluating the user study.*

- Andreas Hartl, **Jens Grubert**, Clemens Arth and Dieter Schmalstieg (2014). *Mobile user interfaces for efficient verification of holograms.* In Proceedings of the IEEE Virtual Reality Conference 2015 (to appear). IEEE [102]. *The enhanced hologram detection and tracking system was implemented by* ***Andreas Hartl****. Together with Andreas Hartl the author of this thesis was designing the user interfaces and was responsible for planning, conducting and evaluating the user studies.*

The development of the *AR* user interfaces for hologram verification also resulted in the following patents (issued, in publication):

- Andreas Hartl, **Jens Grubert**, Dieter Schmalstieg, Gerhard Reitmayr and Olaf Dressel (2014). *Verfahren zur Ausrichung an einer beliebigen Pose mit 6 Freiheitsgraden für AR Anwendungen (Procedure for view-alignment to an arbitrary six degrees of freedom for Augmented Reality applications)* [104].

- Andreas Hartl, **Jens Grubert**, Dieter Schmalstieg, Gerhard Reitmayr and Olaf Dressel (2014). *Aufnahme der SVBRDF von blickwinkelabhängigen Elementen mit mobilen Geräten (SVBRDF capture of view-dependent elements with mobile devices)* [105].

This thesis also investigates *AR* user interfaces for electronic displays in mobile contexts such as public signage systems. The following publication investigates how to minimize the cost of deploying *AR* experiences to public displays and how to enable perceptually beneficial user perspective rendering for those displays:

- **Jens Grubert**, Hartmut Seichter and Dieter Schmalstieg (2014). *Towards user perspective augmented reality for public displays.* In Proceedings of the IEEE International Symposium on Mixed and Augmented Reality 2014 (pp. 339–340). IEEE [90]. *The author of this thesis was the principal investigator and leading the design and implementation of the technical prototype as well as the structuring and writing of the article.* ***Hartmut Seichter*** *contributed by provisioning technical components, e.g., for NFT, needed for the prototype.* ***Dieter Schmalstieg*** *helped to streamline the paper content.*

Turning from large public displays to small personal wearable displays we investigated how *AR* user interfaces can benefit interaction across multiple displays on and around the body:

- **Jens Grubert**, Matthias Heinisch, Aaron Quigley and Dieter Schmalstieg (2015). *MultiFi: multi fidelity interaction with displays on and around the body.* In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems 2015 (pp. 3933–3942). ACM [85]. *The author of this thesis was the principal investigator, leading the design and evaluation of the prototypes and supervised the work of* ***Matthias Heinisch****, who primarily implemented the prototypical system.* ***Aaron Quigley*** *and* ***Dieter Schmalstieg*** *contributed through discussions about the concept and evaluation of the system and helped to streamline the paper content.*

Within the user studies in this thesis a triangulation of quantitative and qualitative methods was targeted. While investigating selected user studies it became apparent that the tracking quality of *AR* system is often neglected as a potential confounding factor. Consequently, we reflected on the potential effects tracking can have on the outcome of *AR* focused user studies.

- Alessandro Mulloni, **Jens Grubert**, Hartmut Seichter, Tobias Langlotz, Raphael Grasset, Gerhard Reitmayr and Dieter Schmalstieg. *Experiences with the impact of tracking technology in mobile augmented reality evaluations.* In the MobiVis workshop at the International Conference on Human-Computer Interaction with Mobile Devices and Services 2012 [168]. *The author of this thesis was, together with* ***Alessandro Mulloni****, the principal investigator. He contributed with the analysis of his previous user studies as well with the structuring and writing of the article. The remaining authors contributed by reflecting on studies in which they were involved.*

2 Related Work

This chapter presents related work that is central for understanding the context of this thesis. Since this thesis focuses on mobile *AR* interaction with information surfaces, a history of *AR* is presented, followed by a review of interaction with printed and electronic information surfaces. This leads to a discussion of hybrid interfaces, which combine *AR* with alternative user interface elements. Furthermore, relevant user studies are presented, which encompass both *AR* and alternative user interfaces for interacting with information surfaces. Finally, it is summarized how the notion of context-awareness was investigated in prior work, as considering the context of interaction is central to the presented mobile *AR* user interfaces in this thesis.

2.1 Towards mobile AR

The vision of overlaying digital information over physical environments dates back over 100 years [16]. First implementations of this vision appeared in the 1960s. Ivan Sutherland's "The sword of Damocles" is considered one of the first working *AR* and Virtual Reality systems, incorporating a fully six Degrees of Freedom (DOF) tracked Optical See-Through (OST) Head-Mounted Display (HMD). In 1992, Caudell and Mizell introduced the term Augmented Reality [44]. They used *HMDs* to guide workers in the manual manufacturing processes. Interaction with the system was envisioned through voice control and a hip mounted indirect input device. In 1993, Feiner et al. also explored the use of head-mounted displays for maintenance tasks with their KARMA (Knowledge-based Augmented Reality for Maintenance Assistance) system [68]. The system relied on tracking the user's head position for rendering 3D graphics. Also, in 1993, Fritzmaurice introduced a handheld device (the "Chameleon Lens") to augment physical information spaces [70]. The system was capable of rudimentary object selection through raycasting from the screen center. Similarly, Rekimoto explored handheld *AR* in his NaviCam system, which combined a palmtop display with vision-based fiducial tracking [194]. Rekimoto showed that a target acquisition task could be performed significantly faster with the handheld device compared to a head-worn display. Azuma presented a first survey of *AR* in 1997 [5]. According to Azuma, *AR* systems are defined through three main characteristics: *AR* (1) combines real and virtual, (2) is interactive in real time and (3) is registered in three dimensions. Towards the end of the 1990s, several mobile *AR* systems for outdoor use emerged as wearable (i.e., backpack) variants of desktop systems, notably the Touring Machine [67], M*AR*S [120] and Tinmith [184]. The touring machine combined a *HMD* with a handheld display with touchpad used as indirect input device. Starner et al. presented the "Remembrance Agent", a wearable and context-aware *AR* system which combined *HMD* and sensing [229]. While it was mostly a text-based system, it demonstrated, amongst others, finger tracking for input and face recognition. In 1999, Spohrer presented the idea of the Worldboard, a global infrastructure to associate information with places: content getting geo-referenced (rather than a URL), and being visualized with *AR* (rather than a 2D HTML renderer) [228]. Similarly, Kooper et al. presented the Real-World Wide Web as an information space of World Wide Web that is perceived using *AR* [136]. Similarl work have also explored non-visual direct augmentation, such as geo-located post-it [195] or audio augmentation [20].

Since 2000, Glspda started to gain sufficient processing power to perform relevant computations locally or at least to integrate tracking results from a server at interactive frame rates. Consequently, researchers started to turn their focus from desktop size back-pack systems to these smaller PDAs. For example Newman et al. presented the BatPortal, a wireless PDA-based *AR* system using radio-frequency based tracking in a building [171]. In 2001, Vlahakis et al. presented PDA-based system for outdoor environments ("Archeoguide") [251]. It was used in a cultural heritage context and used the Global Positioning System (Global Positioning System (GPS)) for the registration of 3D models on ancient artifacts. In 2003, Wagner and Schmalstieg adapted *AR*ToolKit [131], a fiducial-based pose tracking library, to off-the-shelf PDAs [253]. In 2006, Raskar et al. also used PDAs for a Spatial Augmented Reality (S*AR*) system on handheld devices ("iLamps) [188]. It used a handheld projector-camera system to estimate the the display surface geometry and subsequently project augmentations onto the surface.

So far, most of the developed systems relied on either external tracking systems (radio-frequency based, optical outside-in), imprecise sensors such as *GPS* and compass or on visual tracking of simple fiducials. In contrast, in 2006, Reitmayr et al. introduced hybrid tracking for *AR* in urban environments [193]. It combined a model-based edge tracker with gyroscope, gravity and magnetic field measurements. In 2007, Klein and Murray introduced Parallel Tracking and Mapping (PTAM), an approach for concurrent Simultaneous Tracking and Mapping (SLAM), by separating the mapping and tracking tasks in two different threads [134]. The system was later ported to mobile phones and quickly became used in the *AR* community [135]. In 2008, Wagner et al. presented one of the first *NFT* approaches suitable to run at interactive framerates on mobile phones [252].

In 2007, Apple presented the iPhone, a soon to be popular smartphone, and, in 2008, a novel distribution platform, the Apple App Store, for distributing mobile applications ("apps"). This platform, as well as similar distribution platforms such as the Google Play store, became relevant for distributing *AR* apps. In the same year, *AR* browsers started to emerge. *AR* browsers provide access to location-based information by overlaying graphical symbols such as labels and icons onto a live camera view of the environment. In the first generation *AR* browsers, such as Wikitude[1] or Layar[2], registration was achieved through the use of *GPS* and compass data. Those *AR* browsers quickly became adopted by consumers, soon exceeding millions of downloads[3]. Academic projects started to explore the concept of *AR* browsers, too. M*AR*A was one of the first mobile *AR* browsers using inertial sensors [126]. In 2011, MacIntyre et al. introduced the *AR*GON browser, which used a new data format for managing interactive *AR* content based on existing web ecosystem [156]. Also at that time, first standardization efforts were initiated[4]. With *NFT* techniques becoming (freely) available in *AR* Software Development Kits (SDKs), such as Qualcomm Vuforia[5], new use cases for *AR* browsers involving information surfaces were commercially explored. Specifically, augmented print solutions, i.e., augmentations of printed information surfaces, such as magazines or posters for advertising purposes, were explored by companies like Layar and Blippar[6].

There is also a considerable amount of work investigating the interaction between *ML* interaction on handheld devices and large electronic information surfaces. The metaDESK used both active and passive tangible *ML* for tabletop interaction [245]. Much later, the PaperLens reduced the infrastructure to a projection on paper and a table surface, but still required a calibrated stationary setup [226]. Alternative approaches allow tracking of mobile devices on tabletop systems [178], again relying on an external tracking solution. Virtual Projection does not need stationary tracking hardware, but instead proposes a client-server approach [17]. Mobile clients send a video stream to the server, which is responsible for tracking. This approach requires a bi-directional network connection, which may be hard to accomplish in public settings. Moreover, network bandwidth consumption and server load increases linearly with the number of clients and thus does not scale well.

Also, relevant for this thesis is a user-perspective rendering on handheld *AR* devices. Baricevic defined user-perspective rendering as the “geometrically correct view of a scene from the point of view of the user, in the direction of the user’s view, and with the exact view frustum the user should have in that direction” [14]. Hill et al. called this approach “Virtual Transparency” [113]. Copic et al. indicated that users expect user-perspective rendering in *AR*, i.e., the *AR* device to act as a transparent frame [48]. Current implementations of user-perspective rendering either rely on distorting the video feed of the back-facing camera [113, 243] or are using coarse 3D reconstructions [14]. Both approaches can suffer from visual artifacts, as the acquisition of the real world data through cameras or reconstruction is imperfect.

2.2 Mobile interaction with information surfaces

While this thesis focuses on *AR* as user interface for information surfaces, further interaction metaphors are also relevant, as they potentially allow interaction in circumstances where *AR* is not a suitable choice. Hence, this section will investigate related mobile

[1] https://www.wikitude.com/, last retrieved 20.04.2015.

[2] https://www.layar.com/, last retrieved 20.04.2015.

[3] https://www.layar.com/news/blog/tags/stats/, last retrieved 20.04.2015.

[4] http://www.perey.com/ARStandards/, last retrieved 20.04.2015.

[5] https://www.qualcomm.com/products/vuforia, last retrieved 20.04.2015.

[6] https://blog.blippar.com/en/blog/176-understanding-roi-in-ar, last retrieved 20.04.2015.

interaction techniques for printed and digital information surfaces. This interaction can be seen as subspace of Mobile Interaction with the Real World (MIRW), a term coined by Rukzio for research that investigates the "interplay between users and physical objects in the proximity using handheld devices as mediator for the interaction" [207]. *MIRW* itself can be seen as an intermediate step to Weiser's ubiquitous computing vision [257] who explicitly stated that Ubiquitous Computing "will not require that you carry around a *PDA*" [258]. At the time of writing of this thesis, consumer-oriented *OST* displays such as Google Glass[7], Microsoft Hololens[8] or Magic Leap[9] are (about to be) probing the market, so far most related work has concentrated on handheld devices. Handheld pico-projectors (e.g., [49, 98]) may be an alternative output channel, but they are beyond the scope of this thesis.

Interaction with information surfaces through mobile user interfaces on handheld devices has been considered from several viewpoints.

On the one hand, sensing technologies can establish a link between physical artifacts and digital information. They typically encompass visual tags (e.g., QR codes), radio-frequency-based tags (e.g., radio-frequency identification [255] and near-field communication), recognition of visual features of the information surface itself through computer-vision-based object recognition [80] or, in the case of digital information surfaces, recognition of imperceptible codes (e.g., [259]).

On the other hand, sensing techniques for recognizing user input aimed at the handheld device itself play an important role. Besides touch screens, sensors, employed for interaction typically found on commodity handheld devices, encompass cameras, acceleration and orientation sensors. Before becoming commonplace on handheld devices, those sensors were already investigated on *PDAs* in 2000 by Hinckley [115]. Besides input on the device itself, around-device interaction was explored, i.e., input to handheld devices using the surrounding space. Sensors used here encompass for example, infrared sensors [39, 137], microphones [260], magnetometers [4, 99], cameras [224, 254] or depth sensors [223]. More recently, depth sensors are being miniaturized and integrated into handheld devices, as demonstrated e.g., by Google[10] or Intel[11].

Given these sensing technologies, there are various interaction tasks that can be performed. Here, we concentrate on tasks relevant for interacting with information surfaces through handheld devices. An overview of other atomic tasks, which can be performed on the mobile phone without additional physical artifacts, is presented by Ballagas et al. [11].

Retrieving information is a popular use case that was investigated by several researchers. RFID and NFC tags were explored to retrieve (or add) information through touching (e.g., [73, 162, 208, 209, 255]). Interacting with services, such as buying a ticket at a movie poster, was also explored [34]. Several studies showed that pointing was preferred over using the on-screen user interface of mobile phones (e.g., [35, 162, 212]).

[7] https://www.google.com/glass/start/

[8] http://www.microsoft.com/microsoft-hololens/en-us, last retrieved 20.04.2015.

[9] http://www.magicleap.com, last retrieved 20.04.2015.

[10] https://www.google.com/atap/projecttango/, last retrieved 20.04.2015.

[11] http://www.intel.com/content/www/us/en/architecture-and-technology/realsense-overview.html, last retrieved 20.04.2015.

However, depending on the number of elements that can be selected, simplistic on-screen user interfaces might still be more efficient [41]. This is due to the fact, that users might need several attempts to select an item through touch [41].

Touching requires users to be at close physical proximity to the information surface. Using pointing as an interaction technique allows to expand the interaction radius considerably. Instead of having direct contact with a tag, users aim with the camera of the mobile device at a visual tag. Several works have investigated visual tags as means to retrieve information from physical objects (e.g., [30, 43, 133, 153]).

An intermediate technique is scanning, which has an operational range between touching (contact with the surface) and pointing (at large distances). Depending on the employed sensing technique (e.g., Bluetooth, infrared), it requires user to be at close proximity to a physical artifact without the need to physically touch it.

Several studies [36, 162, 208, 209, 211] compared these three approaches of touching, scanning and pointing, but could not come up with universal recommendations. The findings indicated the the most suitable interaction technique depends on context factors such as location, motivation, activity or required reliability (e.g., in the case of drug identification).

In contrast to the relative short duration of the previously presented discrete interaction techniques, continuous d techniques (partly in combination with discrete ones) can support more complex interaction tasks such as navigation of an information space or object manipulation.

Several works have investigated how handheld displays can be used to continuously interact with physical information surfaces, with a focus on situated electronic displays. For example, Ballagas et al. demonstrated how to control a remote cursor on a distant display through spatial sensing [12]. Boring explored further techniques to control a remote display cursor (scrolling, tilting or translating the handheld device) [33].

In 2010, Boring et al. explored how to move content on and across electronic displays at a distance, using pointing with a handheld display [31]. The authors implemented a number of improvements over a naive raycasting approach, which would not work reliably. Specifically, they allowed to virtually zoom into a remote display to enlarge the target selection area and could temporarily freeze the camera view for more convenient poses while retaining a live stream of the target display. Baldauf et al. investigated how to transfer files between a private handheld and a remote public display through pointing [10].

Some works also investigated multi-user interaction at public displays. Boring et al. extended the concept of touch projector [31] to allow multi-user interaction at a media facade [32]. Users could collaboratively (or competitively) draw images on a low resolution media facade. Baldauf presented the “augmented video wall”, which allowed multiple users to concurrently overlay private views (videos) onto a public display [7, 9].

Besides interacting with physical artifacts, interaction techniques for navigating virtual information surfaces are also reviewed here. Specifically, *SP* and Dynamic Peephole (DP) interaction have been popular for navigating virtual information spaces such as digital maps [264]. While *SP* interfaces typically move a scene behind a fixed virtual window (i.e., traditional pan and zoom using touch input), *DP* interfaces keep the information space fixed and move a viewing window (or virtual camera) over it by sensing the spatial input of users.

2.3 Hybrid user interfaces

The previous two sections described related work encompassing handheld *AR* and alternative (non-*AR*) user interfaces for interacting with physical objects. This section provides an overview on user interfaces that aim at combining the strength of multiple interaction metaphors into one experience.

In this thesis, hybrid user interfaces are understood as the combination of *AR* with alternative user interfaces. They have been considered already over 10 years ago by Billinghurst et al. [26]. Their MagicBook combined illustrations in a real book with *AR* and immersive Virtual Reality views. Preceding work of Feiner et al. coined the term hybrid interface with a slightly different connotation, namely to combine different VR and desktop devices in one physical reference space [69]. Later the notion of transitional interfaces, which allows fluidly changing between interfaces, was introduced [78]. Recent examples of transitional interfaces include zooming interfaces for *AR* Browsers [167], view transitioning for distributed outdoor cameras [249] and indoor navigation [163, 164]. An overview on combinations of *AR* with complementary interfaces (like maps, world in miniature, distorted camera views and virtual environments) can be found in the survey of Grasset et al. [79]. The common theme of many existing hybrid and transitional user interfaces is that they are grounded in one (potentially large and distributed) physical reference frame.

Most of the previous work concentrated on combining several interface metaphors on a single input and output device (either a single handheld device or an *HMD*). However, today mobile users often have access to several devices at once. Besides having a smartphone or a tablet, available smartwatches and smartglasses are becoming popular. These novel devices have individual benefits and drawbacks in mobile interaction scenarios. For example, today's dominant handheld devices, smartphones and tablets, have a high access cost in terms of the time and effort it takes to retrieve and store the device from where it typically resides, such as one's pocket. This cost reduces the usefulness of a device for micro-interactions, such as checking the time or one's inbox. In contrast, wearable devices such as a smartwatch or *HMD* lower the access cost to a wrist flick or eye movement.

However, interaction with these always-on devices is encumbered by their low fidelity: limited screen and touch area, low resolution and poor contrast limit what users can do. Currently, HMDs require indirect input through touch devices, while high-precision spatial pointing is not yet commercially available.

A recurring topic for wearable displays is the extension of display real-estate using virtual screen techniques [66, 70, 192]. Recently, Ens et al. [64] explored the design space for a body-centric virtual display space optimized for multi-tasking on HMDs and pinpointed relevant design parameters of concepts introduced earlier by Billinghurst et al. [25, 27]. They found that body-centered referenced layouts can lead to higher selection errors compared to world-referenced layouts, due to unintentional perturbations caused by reaching motions.

Users with multiple devices tend to distribute tasks across different displays, because moving between displays is currently considered a task switch. For some forms of interaction, a tight spatial registration may not be needed. For example, Duet combines handheld and smartwatch and infers spatial relationships between the devices based on local orientation sensors [45]. Similarly, Billinghurst et al. [38] combine handheld and *HMD*, but use the handheld mainly as an indirect input device for the *HMD*. Specifically, handheld and *HMD* have no spatial knowledge of each other. Stitching together multiple tablets [116] allows for interaction across them, under the assumption that they lie on a common plane. Several other approaches combine larger stationary displays with handheld displays through spatial interaction [19, 31]. The large stationary displays make virtual screens unnecessary, but restrict mobility. The same is true for the work of Benko et al. [21], who combine a touch table with an *HMD*. Yang and Widgor introduced a web-based framework for the construction of applications using distributed user interfaces but do not consider wearable displays [263].

2.4 User studies of spatially-aware mobile interfaces

This section gives an overview of user evaluations in the field of mobile *AR* and other spatially-aware user interfaces that are relevant for interacting with information surfaces.

Controlled studies of *ML*, *SP* and *DP* interaction encompass fundamental interaction tasks such as target acquisition tasks and visual search tasks (finding a target object among distractors) and higher level tasks such as navigation.

Mehra et al. compared *DP* and *SP* metaphors for line-length discrimination using a desktop PC interface with mouse input [160]. Their results indicated that *DP* interfaces are superior to *SP* interfaces for tasks in which spatial relationships matter and display size is limited. In 2008, Rohs and Oulasvirta investigated target acquisition performance with *ML* and *DP* interfaces on a handheld device [199] and formulated a two part pointing model for *ML* including coarse physical and fine-grained virtual pointing. They also validated their model in a real-world pointing task for varying target shapes and visual contexts [200]. Cao et al. investigated peephole pointing for dynamically revealed targets [40] using a desktop PC and graphics tablet. The authors focused on a one-dimensional pointing task both for coupled cursor position (fixed on the screen center) and decoupled cursor position (independent of screen position). These fundamental target acquisition studies are important as building blocks for designing spatially-aware user interfaces. However, to our knowledge, human movement models like Fitt's law cannot easily be employed to predict performance of exploratory map navigation tasks. Those map related tasks involve building up survey knowledge and path planning in the presence or absence of a physical map [214].

To this end, Rohs et al. compared *ML*, *SP* (via joystick control) and *DP* interaction for explorative map navigation [201]. They evaluated performance, motion patterns and user preferences for a locator task. They found that both *DP* and *ML* interaction outperformed *SP* navigation in terms of *TCT* and degree of search space exploration but did not find significant differences between *DP* and *ML* interaction. Rohs et al. extended their previous study to include the impact of item density on *ML* interaction [201]. They found that the effectiveness of the visual context (*ML*) decreases with increasing item density compared to *DP*. In their studies participants generally preferred *ML* over *DP* interaction. They also found, that the availability of visual context (in the *ML* condition) led to more guided search patterns, whereas the *DP* condition resulted in search patterns that uniformly covered the map. Technical limitations of the studies included the small operational range between handheld device and map (6-21 cm) and the low update rate of the tracker of only 10 Hz. Mulloni et al. indicated that users adopt their behavior to the capabilities of the available tracking technologies for *AR* interaction [168]. Hence, it seems advisable to conduct comparisons between interfaces when the underlying technologies change significantly, as it is the case with current *AR* tracking technologies. As of 2015, computer vision-based tracking technology can be deployed in real-world environments supporting tracking with 30 Hz update rate and a vastly wider operational range.

Goh et al. investigated usability and perceptual aspects of three interfaces for searching and browsing geolocation-based information including *ML*, *SP* and list views [76]. Their results indicated that for searching, performance was similar across all three interfaces but for browsing, the map performed significantly worse than the list and *AR* interfaces. Also, the *AR* interface was always ranked last in terms of usability despite its better performance when compared to the map. However, Goh et al. did not address aspects of user experience measures beyond usability. Dünser et al. compared *AR* to *SP* interfaces for navigation to *POIs* [61]. They found no performance differences between both types of user interfaces, but they indicated that the *AR* interface could be less useful in certain contexts. In another application task oriented study, Yee et al. compared a peephole interface to a conventional pen operated scrolling interface in a performance oriented user study for selection, route planning and drawing tasks. The authors indicated mixed results (no significant differences in error rates or for the route planning task) [264]. Baldauf et al. compared the performance of two orientation-aware (including *ML*) and two orientation-agnostic techniques for interacting with public displays through a smartphone in pointing, drag and drop and drawing tasks [8]. Their results indicated, that *ML* interaction is well suited for spontaneous pointing tasks with short interaction periods. While *ML* interaction could not outperform an orientation-agnostic alternative, participants found *ML* interaction more intuitive and fascinating. Recently, Pahud et al. compared *DP* and *SP* for map navigation [180]. No performance advantage for *DP* could be identified for their selection tasks, where the participants had to navigate (by panning and/or zooming) to locate a specific target on a map, before selecting it. However, they observed that *DP* outperformed *SP* for repetitive back/forth navigation and selection tasks between two known targets. This observation would reinforce the opportunity to design *DP* experiences such as virtual shelves [151], or tool menus in specific locations in space. Pahud et al. also mentioned that *DP* seems to also have an opportunity with compound tasks such as navigate and trace. In contrast to the work of Pahud et al., Spindler et al. found that an *DP* interface could significantly outperform *SP* navigation for navigation tasks involving panning and zooming in an abstract information space [227].

While there is a large number of performance-based user studies on spatially-aware displays, to date, there are comparably few studies focusing on qualitative aspects of spatially-aware mobile interaction. Olsson et al. presented one of the few studies that explored users' experiences with mobile *AR* [177]. They conducted an online survey to explore the most satisfying and unsatisfying experiences with mobile *AR* applications.

Their results, confirm research outcomes by us (see Chapter 3)and conclude that mobile *AR* browsers are still mainly used due to their novelty value. Furthermore, qualitative aspects in collaborative settings of mobile *AR* were addressed by Morrison et al. [166]. They conducted field trials using ethnographic observation methods on the collaborative use of handheld *AR* with a single device [166] and later expanded their observations to synchronous use of multiple mobile devices [165]. One finding was that *AR* facilitates place-making and that it allows for ease of bodily configurations for the interacting group. This could indicate enhanced user experiences over traditional user interfaces.

2.5 Context-awareness

Context and context-awareness have been thoroughly investigated in various domains such as ubiquitous computing, intelligent user interfaces or recommender systems. Theoretical foundations about the semantics of context have been discussed in previous work, e.g., [58]. Different taxonomies and design frameworks, e.g., [55, 267] as well as numerous software-engineering models for context and contextual reasoning have been proposed by other research groups, e.g., [110]. In addition, comprehensive reviews of context-aware systems and models were published, e.g., [6, 23, 230]. There have been discussions if capturing context in a general sense is of any use to inform the design (and operation) of mobile and ubiquitous systems as it is tightly bound to the users' internal states and social context [57, 81]. We argue that it is worthwhile to make these various context sources explicit, even though we might not have the means to measure all possible sources, yet (such as users' cognitive state). Within this thesis, we follow the generic notion of context by Dey et al. as "any information that can be used to characterize the situation of an entity. An entity is a person, place, or object that is considered relevant to the interaction between a user and an application, including the user and applications themselves" [53]. Similar to discussions about several context aspects, diverse taxonomies and design frameworks to capture context factors have been proposed. While philosophical aspects of context have been discussed [58, 235], the majority of existing works deals with technology-oriented approaches. For example, in the domain of pervasive systems, Abowd et al. introduced the primary context factors of location, identity, activity and time to address the questions of where?, who?, what?, and when? [53]. The authors also proposed to model secondary context factors, i.e., factors that are subcategories of primary factors (e.g., the e-mail address as subcategory of "what"), which could be indexed by the primary factors. Schmidt et al. proposed a working model for context with the two primary factors physical environment and human factors [218]. They express several cascaded factors and features within these two primary factors. Examples include user habits and affective state, users' tasks, co-location of other users and interaction with them for human factors. The physical environment includes location, infrastructure and physical conditions (e.g., noise, light, pressure). They also suggest considering the history of the context itself as relevant feature.

In 2007, Zimmermann et al. proposed a meta-model for defining context [267]. Specifically, they introduced five categories for expressing context information about an entity: individuality, time, location, activity, and relations. The "individuality" category discerned natural entities from human entities, artificial, and group entities. The activity category encompasses the goals, tasks and actions of an entity. Time describes the history of events. Location covers physical and virtual, quantitative and qualitative (symbolic), as well as hybrid expressions of spatial aspects. Finally, the relation's category describes any possible relation between entities. Social, functional and compositional relations are explicitly mentioned. Dix et al. proposes infrastructure, system context, application domain and finally the actual physical context as factors [55]. The infrastructure context covers the supporting technical infrastructure in which a context-aware mobile device operates (such as the telecommunication network). The system context includes the technical components of the system itself, even if they are distributed in nature. The domain context encompasses the semantics of the application domain (e.g., the situated nature of work that is supported). In addition, the domain context also covers user related information. Finally, the physical context covers the physical properties of the space the system is operated in (such as light, temperature).

Hong et al. proposed a user-centric model with six fundamental context parameters of who, when, where, what, how, and why (5W1H) [121]. "Who" captures basic user information such as name, or gender. "When" encompasses time information (season, time of day). "Where" captures location information (in different granularities, ranging from Cartesian coordinates to geographic regions) "What" includes "relevant objects", specifically applications, services, commands. "How" captures the relevant processes, such as sensor signals or current user activities. Finally, "Why" tries to capture users' intentions and affective state. Hong et al. also propose a categorization into preliminary context (raw sensor measurements) integrated context (inferred information, specifically from sensor fusion), and final context (information processed by the application, trying to encompass higher level reasoning about users' intentions). On a meta level, context can be divided in primary, integrated and final context [121]. Preliminary context considers raw measured data. Integrated context contains accumulated preliminary contexts and inferred information. Final context is the context representation received from and sent to applications. For example, a raw measurement could be provided by a linear accelerometer of a mobile device, which is combined with other sensor measurements of gyroscopes and magnetometers to deliver an integrated rotation measurement. Combined with location data and audio level measurements, the system can infer a "meeting situation" and automatically mute the mobile phone. This three-level categorization follows models about human perception, which assume a multi-layered perception pipeline, e.g., for human vision divided into early, intermediate and high-level vision [108].

Thevenin and Coutaz introduced the sub-term "plasticity". Plasticity "is the capacity of a user interface to withstand variations of both the system physical characteristics and the environment, while preserving its usability" [240]. Hence, it can be seen as a focus of context-awareness on the system level. They identified three dimensions: 1) adaptation source, 2) adaption targets and 3) the temporal dimension of adaptation [240], which can be extended by a fourth dimension, 4) the controller (i.e., the user or the system) [139]. While plasticity concentrates on keeping a system usable in varying usage scenarios, context-aware systems might also offer new services or functionalities depending on the user's situation.

The approaches presented here use general notions of context factors, which allow them to address the problem space of context-awareness on an abstract scale. It is noteworthy that most taxonomies agree on those top-level factors (human factors, technological factors, environmental factors, temporal factors). However, we believe that extending those top-level factors with further sub-categories can ease informing the design of real interactive systems. Specifically, for the domain of *AR*, which by its nature combines attributes of the physical environment and digital information, a comprehensive overview of how context-awareness is addressed and which context factors are relevant for interaction is missing to date. We highlight the fact that by their nature *AR* interfaces are context-aware as they use localization information with six *DOF* to integrate digital information into their physical surrounding. Hence, for this thesis, we concentrate on research that investigated context factors other than spatial location.

2.6 Summary

This chapter provided an overview of related work in the fields of mobile *AR*, mobile interfaces for interacting with physical objects in real-world environments and hybrid user interfaces. Further, we presented an overview of related user studies and introduced how the notion of context-awareness has been understood in previous literature.

Reflecting on the related work, we see that mobile *AR* apps such as *AR* browsers and augmented print apps seem to be a commercial success and that compelling use cases also exist for interacting with electronic displays. Furthermore, alternative mobile user interfaces for information surfaces have been studied. However, both mobile *AR* and alternative user interfaces have mainly been evaluated in performance-oriented studies. Specifically, it remains unclear in which contexts of use mobile *AR* user interfaces are a suitable choice and when alternative mobile user interfaces should be preferred. More specifically, there is lack of scientific evidence that mobile *AR* can benefit consumers in interacting with information surfaces beyond a short term hedonic value (the “wow-effect”). Looking at the various notions of context-awareness, it is also understandable that it will be challenging to design user interfaces that are a suitable choice for all possible contexts. Still, it is worthwhile to further study which context factors could be relevant for mobile *AR* interaction with information surfaces and how mobile *AR* and alternative user interfaces could deliver utilitarian or hedonic value to users in those contexts. The upcoming chapters are dedicated to these investigations.

3

Towards Context-aware Mobile AR

Contents

3.1 Context-aware AR survey 27

3.2 AR browser survey 41

3.3 Information access at event posters 64

3.4 Summary 65

This chapter presents surveys that further motivate the need to consider context factors in the study and design of mobile *AR* user interfaces for information surfaces. The chapter starts with a literature survey on how context-awareness has been considered in *AR* systems. Then, a user survey about the use of *AR* browsers by early adopters is presented. While this survey is concerned with first generation *AR* browsers for location-based experiences, its findings have relevance for the interaction with information surfaces. Finally, a survey on information access at large printed information surfaces, specifically posters, follows. In conjunction with findings from further user evaluations presented in the subsequent chapters, these surveys indicate the need for context-aware mobile *AR* user interfaces for information surfaces.

3.1 Context-aware AR survey

The rise of mobile and wearable devices, increasing availability of geo-referenced and user generated data and high speed networks spurs the need for user interfaces, which provide the right information at the right moment, and at the right place. *AR* is one such user interface metaphor, which allows interweaving digital data into physical spaces and through this aims at providing relevant information on the spot.

AR applications are usually grouped into three components: A tracking component, a rendering component, and an interaction component. All of these components can be considered as essential. The tracking component determines the device or user position in six *DOF*, which is required for visual registration between digital content and the physical surrounding. Based on tracking data the scene (e.g., 3D models and camera images representing the physical world) is composed in the rendering component. Finally, the interaction component allows the user to interact with the physical or digital information when using the fsystem.

Initially, *AR* researchers addressed technical challenges in *AR*, however, in recent years *AR* research switched focus from basic tracking and rendering algorithms to human-centered issues in consumer and industrial contexts. Given the nature and definition of *AR*, location has been handled as major context source for *AR* but there are a multitude of other context factors that have an impact on the interaction with an *AR* system [2, 157]. Generally, context can be seen as being "any information used to characterize the situation of an entity. An entity is a person, place or object that is considered relevant to the interaction between a user and an application, including the user and the application themselves." Similarly, context-awareness is defined as "the facility to establish context" [53].

Over the last years *AR* moved more out of the lab environments into the real world. Also, companies have started to roll out *AR* apps to consumers, which are downloaded by millions of users and used in a multitude of mobile contexts [86, 141]. For example *AR* Browsers, applications browsing digital information that is registered to places or objects using an *AR* view, are used among other purposes for navigation in indoor and outdoor environments (by augmenting routing information), marketing purposes (augmenting interactive 3D media on magazines, posters or products), mobile games (by augmenting interactive virtual characters registered to the physical world) or exploring the environment as part of city guides (e.g., retrieving Wikipedia information that are augmented in the users' view)[142].

As *AR* is increasingly used in real-world environments there is a need to better understand the particularities of *AR* interfaces in different contexts going beyond location. These particularities are often based on the tight spatial link between the interactive system and the physical environment and its implications on visualization and interaction techniques for *AR* applications. This link is also one key factor, which distinguishes *AR* applications from other (potentially context-aware) interfaces for Mobile and Ubiquitous Computing. Hence, it is worthwhile to study the role of context specifically for *AR* and to highlight distinct characteristics that are unique to *AR*. We contribute to this field by providing a) a taxonomy for context-aware *AR* systems, b) a comprehensive overview of how existing *AR* systems adapt to varying contexts and c) by identifying opportunities for future research on context-aware or adaptive *AR* systems. Through this we hope to bring together research from different fields related to this topic (e.g., Pervasive Computing, Human-Computer Interaction, Intelligent User Interfaces, *AR*, Psychology) while also raising awareness for the specific characteristics of context-awareness in *AR*.

3.1.1 A Taxonomy for context-aware AR

Existing taxonomies from the ubiquitous computing domain captured several viewpoints, mostly technology focused, but also address phenomenological aspects. Most of them are coarse (typically only having one to two levels of context factors), leaving the association of finer grained factors to the researchers who apply the taxonomies [218]. For the domain of *AR* our goal was to identify a detailed classification of context sources. This is mainly needed for two reasons. Firstly, because context-aware *AR* approaches often focus on one single specific context aspect instead of integrating a larger group of factors. Thus, a finer granularity makes it easier to discuss existing works on context-aware *AR* and sorting them into the overall taxonomy. Secondly, the finer granularity of the new taxonomy allows us to identify underexplored research areas in particular in the field of context-aware *AR*.

Methodology For creating the classification we used a mixed method approach that combined high level categories of previous taxonomies with bottom up generation of individual categories.

Specifically, we re-used the high level categories of context sources, context targets and context controllers proposed in previous work [152, 240]. Context sources include the context factors to which *AR* systems can adapt. Context targets addresses the question "what is adapted" and corresponds to the "adaptation targets" category previously proposed [240]. This domain describes which part of the *AR* system was the target of the adaptation to external context factors (e.g., the visualization of an *AR* application). Context controller deals with the question "how to adapt?" and corresponds to controller of the adaptation process in previous work [152]. It identifies how the adaptation is implemented: implicitly through the system (adaptivity) or explicitly through user input (adaptability).

Furthermore, for the category context sources we re-used high-level concepts that broadly cover general entities in Human-Computer Interaction [112], which were also employed in taxonomies in the mobile and ubiquitous computing domains (e.g., [218]): human factors, environmental factors and system factors (see Figure 3.1, left).

In addition, we created individual classifications through open and axial coding steps [231]. Specifically, a group of domain experts in *AR* individually identified context factors relevant to *AR*. Those factors were partially but not exclusively based on an initial subset of the surveyed papers. Then those individually identified factors were re-assessed for their relevance to *AR* in group sessions. These group sessions were also used to identify relations between factors and to build clusters of factors that were integrated into the high-level concepts derived from previous work (eventually leading to the presented taxonomy). It became clear that some factors could be seen as part of several parent factors depending on the viewpoint. For example, information clutter can be seen as an environmental factor (a characteristic of the environment) but can also be treated in Human Factors (e.g., attention deficit caused by information clutter). Hence, we want to highlight the fact that while we see the number of context factors as saturated there are other valid hierarchical relations between the factors than the one we present here.

In the following, we will discuss these domains more in detail. In particular, we will discuss factors for which we could identify existing publications while unexplored factors are only briefly mentioned and discussed with more details in the future directions section.

3.1.1.1 Context sources

The high-level categories for context sources human factors, environmental factors and system factors together with their sub-categories are discussed next. They are depicted in Figure 3.1, left and Figure B.2 in Appendix B.

Human factors The human factor domain differentiates between concepts that employ personal factors and social factors as context sources. The difference between both is that personal factors are context sources focusing on an individual user, while social factors concern the interaction between several people (who are not necessarily users of the system).

Personal factors encompass anatomic and physiological states (including impairments and age), perceptual and cognitive [237], as well as affective states. We also separately include attitude (which can be seen as a combination of cognitive, affective and behavioral aspects) and preferences. Another context source that we identified within this sub-category is action/activity (as understood as a bodily movement involving an intention and a goal in action theory). Action/activity addresses both in-situ activity as well as past activities (accumulating to an action history).

Social factors Within the category social factors, we identified two sub-categories: social networks and places. Social networks are understood as a set of people or organizations and their paired relationships [256]. Place can be understood as the semantic of a space (i.e., the meaning which has been associated to a physical location by humans). Social and cultural aspects influence how users perceive a place and one physical location (space) can have many places associated with it. Previous research has shown that place can have a major impact on users behavior with an interactive system in general [2] and with mobile *AR* systems in specific [89].

Environmental factors The domain of environmental factors describes the surrounding of the user and the *AR* system in which interaction takes place, i.e., external physical and technical factors that are not under control of the mobile *AR* system or the user. In order to structure environmental factors, we took the viewpoint of perceptual and cognitive systems. In particular, we rely on the notion of "scene" which describes information that flows from the physical (or digital) environment into our perceptual system(s) in which it is grouped and interpreted.

It is important to note that the sensing and processing of scene information can be modeled on different processing layers of a system ranging from raw measurements, derived measures that rely on a priori knowledge but there is no consensus on which level certain abstractions of information actually take place. For example there are various theories about the process of human visual perception [154, 158], which are specifically popular for computer vision based analysis of the environment but differ in how they are modeled. in this part of this thesis, we differentiate between raw and derived measures (including inferred measures). Raw measures are provisioned by sensors of the mobile *AR* system (e.g., through a light sensor). Derived measures combine several raw measurements (e.g., gyroscope + magnetometer for rotation estimates) and potentially integrate model information to infer a situation.

Within the domain of environmental factors, we distinguish between physical factors, digital factors and infrastructure factors.

Physical factors describe all environmental factors related to the physical world, for instance movements of people around the user. We explicitly differentiate between raw physical factors and derived (combined) physical factors.

Raw factors include factors that can be directly perceived via human senses (such as temperature) or sensor measurement (such as time points or absolute locations in a geographic coordinate system such as WGS84). Derived factors combine several raw factors or derive higher-level factors from certain low level factors (i.e., amount of people in the environment based on recorded environment noise). One example for a derived factor is the spatial or geometric configuration. Spatial or geometric configuration of a scene describes the perceived spatial properties of individual physical artifacts (such as the extend of a poster), the relative position and orientation of physical artifacts to each other and topological properties such as connectivity, continuity and boundary. There are a number of quantitative and qualitative approaches, which try to infer human behavior in urban environments based on spatial properties (e.g., space syntax [114], or proxemics [95]).

Another environmental factor is time. We included time as raw measure, such as a point in time, but also as derived measure (e.g. time interval as the difference between time points). It is important to note that while time may seem trivial on the first sight, it can be a highly complex dimension. Hence, more attributes of time could be of interest. For example, important attributes are temporal primitives and the structure of time [71]. Temporal primitives can be individual time points or time intervals. The time structure can be linear (as we naturally perceive time), circular (e.g., holidays such as Christmas as recurring events) or branching (allowing splitting of sequences and multiple occurrences of events).

The combination of spatial and temporal factors leads to the derived factor of presence (or absence) of physical artifacts in a scene. In particular in mobile contexts, presence is an influential factor due its high dynamic. In mobile contexts it is likely that interaction with a physical artifact is be interrupted and that artifacts become unavailable over time (e.g., an advertisement poster on a bus which stops at a public transportation stop for 60 seconds before moving on). These interruptions happen frequently [238] so *AR* systems should be ready to cope with them.

Other derived factors include motion of scene objects, and interpreted visual properties of a scene. Both factors could for instance be used to decide if a scene object is suitable for being augmented with digital information.

Digital factors. In contrast to physical factors, the second category of environmental factors focuses on the digital environment. Due to the immersive character of *AR* systems, several problems during the usage of *AR* system are directly related to the information presented. The characteristics of digital information such as the quality and the quantity have a direct influence on the *AR* system. Digital information is often dependent on other context sources such as the physical environment. For example, the amount of Wikipedia articles accessible in a current situation can be dependent on the specific location (tourist hotspot or less frequently visited area). However, when it comes to the information presentation as it is achieved through *AR* digital information can in fact be seen as a separate context source that targets the adaptation of user interface elements. Relevant attributes of digital information are type, quality, and quantity of digital information items. As an example the *AR* system could adapt to the quantity of available digital information by adjusting a filter as well as it could adapt to the quality of digital information (e.g., the quality/accuracy of their placement) by adapting their presentation (i.e., similar to adapting the presentation when using inaccurate sensors [96]).

Furthermore, even the results of presentation techniques themselves (e.g., clutter or readability) have been considered as context factors. The latter factors can be seen as integrated context factors [121], which only occur due to the interaction between preliminary factors (quality of information, perceptual abilities of the user) and a processing system. It should also be noted that this processed information category is naturally connected to other categories such as the perceptual and cognitive capabilities of a user or the technical characteristics of the display (e.g., resolution or contrast of a *HMD*).

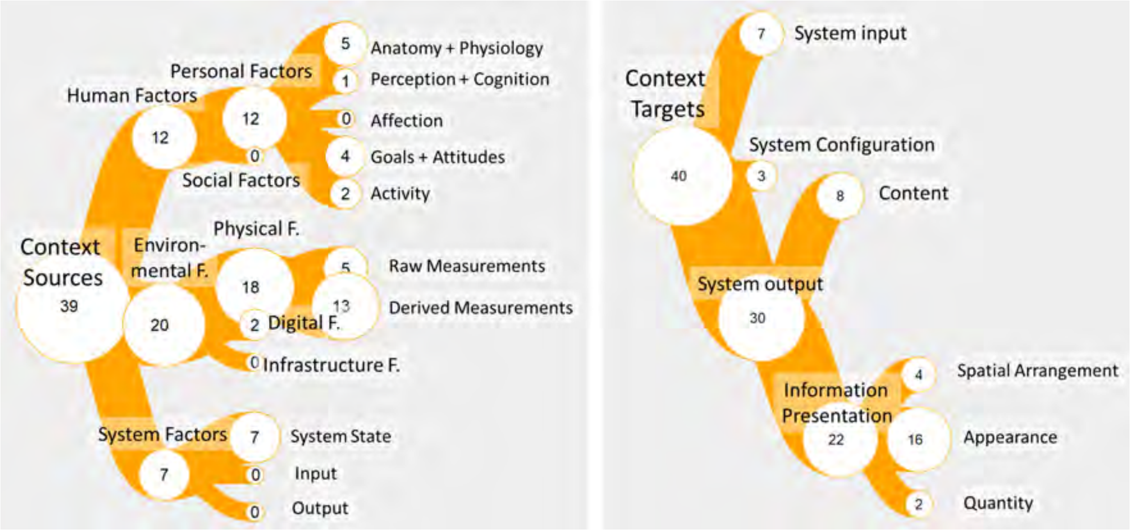


Figure 3.1: Context sources (left) and targets (right) relevant for *AR* interaction. The numbers in the circles indicate the amount of papers in the associated category, papers can be present in multiple categories.

Infrastructure factors. Nowadays, many *AR* applications are mobile applications that can work in different environments. This is enabled by an infrastructure that is used by the *AR* application. In distributed systems in which *AR* interfaces are often employed it might be hard to draw the line between the interactive system itself and the wider technical infrastructure. At a minimum we consider the general network infrastructure, specifically wide area network communication, as part of the technical infrastructure. For practical *AR* applications the reliability and bandwidth of a network connection are of high importance as digital assets are often retrieved over the network.

System factors Technical sources of context can concern the interactive system. As mentioned earlier we leave out infrastructure components that are used by the interactive system but not necessarily part of the system (e.g. networks infrastructure). The main system factors an *AR* system can be aware of is the general system and its capabilities such as the device the *AR* application is running on, the current state of the *AR* system, factors evolving around, the general output component of an *AR* system, and the general input component of an *AR* system.

System state. One system factor is the interactive system itself. For instance, computational characteristics such as the platform, computational power or battery consumption can be used for adaptation as these are strongly connected to the system. In particular for *AR*, both sensors (such as cameras, inertial measurement units or global positioning system sensors) and their characteristics (*DOF*, range, accuracy, update rate, reliability) contribute to the system state.

Output factors describe the different varieties of presenting information to the user. Typically, systems adapt to visual output devices, such as different display types, varying resolutions, sizes or even spatial layout for multi-display environments. But output factors also include other modalities such as audio or tactile output.

Input factors. In contrast to output factors, input factors describe different possibilities how users can give input to the *AR* system. Typically, input is done via touch gestures, but it also includes gestures in general, mouse input or speech. Depending on which kinds of input modalities are available, the system could adapt its operation.

3.1.1.2 Context targets

Based on the analysis of context sources the system applies changes to targets, which are parts of the interactive system [240]. Major categories that can be adapted in an *AR* system (and most other interactive systems) are the system input, output and the configuration of the system itself (see Figure 3.1, right and Figure B.1 in Appendix B).

For system input, the interaction modalities can be adapted. For example, the input modality of an *AR* system could be changed from speech input to gesture-based dependent on the ambient noise level but also based on user profiles or environments (e.g., public vs. private space).

Other approaches that adapt the input could optimize the position and appearance or type of interactive input elements (e.g., increasing the size of soft buttons based on the environment, optimizing the position of user interface elements or the intensity of haptic feedback based on the information from the physical environment).

For *AR*, the adaptation of information presentation is an important subgroup. Given that *AR* has an emphasis in visual augmentation a main target for adaption is an adapted graphical representation. Here, spatial arrangement of *AR* content (e.g., label placement [77]), appearance changes (e.g., transparency levels [128]) or filtering of the content amount (e.g., removing labels [125] or adjusting the level-of-detail [54, 225]) have been studied. An example for adapting a complete user interface (input and output), would be an *AR* route navigation system, which operates by overlaying arrows on the video background at decision points. If the tracking quality degrades the depicted arrows visualization can be adapted [181]. In addition, an alternative user interface could be activated (e.g., an activity-based guidance system [169]).

3.1.1.3 Controller

As third major aspect of context-aware *AR* systems we investigated how context targets are adapted based on input from context sources. As in other context-aware systems the adaption can be conducted implicitly through the system (adaptivity) or explicitly through user input (adaptability). Implicit adaptation mechanisms automatically analyze context sources and adapt contexts targets accordingly based on model knowledge and rule sets. For example, a popular model for the analysis of scene content in *AR* is the saliency-based visual attention model by Itti et al. [123].

3.1.2 Survey on existing approaches for context-aware AR

In the following section we will discuss existing works in the field of context-aware *AR* following the taxonomy we created earlier in this part of this thesis. We categorize the existing works based on identified context sources while giving further information on the context targets and controller aspects in the text.

3.1.2.1 Human factors

The first category of context sources that we will use to discuss existing research concerns factors that are directly related to the user's state. While we identified two sub-domains within the domain of human factors, namely personal and social factors, only personal factors have been considered in previous research. We classified existing research in this field into three subcategories: anatomic and physiological states, perceptual and cognitive states, and activity.

Anatomic and physiological states Several groups investigated how to adapt an *AR* system to the user's state, which is approximated through biophysical readings or through user profiles. While potentially useful for various application domains in particular medical *AR* applications used for rehabilitation were investigated. As such Dünser et al. presented an *AR* system for treating arachnophobia (fear of spiders) by using virtual spiders overlaid in the patient's proximity [63]. Based on physiological sensor readings such as heart rate but also by tracking and analyzing the patient's gestures, the system adapts the graphical representation and animation of the virtual spider, which both affect the exposure of the patients fears. Unfortunately, parts of the presented work were in a conceptual state and details and how to track and analyze the patient's gestures were not provided. Lewandowski et al. [148] focused on a mobile system for evaluating and aggregating sensor readings. They presented the design of a portable vital signs monitoring framework. The system "'aggregates and analyses data before sending it to the virtual world's controlling device as game play parameters"' [148]. Sinclair and Martinez created a museum guide that adapts to age categories (adults or children) [222]. Based on the type of user the system reduces (children) or increases (adults) the amount of displayed information. The system uses the assumption that adults prefer more details while children need less information. Xu et al. used bio-sensor readings (e.g., pulse measurements) as part of an integrated attention model for *AR* applications in the cultural heritage domain [262]. They adapted the visual presentation of artwork information based on an integrated "interest model".

Perceptual and cognitive states Besides bio-sensor readings Xu et al. also employed visual attention measures through an eye tracker to infer visual attention [262]. Specifically, they employed eye fixations as one parameter in their interest model. In addition, the authors used audio sensors to identify if the user was talking to a nearby person or concentrating on the artwork and to identify crowded locations.

Attitude and preferences Hodhod et al. presented an *AR* serious game for facilitating problem solving skills [117]. The authors adapt the gameplay based on a student model that holds information about a student's learning style and ability level. Similar, Doswell presented a general architecture for educational *AR* applications that takes into account the user specific pedagogical models [56]. These pedagogical models influence the information and explanations that are displayed to the user.

Activity Stricker and Bleser presented the idea of gradually building knowledge about situations and intentions of the user using an *AR* system to adapt the system based on these context sources [232]. As a first step, they propose to determine body posture and to analyze the users' environment. Both together are used as input to machine learning algorithms to derive knowledge about the situation and intentions of the user. Stricker and Bleser propose to use the users' activity to create an unobtrusive and adapted information presentation that fits to the users' actual needs. However, their work entirely focuses on tracking of posture and environment together with the machine learning while the adaption is only conceptually presented.

3.1.2.2 Environmental factors

While *AR* applications are in general dependent on their current position and consequently their environment, some works go beyond that by actively analyzing the environment to adapt the system. Analysing the environment and adapting the system based on the gained information can be utilized in various ways for Augmented Reality. One can think of *AR* applications that analyze the shape or structure of the environment to use it for example to optimize the position of augmentations. As described in the taxonomy, we identified three subdomains in the category of environmental factors - physical factors, digital factors and infrastructure factors. Adapting the *AR* system based on physical structures includes the above mentioned example of analysing the shape of the physical environment but also noise or other characteristics of the environment that can be sensed, measured or derived from these measured environment factors. Digital factors are context sources that relate to the digital environment for example the amount of digital information in the environment. The last subdomain in environmental factors are infrastructure factors where we considered the technical infrastructure installed in the environment and used by the *AR* system as an important context source. This could include the availability of wide area networks but also other technical infrastructure elements that are part of the environment and not of the system. While there is a large amount of previous research investigating how to use physical factors to adapt an *AR* system, there are only few works on how to use digital factors and none that uses the infrastructure as context source.

Physical factors Barakonyi et al. presented a framework that uses animated agents to augment the user's view [13]. The *AR* agents make autonomous decisions based on processing ambient environment measures such as light or sound.

Henderson and Feiner [109] presented the idea of Opportunistic Controls. In their work they adapt the interaction implemented in a tangible interface based on the appearance of the environment. The system utilizes existing physical objects in the users' immediate environment as input elements. Xu et al. employed measurements of environmental noise to adapt their user interface and displayed content in an *AR* museum guide [262]. If a certain threshold is reached the tour route is changed the user is guided away from the noisy location. Grubert et al. proposed to employ hybrid user interfaces, a combination of *AR* and alternative user interfaces, for interacting with a printed poster in mobile contexts [83]. A key observation of their research was that users might not always prefer an *AR* interface for interacting with a printed poster [88, 89] in gaming contexts or even benefit from it in touristic map applications [88]. Hence, the authors propose to allow users to explicitly switch between *AR* and alternative user interfaces. They also discussed the possibility to detect when a user moves away from a poster (through analyzing tracking data) and subsequently automatically switching between *AR* and alternative interface (such as a zoomable view) [83].

In particular, in video-based *AR* it is popular to use video images not only for overlaying and tracking but also for computing visual aspects about the physical environment of the user. We map these methods to the dimension of physical environment context factors within the sub-domain of derived visual measurements. These methods often address either the problem of information clutter or readability and use information presentation as context targets, such as spatial arrangement or appearance of user interface elements. For instance, Rosten et al. introduced an approach that spatially rearranges labels in order to avoid that these labels interfere with other objects of interest [204]. For this purpose, their method extracts features from camera images and computes regions appearing homogeneous (not textured) to allow for integration of new digital content in these regions. Similarly, Bordes et al. introduced an *AR*-based driver assistance system which analyses road characteristic and position of road markings as context source for adapting visual representation of navigation hints [29]. They focused on readability of overlaid information in particular when using reflective screens for creating the *AR* experience (in their example the windscreen of a car). A related approach was used by Tanaka et al. for calculating the most suitable layout for presenting digital information on an *OST HMD* [239]. In their approach, feature quantities for different display regions based on RGB colour, saturation and luminance were calculated. Another related method has been proposed by Grasset et al. [77] and focuses on finding the optimal layout of labels for *AR* browsers. This method again uses information clutter as context source. Information clutter is measured not only using edges [204] but using salient regions in general for determining regions that contain less important information [1].

Another problem that is caused by the composition of digital and physical information is reduced readability. While readability also depends on human factors we consider them as constant during the time of interaction. Hence, the properties of the physical scene have a major impact on the readability. Methods that address this problem often use readability measures as context source and adapt the information presentation as context target. For instance, Gabbard et al. suggest to analyze the readability of labels and to adjust their presentation, such as font colors [72]. For this purpose, they performed a user study that investigated the effect of different background textures, illumination properties and different text drawing styles to analyze user performance in a text identification task. While this work does not present a fully adaptive solution to the readability problem, the results delivered important findings about readability as a context source. In particular, in outdoor environment text readability is a big problem as those environments are less restricted than controlled indoor environments. In order to address this problem, Kalkofen et al. proposed to use various measures of the physical and digital environment e.g., acquired through image-based features properties of environmental 3D models, to adjust the visual parameters or material properties in an *AR* outdoor application [129]. Later, this idea of using different context sources for adjusting the information presentation in *AR* was extended by Kalkofen et al. for the concept of X-Ray *AR* [128]. X-Ray *AR* allows for instance to reveal occluded subsurface objects for subsurface visualization [268]. One main challenge for this kind of visualization is the preservation of important depth cues that are often lost in the process of compositing digital and physical information. Kalkofen et al. addressed this problem with an adaptive approach that uses different context sources in order to adjust the composition between both information sources [128].

Another important physical context factor in *AR* environments is scene illumination, since it may be subject to fast changes in particular in outdoor environments. In order to address this problem, Ghouaiel et al. developed an *AR* application that adapts the scene brightness of the virtual scene according to measures of the illumination of the physical environment (as measured through an ambient light sensor on a smartphone) [75]. Furthermore their system adapts to the distance to a target object and to ambient noise [75]. Dependent on the Euclidean distance to a target object (e.g., a house) the authors adapted the size of the target (e.g., a label), proportionally. Finally, the authors also propose to adjust the level of virtual sound based on the ambient noise level.

Similar, Uratani et al. propose to adjust the presentation of labels based on their distance to the user [246]. In this case the distance of labels in the scene is used as context source to change the appearance of labels. The frame of a label was used to color-code depth while the style of the frame was adapted according to their distance. DiVerdi et al. have investigated a similar concept; they use the distance of the user to objects in the physical world as input to adapt the level-of-detail of presented information [54]. Recently, this research has been extended to the usage of additional spatial relationships in the work of Speiginer and MacIntyre [225].

Digital factors In contrast to physical context factors, digital factors use input from the digital environment as context sources and adapt the *AR* system based on this input. The techniques can be used to overcome the problem of information clutter. For instance, one can adapt the system to the quantity of digital information that is present in an environment (e.g., the number of *POIs* at a specific geolocation). Based on the amount of information, these methods reduce the number of presented information items (such as labels or pictures) or rearrange the presented information to avoid an overload of information. An example for reducing the amount of information has been presented by Julier et al. [125]. Their method uses the amount of digital information both as context source and context target. The method divides the image into focus and nimbus regions. They then analyze the number of objects in the 3D scenegraph representing the digital scene for those individual regions. Based on this analysis they remove 3D objects in the scenegraph for cluttered regions. Mendez and Schmalstieg propose to use context markup (textual description) for scenegraph elements which in turn can be used to automatically apply context-sensitive magic lenses using style maps [161].

3.1.2.3 System factors

Within this section we describe *AR* systems that use system factors as context sources and adapt either to the system state (i.e., computational resources, such as computational power or sensors integrated into the system and their characteristics), the system's output factors (e.g., visual output devices, spatial arrangement of displays or other modalities) or the system's input factors (e.g., availability of input modalities).

There are several works that investigate adaption to the tracking system or use positional error estimates of the tracking system to adapt visual output. A common idea of many existing works that are sensitive in terms of tracking quality is to adapt the graphical user interface based on the error in the position estimate. For example, Hallaway et al. presented an *AR* system for indoor and outdoor environments that uses several tracking systems offering different levels of confidence in the position estimate [96]. In indoor environments, a ceiling-mounted ultrasonic tracking system offering high precision is used. This allowed the overlay of precisely placed labels or wireframe models. However, when the users leave the area covered by this tracker the system makes use of trackers with less accuracy, such as pedometers (in combination with knowledge of the environment) or infrared tracker. In outdoor environments the proposed system makes use of a GPS sensor with inertial sensors for tracking the position. In all these cases the error of the position estimate is larger than the one from the ultrasonic tracker making it impossible to precisely overlay digital information. The proposed system consequently adapts the graphical interface by transition into a world in miniature visualization where the WIM is roughly aligned with the users' position coming from the less accurate trackers employed.Similarly, MacIntyre et al. [155] analyse the statistical error of a tracking system and apply the result using the graphical representation of digital overlays as context target. The developed *AR* system was used to highlight objects and buildings in the environment (e.g., for navigation). The idea to overcome wrongly placed overlays resulting from the tracking error is to grow the convex hull of the digital overlay based on an estimate of the registration error. This guarantees that the digital overlay is still covering the physical object by displaying this modified convex hull and applying other visualization techniques. The results of these work also influenced work by Coelho et al. when they presented similar visualization techniques but this time already integrated into a standard scenegraph implementation [47].

A general approach for using the system state as context source was presented by MacWilliams working on ubiquitous tracking for *AR* [157]. He presented a tracking architecture that adapts the general configuration consisting of several simultaneous running trackers with various update rates and with different precisions. The proposed architecture consequently had to support system analysis at runtime. The system " [...] builds on existing architectural approaches, such as loosely coupled services, service discovery, reflection, and data flow architectures, but additionally considers the interdependencies between distributed services and uses them for system adaptation"[157]. The context target is the graph that is used to connect the different trackers and represents the system configuration.

Verbelen et. al presented a different work for adapting the overall system configuration with the aim to optimize the performance of an mobile *AR* system [250]. Contrary to the work of MacWilliams, they focused on mobile *AR* applications where parts of the computation can be offloaded to a remote server. The overall configuration and computation of the system is adapted to the current workload of the mobile CPU, to the network quality, and the availability of remote servers that can be used to offload certain computations. Depending on the context the *AR* application can offload parts of the tracking computation to a server that sends back the results. Similarly, they also presented how to gracefully degrade the tracking quality when the network connection is lost to meet the capabilities of the local processing power on the device. This process is hidden from the user but aims to improve the overall experience by giving the best performance in terms of tracking quality and speed. While not explicitly mentioning context-awareness Pankratz et al. also dealt with tracking uncertainty as context source [181]. They investigated a number of visualization concepts which apply to route navigation systems. They indicated that error visualizations have the potential to improve *AR* navigation systems but also that it is difficult to find suitable visualizations that are correctly understood by users.

3.1.2.4 Others

There are also works that deal with context-awareness but do so on a general level (e.g., merely claiming that context-awareness is important for *AR*). For example, Shin et al. [220] presented a conceptual work that adapts the content and the general representation with respect to the users' profile and the user history. Caused by the conceptual character of the work no details are provided how to compute the profile and how it is exactly used as context source for adapting the system.

3.1.3 Discussion

Based on the created taxonomy and the reviewed literature, in this section we discuss the current state of context-aware *AR* and opportunities for future research.

3.1.3.1 Summary of existing approaches

While there are potentially many relevant context sources for an *AR* system, research so far has concentrated on selected topics. Specifically, anatomic and physiological factors have been considered [63, 148], visual perceptual factors [262] as well as user preferences or pre-defined proprietary user profiles [117]. Few works have concentrated on the user's activity [232], activity history, attention or affective state. Similar, social factors (such as place or social networks) did not play a major role in existing works on context-aware *AR*.

Regarding environmental factors research has concentrated on both raw and derived visual measurements (such as saliency) of a scene [128, 204]. These works usually aimed at improving the composition of the physical world and digital information so that it is easier to understand. Some works have explicitly considered the spatial configuration of a scene [109] and others the presence or absence of augmented artifacts [83]. Only very few works have concentrated on digital context factors [125]. For system factors the majority of works have concentrated on characteristics of tracking sensors [155, 181] and few on user input and output factors. For context targets most work concentrated on the adaptation of information presentation [77, 125, 204]. Regarding context controllers all presented work used implicit adaptation techniques and only few systems relied on adaptability through explicit user input. To summarize, context-aware *AR* has been investigated only in isolated islands of topics. While there are a number of conceptual works and system papers (where the state of the implementation appears unclear), user studies on the effects of context-aware systems on the user experience of *AR* are rare. Interestingly, despite the fact the tracking is deemed important in the *AR* community adaptive tracking research has only scratched the surface, too.

Despite these isolated islands of investigated topics we argue that several of the demonstrated context sources in context-aware *AR* are specific to *AR* interfaces and are caused by the tight spatial link between the interactive system and the physical environment. Specifically, environmental factors and here in particular the physical factors play an important role for context-awareness. The fact the most *AR* systems use a camera for visual tracking or depth sensors for sensing the environment is further exploited to support context-awareness by capturing additional information about the environment. This specific hardware is often not part of other interactive applications and these specific context factors are consequently less explored in works outside *AR*. Similarly, the larger amount of works using system factors as context sources, in particular the state of the tracking system, is unique to *AR*. Precise tracking is essential for *AR* applications and usually a combination of many tracking sensors is employed for achieving this high precision, making the tracking system an important factor for adapting the system. While other interfaces also used tracking information such as location as context source, the used tracking data has usually less dimensions (e.g., two *DOF* instead of six *DOF* such as in *AR*), less accuracy (e.g., meters instead of millimeters), and results from fewer sensors (e.g., GPS only instead of hybrid tracking using cameras and hardware sensors).

3.1.3.2 Opportunities for future research

Based on the presented taxonomy and existing works, we can identify research gaps and promising research directions. Our taxonomy and surveyed papers show that the context source space is only partially addressed. For example, while visual perceptual issues are addressed by several works for personal human factors the affective state of the user plays no major role in *AR* system adaptation – even though there is a whole research field on affective user interfaces [183] which is relevant for *AR* interfaces, too. Similar, so far the *AR* community has missed to investigate the potentials of using social network services to get more information about the social context in which users interact with an *AR* system. For example, one potential context source could be the crowdedness of a scene, which could be measured either through live video analysis (e.g., using a people detector) or a priori knowledge using social network services (e.g., analyzing the number of tweets about public events in a region). This information could be used to adapt the input capabilities of handheld *AR* systems, e.g., by offering users a more discrete user interface which does not require visible spatial gestures (holding up the handheld device in front of the user while walking through a crowd).Also, no work so far has concentrated on varying infrastructure factors (e.g., the availability of situated displays in public space). Similar, the availability (or lack of) multiple concurrent input and output devices for *AR* interaction has not been investigated. Hence, we see a potential to investigate *AR* interaction beyond a single input and output device such as an individual smartphone.

We also see large potential and even the need for investigating physical factors as context sources in *AR* systems. Examples are adapting to temporal factors (e.g., adapting the visualization based on the brightness of the physical world similar to dark and bright desktop themes). There seems also to be a large potential for mobile *AR* systems to better adapt to the motion of the user or the environment. For example, user interface elements could be adapted to the motion of a user (e.g., label size as the user walks faster). For system factors existing research has largely concentrated on tracking sensor characteristics, neglecting other important system characteristics of mobile devices. One could imagine a mobile *SLAM*, which balances the workload of mapping between a server and the handheld client, based on the computational resources and battery state of the client.

All these future research directions become even more important when head-mounted displays (such as Google Glass) enter the public market. The fact that these devices can be permanently worn and can be used in different contexts while offering only limited controls for manually adapting the interface to the current context requires automatically adapting to the current context. Furthermore, we identified that most of the existing works focus on integrating context source into their system but do not demonstrate which context target is adapted. This indicates that many developed systems seem to be not complete. Looking at application domains for context-aware *AR* a promising area is phobia treatment and simulating psychological effects [63]. Context-aware systems would allow for building a closed-loop approach that adapts to the users' state (similarly as conceptually proposed by Dünser et al. [63]) continuously.

3.2 AR browser survey

Mobile *AR* browsers have become one of the major commercial *AR* applications. Still, real-world usage behaviour with this technology is still a widely unexplored area. We report on our findings from an online survey that we conducted on the topic and an analysis of mobile distribution platforms for popular first generation *AR* browsers. We found that while the usage of *AR* browsers is often driven by their novelty factor, a substantial amount of long term users exists. The analysis of quantitative and qualitative data showed that poor and sparse content, the user interface design or the system performances are major elements influencing the permanent usage of this technology by early adopters.

An *AR* browser is a generic augmented reality application proposing to display geo-located multi-media content using a virtual representation augmented on the vision of the real world (i.e., a camera-image in the context of smartphone technology). *AR* browsers generally access remote resources through web protocols and services (e.g. HTTP Methods, REST), index the content through media streams (termed *channels*, *layers* or *worlds*) and support a variety of MIME formats (html, image, audio, video or 3D model).

AR browsers are not per se new; earlier work such as presented by Feiner et al.[67], Höllerer et al. [118] or Kooper et al. [136] were already introducing the concept of multimedia browser in the real world, either in term of their specific user interface or their system architecture. Differently, the recent progress of pervasive technology (wireless and cellular network infrastructure, web software technology, powerful mobile devices) deliver now a simple way to access and use an *AR* browser on a mobile device, outdoor as well as indoors.

As the awareness of this technology is spreading rapidly in the mind of the public (but also on their own device), the usability and responsiveness of *AR* browsers has never been thoroughly analysed. Precisely, former studies have been generally limited to the testing of some of their components and features (previously developed by academic research), in the context of lab-controlled human factor studies.

In this part of this thesis we describe a survey we conducted in July 2011 as a first step to gather more knowledge about the potential and interest of *AR* technology from the public. Complimentary, we also looked at the evolution and adoption of the technology that can be quantified from mobile distribution platforms, such as Android Market or Apple App Store, where *AR* browser applications can be access, rated or commented. Both of these tools offer us a wider vision on the user behaviour related to *AR* browsers.

After briefly summarizing previous work on this topic, we introduce the experimental design and results of our survey on *AR* browsers. Finally we describe our analysis of adoption and subjective comments of some of the *AR* browsers available in popular mobile distribution platforms before concluding.

3.2.1 Online survey

In this section, we present the experimental design and result of an online survey we conducted from May to July 2011. We will use the term ARB to refer to *AR* browser.

3.2.1.1 Method

We used an online survey to collect data from early adopter of the ARB. It was advertised on several social media channels and via e-mail.

Participants We recruited participants through social network sites (Facebook, Linkedin, Twitter, discussion boards), mailing lists and postings on communication channels of ARB vendors. In total 77 participants (14 female) fully completed the survey, 118 partially answered questions. We report only the results from the completed responses. Most participants were aged between 20 an 40 years (Figure 3.2a).

Material The data was collected with LimeSurvey[1]. Statistical tests were conducted with R[2]. Coding of qualitative data was done in Nvivo 9[3] and Microsoft Excel.

[1] http://www.limesurvey.org, last retrieved 20.04.2015.

[2] http://www.r-project.org

[3] http://www.qsrinternational.com/, last retrieved 20.04.2015.

Procedure Participants were informed about the purpose of the study and the approximate time needed to complete the survey. They were informed that the data was collected completely anonymously; no incentives for taking part in the survey were offered. Participants were asked to answer 28 questions separated in three question groups (namely user background, type and applications, and benefits and drawbacks). The complete questionnaire can be found in appendix A.

3.2.1.2 Results

We present results on selected sections of the survey including participants' backgrounds, usage behaviour, usage scenario, consumed media, feature quality, movement patterns, social aspects and reasons for discontinuing using ARB.

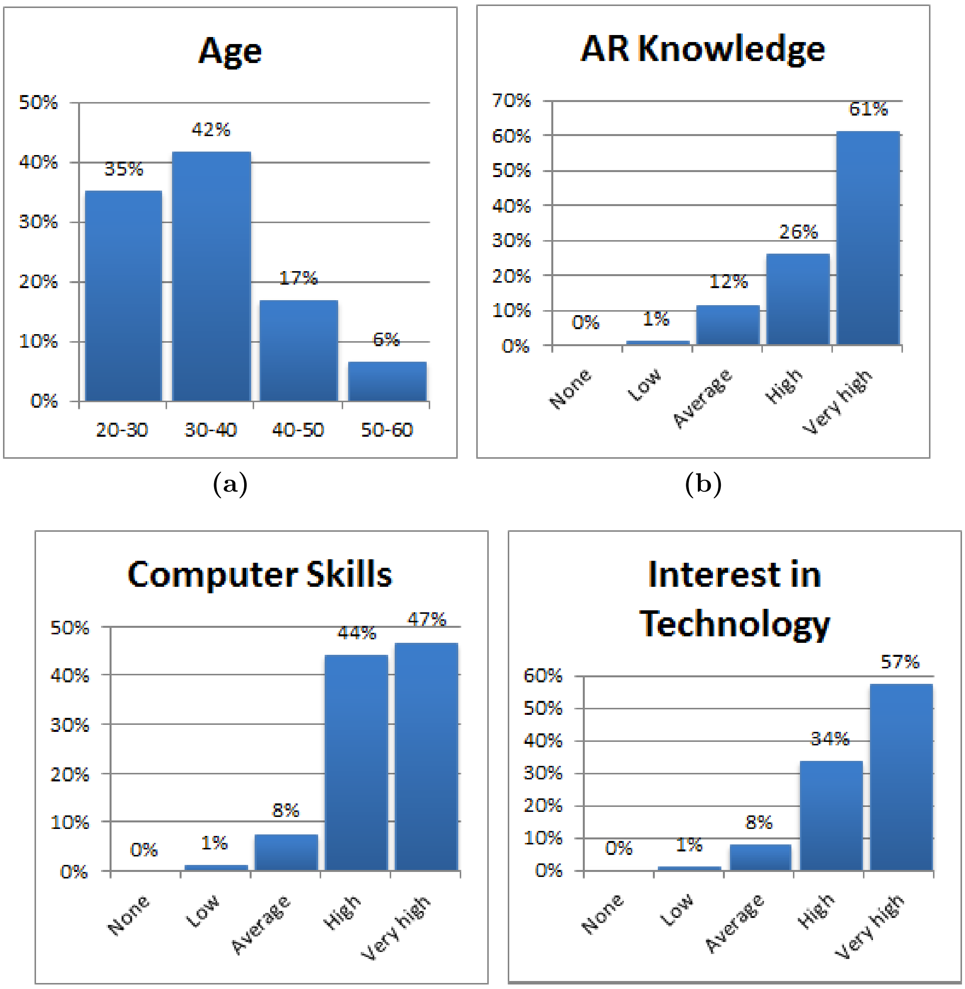
Demographics The recruitment channels of the survey resulted in participants who can be seen as tech-savy people and early adopters of ARB. This is reflected in the demographics that show a high computer literacy and interest in technology of most participants (see Figure 3.2). The participants were allowed to describe their professional status with an open form item. We clustered them in the categories presented in Figure 3.3a.

Application background While there are more than twenty ARB applications out there, three of them were noted as the most popular amongst the participants: Layar , Junaio and Wikitude (see Figure 3.3b). The browsers were mainly used on iOS (54%) and Android devices (42%) with only a few using other platforms. Participants did first hear about ARB mainly through websites an blogs (66%), followed by exploring the distribution platforms (Apple App Store, Google Android Market) (38%) and recommendations by friends (36%) (multiple choices were possible).

Mobile services that were used at least on a daily basis by the participants are Email (83%), Internet Browsing (79%), Social Network Services (71%) and calling (71%) (see Figure 3.4). Games were used on a less than daily basis by 61% (22% used them daily). Navigation applications like Google Maps were used by 58% less than daily and by 41% at least daily. Multimedia content was consumed by 48% daily and by 46% less than daily. These numbers reflect that the majority of the participants employed their phones primarily as communication medium and for general purpose browsing.

Usage time The average session time with an ARB was between 1-5 minutes (see Figure 3.5c). Roughly a third of the participants (34%) tried out the browsers only a few times. On the other hand 42% used the browsers at least on a weekly basis (see Figure 3.5a). The period of active usage was also split into two groups with a third of the participants (33%) using the browsers only for a few days and a third (32%) using them for at least half a year (see Figure 3.5a). In the remainder of this report we therefore also looked for group differences between these high frequency and low frequency users as well as between these long-term and short-term users.

Usage frequency and usage duration have a strong positive correlation (Kendall's $\tau(75) = .55, p < .001$), see Figure 3.6.



(a)

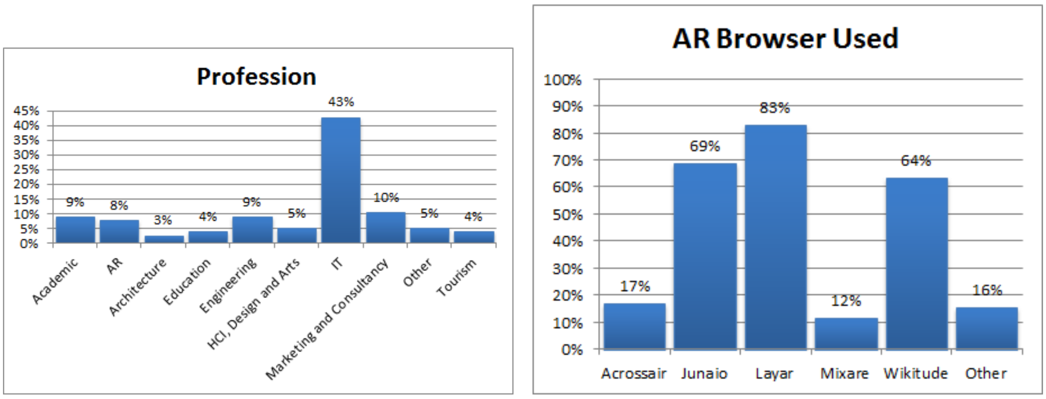
(b)

(c)

(d)

Figure 3.2: Overview of participant's age (a), knowledge of Augmented Reality technology (b), computer skills (c), and interest in technology (d).

As the gathered data was ordinal and failed normality tests (Shapiro-Wilk) we employed non-parametric hypothesis tests (Mann-Whitney U) for testing group differences. A Mann-Whitney U test indicated that professional *AR* users (*AR* knowledge: very high, $n = 47, 61\%$) used *AR* browsers significantly more frequently (Mdn="few times a week") than novel users (*AR* knowledge low to high, $n = 30, 39\%$) (Mdn="5-6 times", "every two months"), $U = 924.5, p = .01$. This test also indicated that professional *AR* users use ARB significantly longer (Mdn="3-6Months") than novel users (Mdn="1-3 Months"), U = 924.5, p = .01 (see also Figure 3.7).



(a) **(b)**

Figure 3.3: Participants' professional status (a) and *AR* browsers used by participants (b).

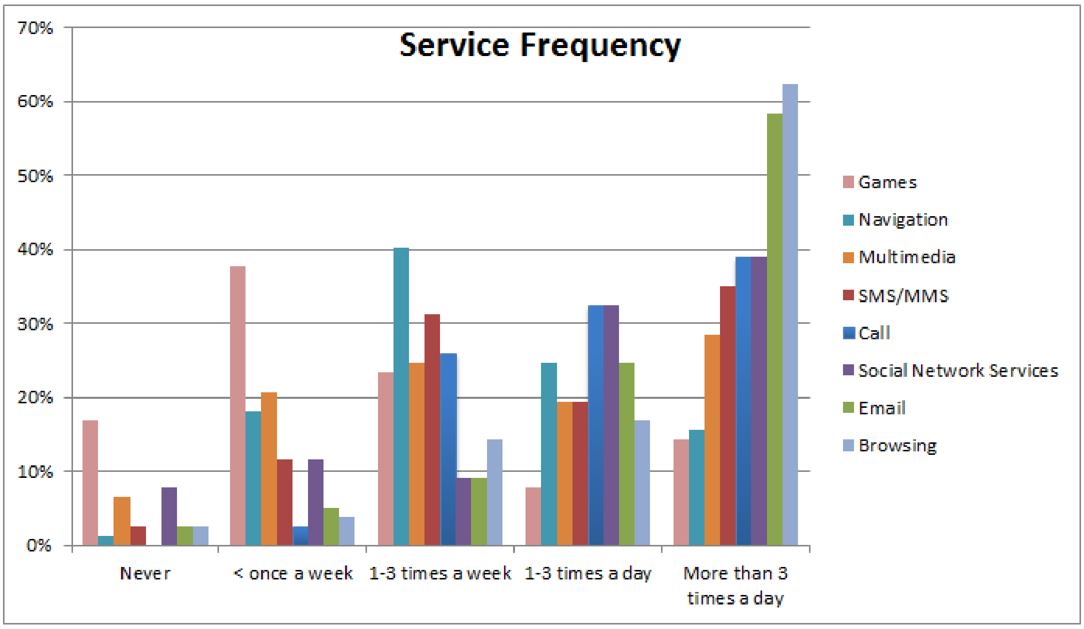
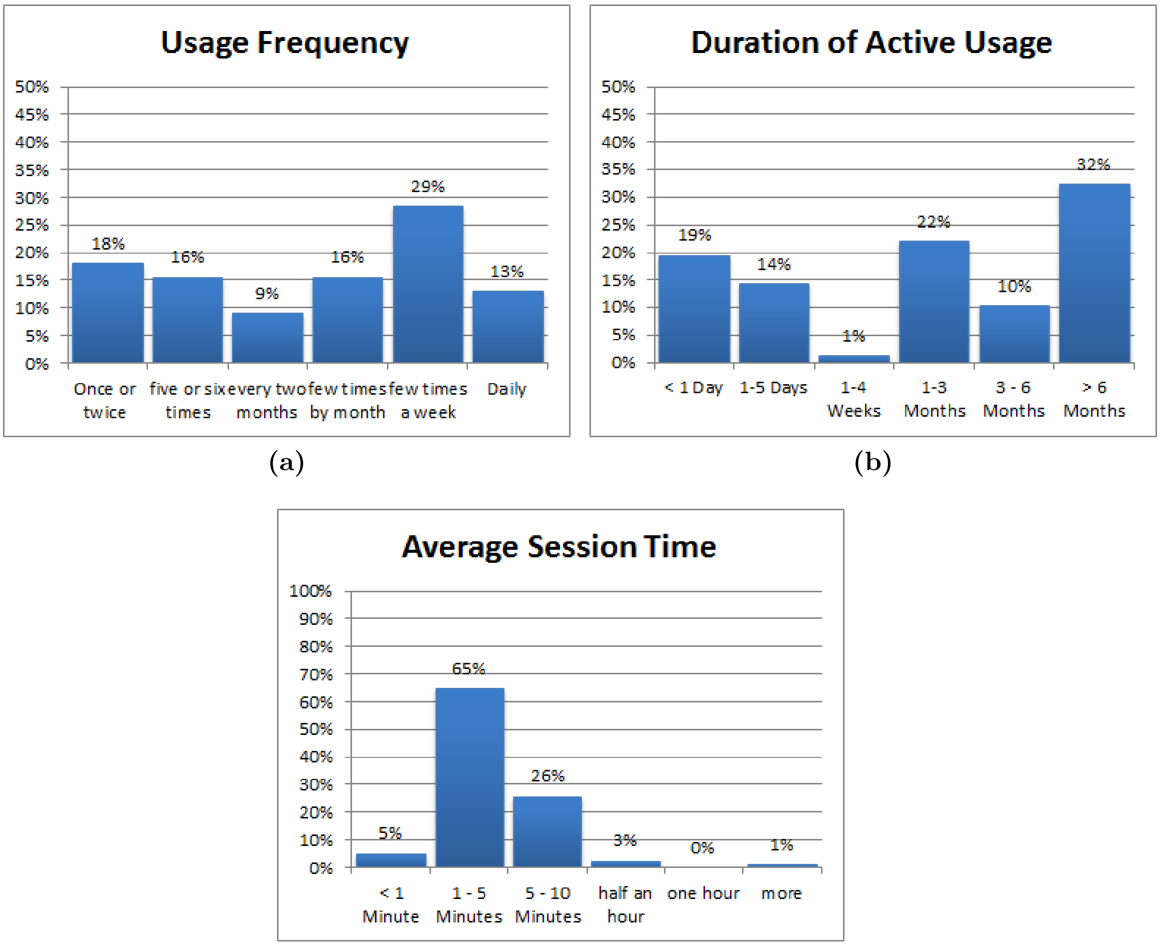


Figure 3.4: Frequency usage of Mobile Services.

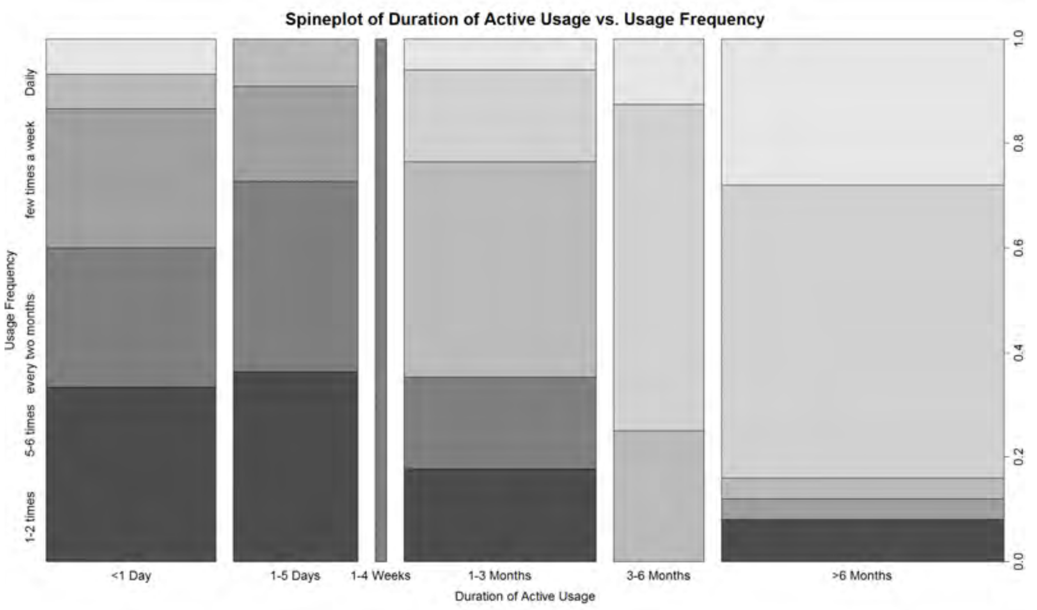


(a)

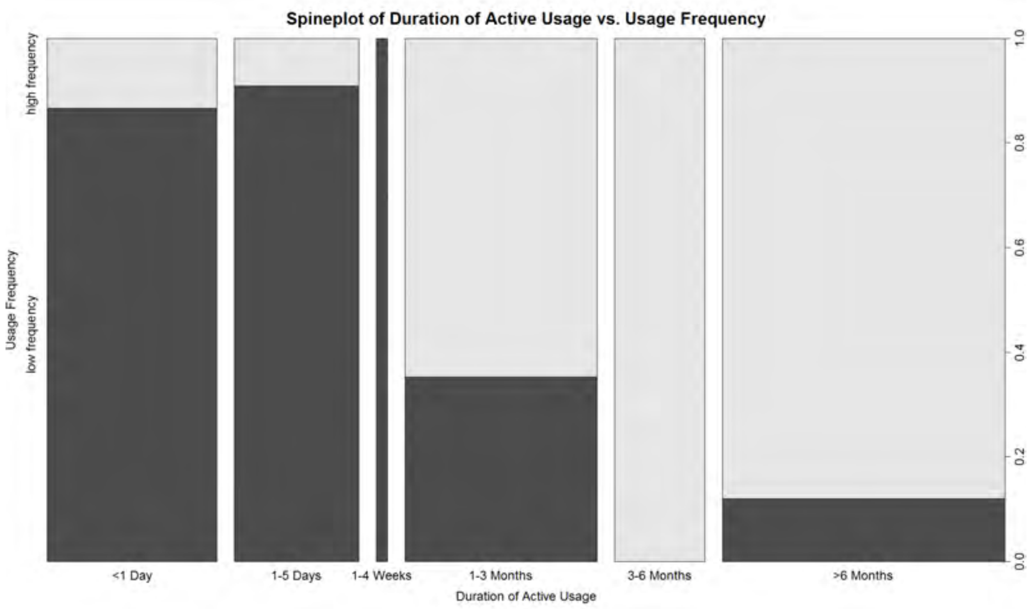
(b)

(c)

Figure 3.5: Usage frequency (a), duration of active usage (b), and average session time (c).

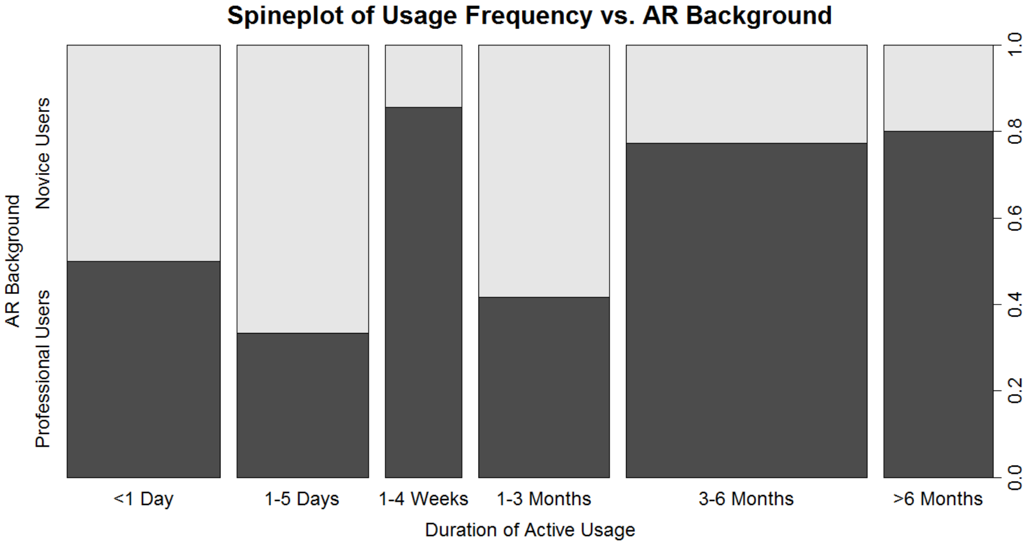


(a) Usage frequency and duration of active usage.

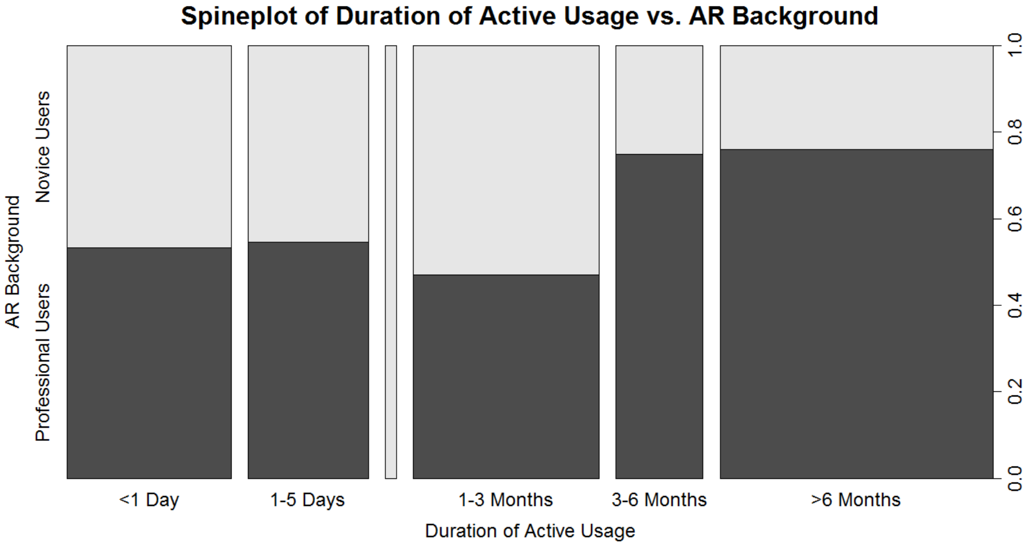


(b) Usage frequency collapsed into high and low frequency users and duration of active usage.

Figure 3.6: Usage frequency and duration of active usage with original (a) and collapsed frequency (b) categories.



(a) *AR* background and usage frequency.



(b) *AR* background and duration of active usage.

Figure 3.7: Spineplots for users with high and low *AR* background w.r.t. usage frequency (a) and active usage duration (b).

Usage scenarios Participants of our survey used the *AR* browsers most often for general purpose browsing and navigation (see Figure 3.8). 31% of the respondents also used the browsers for gaming, 39% in museum settings. The browsers were used outdoors by most (91%) and indoors by half (51%) of the participants. A third of the participants (27%) already used the browsers in a social group, 44% with a few friends, and 57% alone (multiple choices possible). There were no significant effects with respect to age, gender or *AR* expertise.

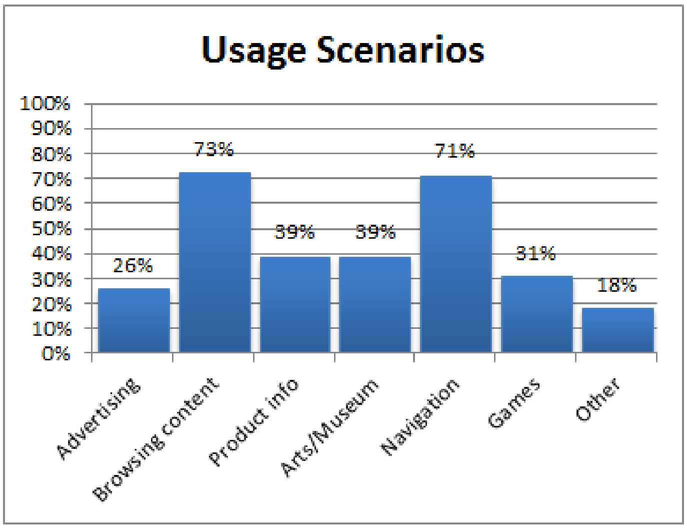


Figure 3.8: Usage scenarios.

Half of the responders rated browsers good to very good for accessing product information (44%) or guidance (47%), a third for browsing content (32%), advertising (31%) or museums (29%) but only 22% for gaming (see Figure 3.9). However, a quarter to a third of the participants was still uncertain of their quality for advertising (26%), museums (29%), and games (29%). This might be explained by the relative low number of participants who used *AR* browsers in these settings. In contrast to the ratings of the current state of *AR* browsers (see Figure 3.9) most participants gave high to very high ratings for the potential of *AR* browsers in the various application domains (see Figure 3.10).

As the gathered data was ordinal we used a rank based correlation measure (Kendall's τ). There are moderate positive rank correlations between current usage and usage potential ratings only for general purpose browsing and navigation (based on Kendall's τ, two-sided, excluding "Don't know") (see Table 3.1). There are no significant correlations for the other application domains.

Consumed media Most participants experienced *POIs* of textual form (77%), followed by 51% who experienced images and 43% of the users consumed 3D content. More complex web content (such as embedded webpages) and videos were experienced by only a third (27%) (see Figure 3.11).

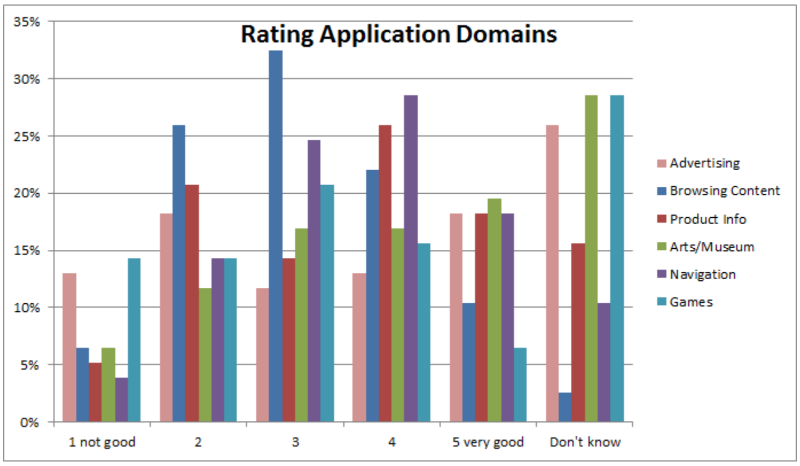


Figure 3.9: Rating of performance of current ARB for application domains.

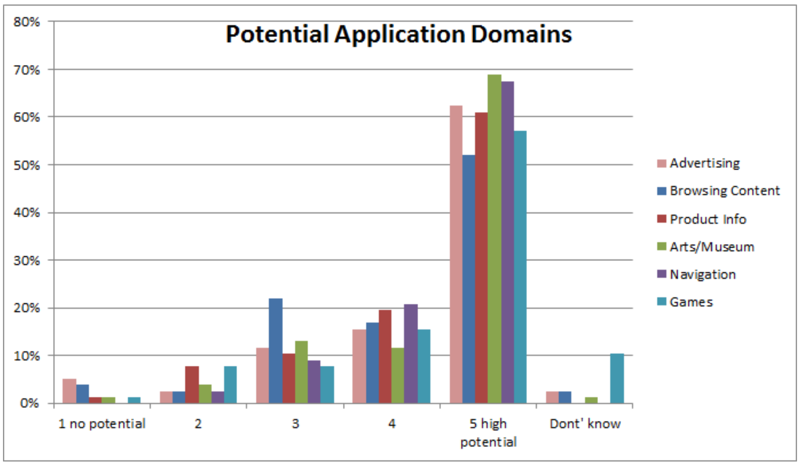


Figure 3.10: Rating of potential of ARB for application domains.

Domain	n	p-value	τ
Advertising	75	.23	.12
Browsing	75	$<$ **.001**	.33
Product Info	77	.934	$-.01$
Arts/Museum	76	.53	$-.06$
Navigation	77	**.017**	.23
Games	69	.89	$-.01$

Table 3.1: Kendall's τ rank correlation between current usage rating and usage potentials.

Feature quality and issue frequency Figures 3.12 to 3.14 show boxplots of rated quality of several features together frequencies of experienced issues with the same features.

A Kendall's τ test revealed moderate negative correlation between rating of feature quality and frequency of experienced issues for position accuracy, position stability (see

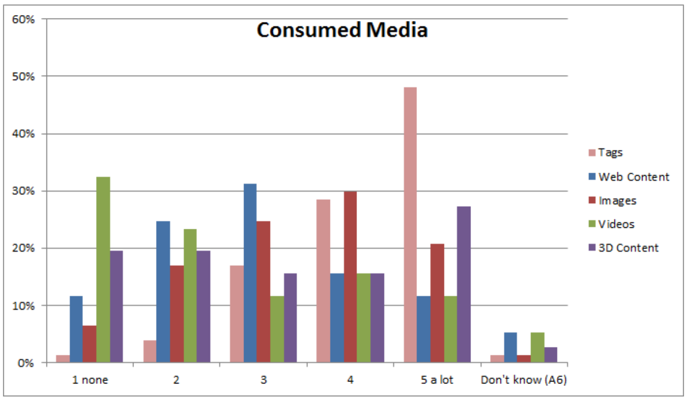


Figure 3.11: Type of consumed media.

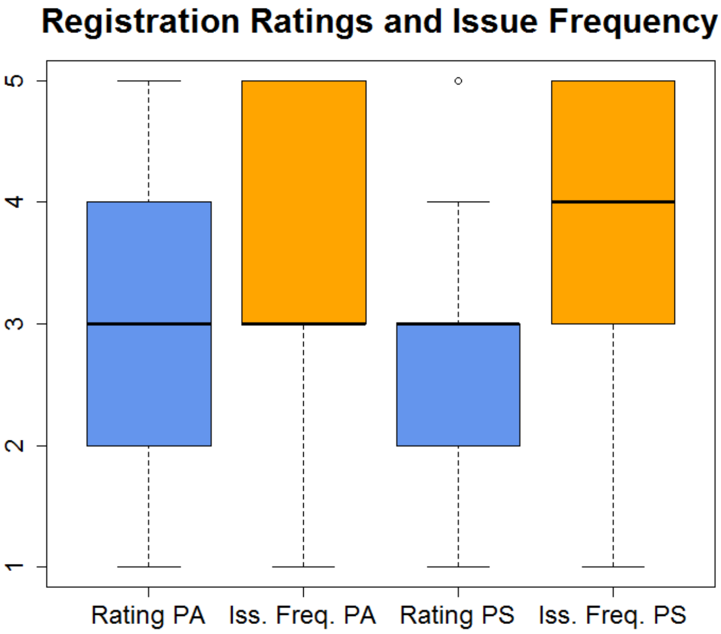


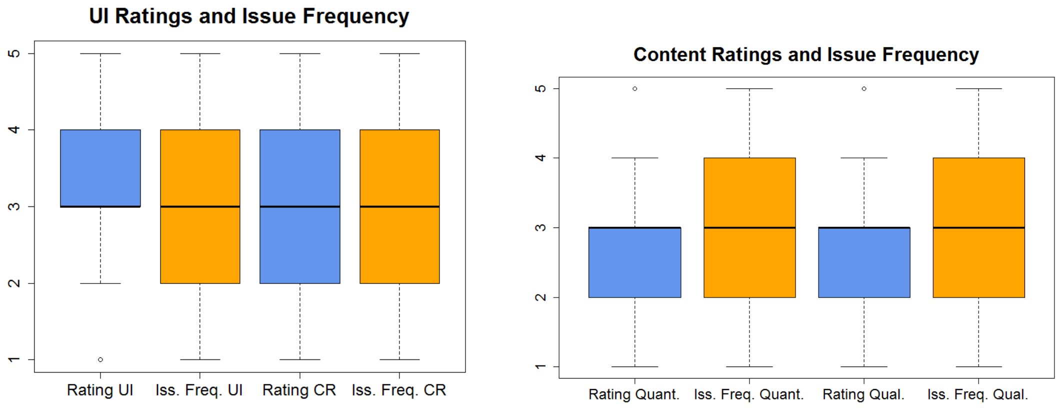
Figure 3.12: Registration quality rating (blue) and issue frequency (orange). PA: Position Accuracy. PS: Position Stability.

Table 3.2).

For the above mentioned features (except for device handiness and weight which have a high rating with low issue frequency) low to modest ratings go along with modest to frequent experiences of issues.

A one-tailed Mann-Whitney U test indicated that professional *AR* users rated content representation significantly lower (Mdn=3) than novel users (Mdn=3, 4), $U = 511, p = .02$.

The test also indicated that frequent users rated position stability significantly higher than non-frequent users (see Table 3.3), as well as content representation. Frequent users



(a) UI: User Interface. CR: Content Representation.

(b) Quant: Content Quantity. Qual: Content Quality.

Figure 3.13: User interface (a) and content related (b) ratings (blue) and issue frequency (orange).

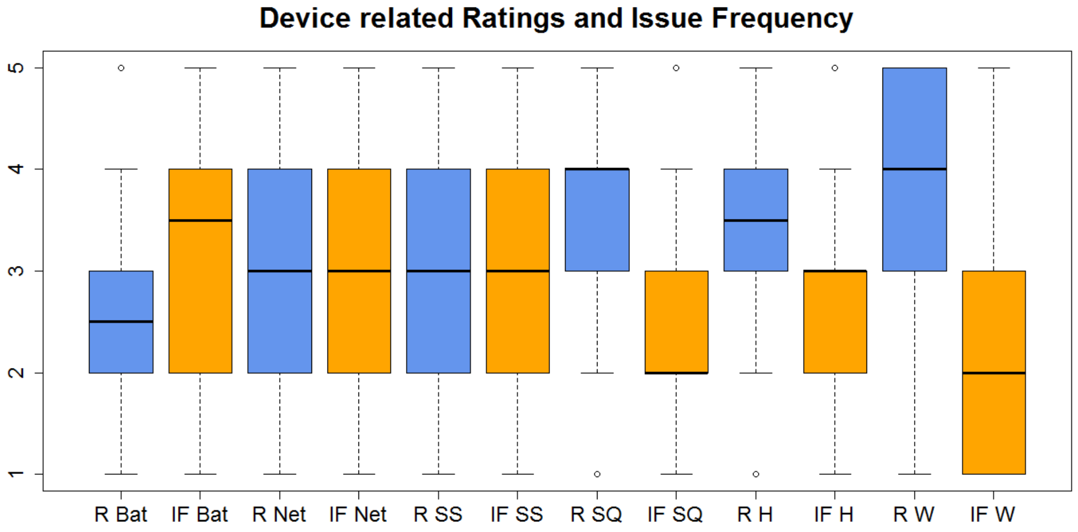


Figure 3.14: Device related quality rating (blue) and issue frequency (orange). Bat: Battery. Net: Network. SS: Screen Size. SQ: Screen Quality. H: Device Handiness. W: Device Weight.

rated content quantity, content quality significantly higher and experienced issues with content quality not as frequent as non-frequent users. In addition issues with content quality did not appear as frequent for frequent users than for non-frequent users (Mdn=3 for both groups), U=538.5, p=.047. For other issues no significant differences were detected.

Looking at the differences between frequent and non-frequent users, a one-tailed Mann-Whitney U test also indicated that long-term users rated position stability, content repre-

Issue	n	Rating Mdn	IF Mdn	p-value	τ
Registration					
Position Accuracy	76	3	3	$<$ **.001**	$-.42$
Position Stability	77	3	4	$<$ **.001**	$-.45$
UI					
Interface Design	77	3	3	$<$ **.001**	$-.44$
Content Representation	76	3	3	**.001**	$-.32$
Content					
Quantity	75	3	3	$<$ **.001**	$-.40$
Quality	75	3	3	$<$ **.001**	$-.45$
Device					
Battery	70	3	3	$<$ **.001**	$-.41$
Network	76	3	3	$<$ **.001**	$-.50$
Screen Size	76	3	3	**.004**	$-.27$
Screen Quality	75	4	2	$<$ **.001**	$-.44$
Device Handiness	76	3, 4	3	$<$ **.001**	$-.50$
Device Weigth	75	4	2	$<$ **.001**	$-.47$
Other					
General	76	3	3	$<$ **.001**	$-.38$

Table 3.2: Kendall's τ rank correlation between ratings of issue quality (low to high) and frequency of issues (never to very often). Interquartile range was 2 for all ratings and issue frequencies (IF Mdn: Issue frequency median).

Rating	n	Mdn f	Mdn nf	p-value	U
Position Stability	76	3	2,3	**.05**	854.5
Content Representation	76	3	3	**.01**	921
Content Quantity	75	3	2, 3	**.0026**	861.5
Content Quality	75	3	3	**.004**	925.5

Table 3.3: Significant differences in feature quality ratings for frequent (f) vs. non-frequent (nf) users according to Mann-Whitney U test. Interquartile range was 2 for all ratings.

sentation significantly higher than non-frequent users (see Table 3.4). For content quantity and content quality there was only a weak significant difference. In addition battery issues were experienced more frequent for long-term users ($Mdn = 4$) than for short-term users ($Mdn = 3$) $n = 70, U = 784.5, p = .018$, as well as device weight issues ($Mdn = 3$ for long-term, $Mdn = 2$ for short-term users), $U = 873.5, p = .023$.

Movement patterns Most of the users were experiencing the application while standing at the same position (78%), combined with rotations (90%). Small movements ($< 5\ m$) were carried out by 57%. Large movements ($> 5\ m$) or multiple large movements were conducted by 48% respectively 42% (see Figure 3.15).

A Chi-squared independence test with Yate's continuity correction indicated significant differences between frequent and non-frequent users for standing combined with rotation,

Rating	n	Mdn lt	Mdn st	p-value	U
Position Stability	76	3	3	**.02**	897
Content Representation	76	4	3	**.008**	930
Content Quantity	75	3	3	.092	568.5
Content Quality	75	3	3	.07	556

Table 3.4: Differences in ratings in feature quality ratings for long-term (lt) vs. short-term (st) users according to Mann-Whitney U test. Interquartil range was 2 for all ratings.

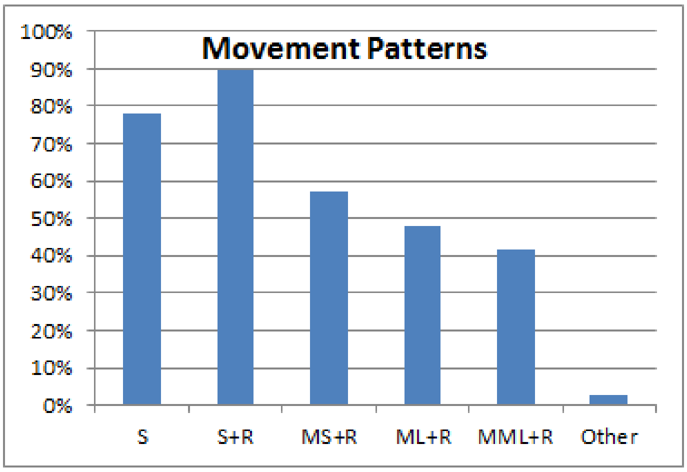


Figure 3.15: Movement patterns. S: standing. S+R: standing combined with rotation. MS+R: small (1-5 m) movements combined with rotation. ML+R: larger movements (> 5 m) combined with rotation. MML+R: multiple large movements (> 5 m) combined with rotation.

$\chi^2(1, n = 77) = 5.47, p = .02$ (see Table 3.5) and multiple large movements (> 5 m) combined with rotation $\chi^2(1, n = 77) = 5.94, p = .01$ (see Table 3.7).

There was also a significant difference in multiple large movements (> 5 m) combined with rotation between long-term and short-term users, $\chi^2(1, n = 77) = 10.05, p = .002$ (see Table 3.7) and professional and novice AR users, $\chi^2(1, n = 77) = 5.55, p = .02$ (see Table 3.8). Furthermore, between professional and novice AR users there were significant differences for small (1-5 m) movements combined with rotation $\chi^2(1, n = 77) = 4.81, p = .03$ see Table 3.10), as well as a weak significant difference for larger movements (> 5 m) combined with rotation $\chi^2(1, n = 77) = 3.35, p = .07$ (see Table 3.9).

This analysis showed that while ARB were used by half of the participants also with large movements, frequent and long term users tend to restrict their movements more then non-frequent or short term users.

S+R	frequent	non-frequent
no	43	26
yes	1	7

Table 3.5: Contingency table for standing combined with rotations (S+R) grouped by usage frequency.

MML+R	frequent	non-frequent
no	24	8
yes	20	25

Table 3.6: Contingency table for multiple large movements ($> 5\ m$) combined with rotations (MML+R) grouped by usage frequency.

MML+R	long-term	short-term
no	21	11
yes	12	33

Table 3.7: Contingency table for multiple large movements ($> 5\ m$) combined with rotations (MML+R) grouped by active usage duration.

MML+R	pro	novice
no	7	25
yes	23	22

Table 3.8: Contingency table for multiple large movements ($> 5\ m$) combined with rotations (MML+R) grouped by AR background.

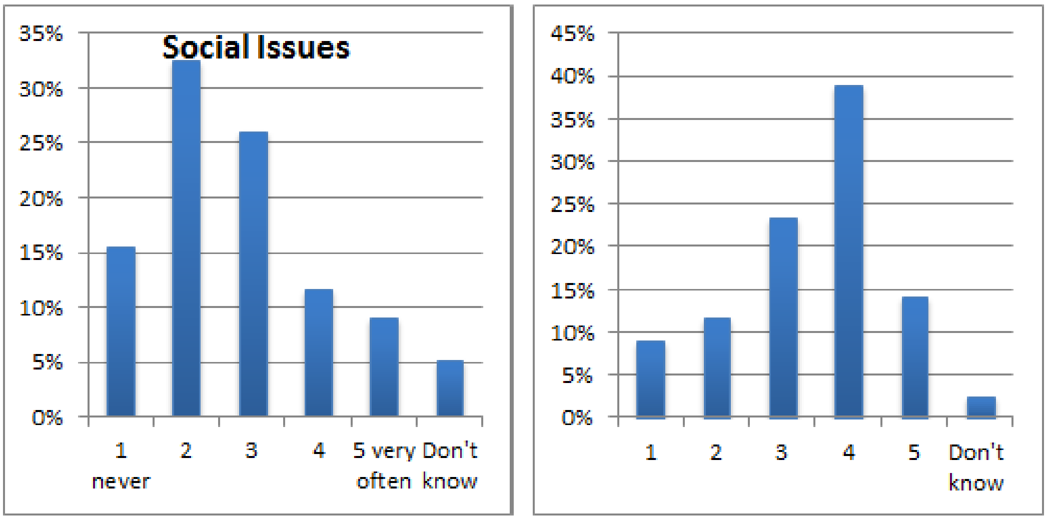
ML+R	pro	novice
no	10	27
yes	20	20

Table 3.9: Contingency table for larger ($> 5\ m$) movements combined with rotations (ML+R) grouped by AR background.

MS+R	pro	novice
no	12	32
yes	18	15

Table 3.10: Contingency table for small (1-5 m) movements combined with rotations (MS+R) grouped by AR background.

Social aspects The majority of the subjects did not experience regular social issues when using ARB and agreed to use the browser despite potential social issues (see Figure 3.16). The majority also did not experience situations (as shown in Figure 3.17) in which they refrained from using the application. A one-tailed Mann-Whitney U test indicated that female users refrained from using AR browser in crowded situations significantly less often than male users (Mdn=1 for both groups), $U = 532, p = .03$. No other significant effects were observed.



(a) Number of occurences of experienced social issues with *AR* browsers (5 point Likert scale. 1: completly disagree. 5: completly agree).

(b) Agreement to use *AR* browsers despite potential social issues (5 point Likert scale. 1: completly disagree. 5: completly agree).

Figure 3.16: Times social issues were experienced (a) agreement to use *AR* browser despite potential social issues (b) ratings.

Qualitative feedback Subjects were asked to provide reasons for withdrawing their usage of *AR* browsers if they did so; 31 (40%) of them provided free text answers. The answers were coded in a data-driven fashion [231] into 12 categories with 46 items. An overview of the reasons for discontinuation of ARB usage can be seen in Figure 3.18a.

Some answers for categories were:

1. Registration:
   - "Sensors are insufficient for suitable overlay"
   - "It is not so reliable. Often the compass and the gps doesn't work"
   - "Not useful as it was not spatially accurate"
   - "Lack of relevance to physical surroundings"
2. Content:
   - "Nothing interesting to see"
   - "No interesting content"
   - "There's not much useful information"
3. Maps:
   - "I don't find it as convenient as just using something like Google Maps"
   - "Google Maps is easier"

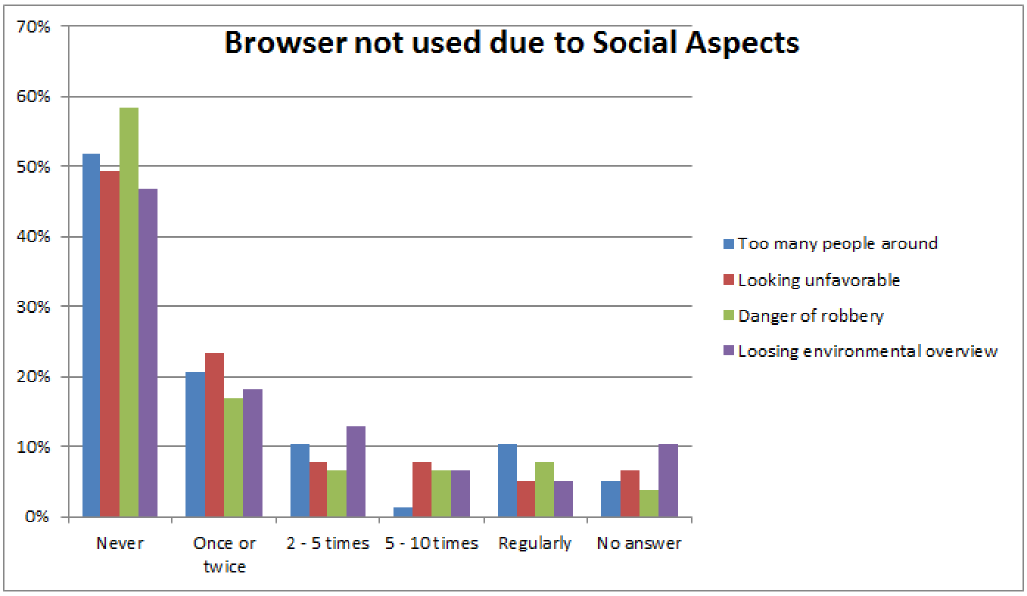


Figure 3.17: Times *AR* browsers were not used in several situations.

- ”No advantage over Google Maps, less useful than Google Maps + internet recommendations for e.g. restaurants”

4. Missing purpose:
   - ”Not much real use-cases”
   - ”Generally I don’t find them very worthwhile (to use privately)”
5. Visual clutter:
   - ”Information is not really helpful to me, because to it is to cluttered”
   - ”Too many *POI* one over the other”
   - ”UI is always cluttered, information is not well structured”
6. Concept:
   - ”There was no need to overlay icons on top of video”
   - ”It is annoying to hold up the phone all the time” (translated from German)
   - ”Holding up phone is unnatural, dangerous in certain circumstances”

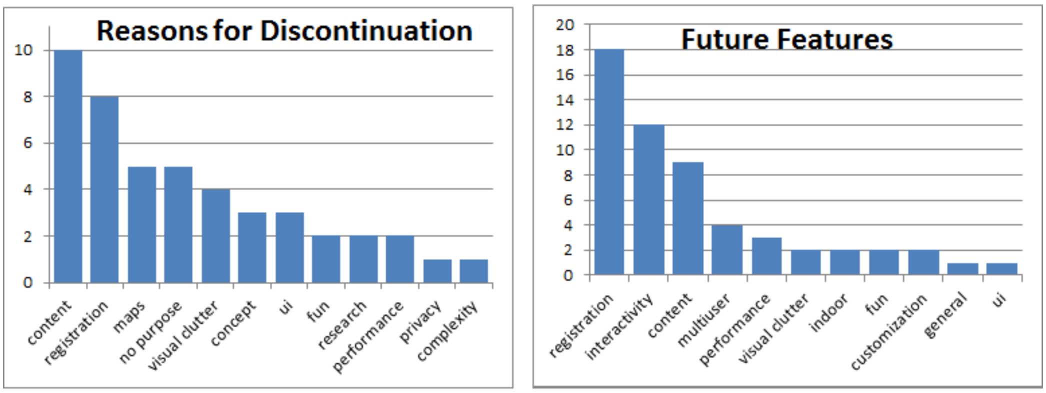
In addition, subjects were asked to provide ideas for future features of ARB; 37 (48%) of them provided free text answers. The answers were coded into 11 categories with 55 items. An overview can be seen in Figure 3.18b.

Some answers for each category were:

1. Registration

- ”Need to find a way to calm down the jumpiness!!! Make it more exact”
- ”Better location accuracy, robust *POI* display”
- ”Better location, better overlay on real world objects”
- ”Vision-based *AR*”

2. Content
   - ”Interesting stuff to see”
   - ”More Content”
   - ”User-generated content”
   - ”Better designed content, more variety in regards to types of documents/files, more tools”
3. Interactivity
   - ”More interactive features (comments, rating, participating)”
   - ”More interactivity”
   - ”3D interactivity”
4. Visual clutter
   - ”Well-arranged content, techniques for remove clutter”
   - ”More interactive, better filters”
5. Multi-user
   - ”IM integration to contact a person if located nearby in the real time”
   - ”Multiuser stuff”



(a) (b)

Figure 3.18: Reasons for discontinuation (a) and requested future features (b).

3.2.1.3 Discussion

Our survey has mainly collected feedback from computer literate persons. Similar to other emerging technologies, like location-based services, users of ARB are early adopters who have a high interest in technology.

On one hand a third of the participants used the ARB just for a few day (five days or less: 33%) and less than six times (34%), indicating a large group of the participant's merely tried out the browsers. On the other hand a 42% of the participants used the the ARB for at least 3 months and 42% at least weekly, indicating a that there is a regular user base of ARB that use them for mere than just 'trying out'. Similar to the usage patterns of other mobile applications [28] ARB are typically used only for a few minutes per session.

Besides general purpose browsing, participants used ARB for navigation purposes most frequently. This could indicate that participants used the ARB as alternative to map-based navigation methods. While the participants gave high ratings for the potential of ARB for wide range of application scenarios the ratings of the current performance of ARB in these domains (except general purpose browsing and navigation) did not correlate. This could indicate that people have high expectations in the ARB which are not met yet.

Augmented Reality leverages it's potential with accurate spatial registration of virtual content to real-world scenes in real-time. If the geometric registration between real and virtual objects is weak the semantic link between the two might become unclear as well. Currently, the consumed content in ARB is mainly of simple form, such as textual tags (77%) or images (51%). Even if 3D content is available (as consumed by 43% of the participants) it is still mainly registered with two *DOF* (longitude, latitude). This can result in meaningless or cluttered overlay of content on the ARB screen. Our study results indicate that content and registration issues are a factor for discontinuing the use of ARB. But even frequent users do rate the quantity or quality of available content only as average. Another common issue with the use of ARB is the large power consumption that results in perceived issues with the battery life of mobile devices. Registration, content and the interaction with that content were also among the most requested features for future versions of the ARB.

ARB were used by half of the participants also with large movements, but frequent and long term users tend to restrict their movements more then non-frequent and short term users, possibly adapting to the difficulties that arise when reading the information while moving. Previous studies investigated the reading performance of simple text while walking (e.g., [170, 216]) or automatic determined text readability over different backgrounds ([150]) but the impact of a changing camera image together with a possibly jittering augmented information while walking has not been investigated so far and should be explored further.

Generally, participants experienced no social issues when using ARB regularly.

3.2.2 Mobile distribution platform analysis

To complement our online survey we analysed the customer feedback available from the two dominant mobile software distribution platforms: The Apple App Store and the Google Android Market. We looked at the ratings and user comments for both stores and thus for some of the most popular *AR* browsers. As rating and commenting require users to authenticate, being limited to only one entry, this filtered information (no profanity, nominative) can provide us some interesting insights in the popularity of these *AR* browser applications.

3.2.2.1 Method

To collect the data, for the Apple App Store we used the AppReviewsFinder software[4]. For Android Market we used the data from the official Android Market homepage [5]. Data from both stores were gathered in June 2011 and represent the feedback given until then. Please note that the type and amount of information that can be retrieved from both distribution platforms are not symmetric. For example you can access country specific statistics for the Apple App Store while there are no country specific statistics available for the Google Android Market. It was also not possible to retrieve all user comments from the Android Market, limiting our analysis for this type of data to the Apple App store. Certain precise information are available to the developers of the software only (e.g., total numbers of downloads) and official information are only a rough indicator. We consequently decided to not evaluate some of these information. The presented numbers of downloads is also biased by the fact that some smartphone manufacturers have pre-installed some of these *AR* browsers but also are included in the total number of downloads, despite the fact that users never explicitly downloaded them. We also restricted our analysis to solely focus on the current state of *AR* browsers on these distribution platforms at a specific period of time, and not considering the temporal aspect (e.g., trends over time for download, comments, adoption for specific countries).

3.2.2.2 Results

We describe here our review of the ratings in both distribution platforms and a deeper analysis of the comments in the Apple App Store for different ARB.

Ratings At the time of our study, we collected - for the different *AR* browsers - about 70.000 ratings for the App Store (multi-countries); about 30.000 ratings for the Android Market were available. Both mobile distribution platform use a 5 stars rating system (5 stars are very good, while 1 star is very poor).

On the Apple App Store we identified five ARB that are prominent in terms of users-base and countries they are available. Based on the numbers of ratings they are SekaiCam (27364 ratings), Layar (23385 ratings), Acrossair (9150 ratings), Wikitude (5443 ratings) and Junaio (3382 ratings). Oppositely, there are only two ARB that achieved more than

[4] http://www.massycat.co.uk/iphonedev/AppReviewsFinder/, last retrieved 20.04.2015.

[5] https://market.android.com, last retrieved 20.04.2015.

1000 user ratings on the Google Android Market: Layar and Wikitude. For both the number of ratings nearly matches the ones from the Apple App Store.

The analysis of the gathered data showed that the average rating for all major ARB was very similar (overall average 2,49 stars) and also the differences in the average rating can be nearly ignored (Max: Layar 2,62 stars, Min: 2,39 stars Junaio). While examining the Android Market data it showed up that except SekaiCam all other applications rated significant higher on the Android platform (average 3,65), which can be caused by stability problems on the certain platforms or certain expectations that are platform dependent (see Figure 3.19). As an example many iOS users have higher expectations regarding the implemented interface and the application quality as both have so far been on a higher level for applications running on iOS.

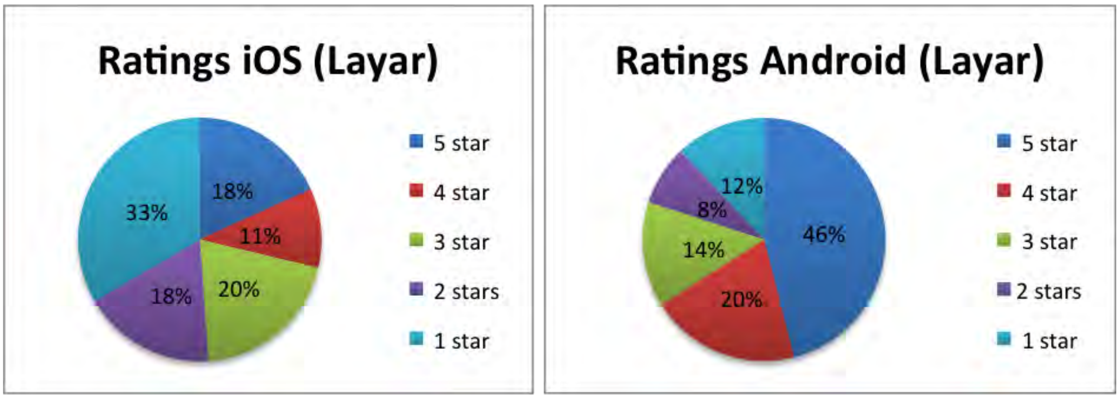


Figure 3.19: Difference of user ratings om both platforms based on Layar as example case (5 stars are very good, while 1 star is very poor).

The average rating is always the results of rather mixed ratings for all examined ARB as the standard deviation ranges from 1,38 (Wikitude) to 1,59 (Junaio) saying that many users gave very high or very low scores.

Based on the users feedback in the Apple App Store it is also possible to analyse the difference in ratings between countries. In general there is for all applications only a small deviation in the rating between the countries (Min. Layar SD = 0.38, Max. Junaio SD = 0.63). This is also reflected in the standard deviation of the ratings for each country, which are all nearly the same and showed that there are no significant effects that are based on cultural differences.

However, it is noticeable that for all countries with more than 100 ratings (to compensate outliers) South Korea was always in the top group of top ratings, while France was always among the countries with the lowest average ratings. But since the differences between the best and the worst ratings per country were only minor this can only be seen as a weak trend. Furthermore, it SekaiCam got in average lower ratings in German speaking countries (Germany and Austria) but again the difference was small (though noticeable) and could indicate content issue or a bad localization. Based on the total number of ratings the most feedback came from users out of the USA followed by Japan, UK, Germany, South Korea and France with each application getting a relatively big number of ratings from the country of origin (Acrossair/UK, Junaio/Germany, Layar/Netherlands, SekaiCam/Japan and Wikitude/Austria).

Comments We analysed 1135 comments from some of the most common western languages (English, German and French language) for all major ARB on the Apple App Store.

Analyzing the content of the comment, we categorized them in different groups, removed the basic and rhetorical liking type of comments, focused strongly on comments with a negative connotation or arguing about specific aspect of ARB. In result, we obtained 5 major clusters (some with subgroups): applications crashes, content availability, user interface and visualization (contains comments about the graphical interface as well as the used visualization of the content), tracking quality and general performance (contains comments regarding perceived performance, problems with network performance or comments regarding power consumption). An analysis regarding the occurrences in our dataset can be seen in Figure 3.20.

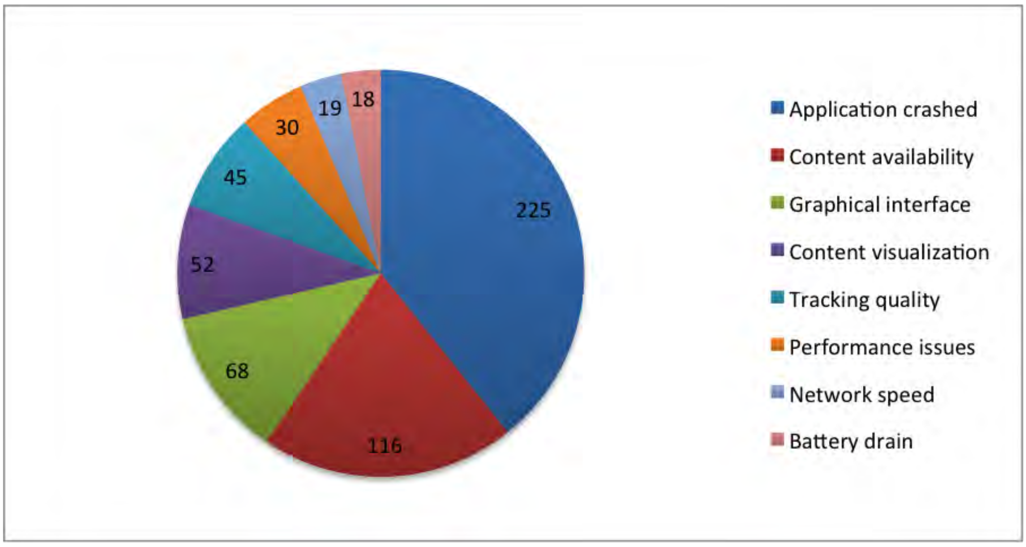


Figure 3.20: Result of clustering the total 1135 comments of the Apple App Store by focusing on negative connotations.

In the following we present a deeper analysis of the clustered comments.

1. Application crashes: From the total amount (1135 comments) 225 comments contained complaints about regular crashes. This is by far the biggest category of complaints, which is also an indicator while the ratings were so mixed between 1 star and 5 stars as most people with repeating crashes gave 1 star. It shows up in the comments that especially maintaining software version for every new system version or new hardware can be quite challenging.

2. Content availability: The second biggest category of complaints was regarding the availability of content. Many people expressed their disappointment with the amount and quality of available content. This ranges from no available content at all ("There

were hardly any *POIs* in Charlotte, NC") to very limited amount of content ("I looked for *POI* near me and all it came up with was a Post Box in the next street"). Furthermore many users had certain expectations regarding the content that were not fulfilled. Some users complained that the content is still not up to date ("Then I tried supermarkets, and it found one non-existent supermarket in our town") or needs to be paid.

3. User interface and visualization: Another problem that was raised in several comments was the quality of visual representation. Firstly, in form of the graphical interface (menus and buttons) that was considered several time as not very intuitive or not nice enough compared to other iOS apps. Furthermore, many people complained about the visualization of the displays content (such as *POIs*), which can become unreadable if to many *POIs* are in close proximity ("It stacks up results until you need to point at the sky to read them") or have a general low quality ("Can't wait until *AR* has real graphic experiences").

4. Tracking quality: Some people addressed in their comments problems with positioning accuracy that are usually caused by a bad GPS signal or an inaccurate orientation estimate ("I played with this app near my home town and it misidentified the location of our closest hospital - it was WAY off").

5. General performance: Only a few people had problems with the general performance or the speed of the necessary network connections. However, some people suggested a caching mode, which would help users in foreign countries (e.g. tourists) to use the application even if they don't use a (expensive) 3g connection by prefetching and caching the results when a connection is available.

   To our surprise only a small amount of users commented about the drain of battery caused by most *AR* browsers ("Tremendous drain on battery life. Actually causes my 3gs to heat up a lot", "But if it's gonna kill my battery, it has no place on my phone."), which we think originates in the fact that only a few people used *AR* browser for a longer time and consequently have experienced that sudden loss of battery power.

Beside these problems that were addressed in the comments many users also gave a good feedback that was often also justified with the fact that most *AR* browser are free to download. Many people also expressed their general interest as they identified the potential. We often read sentences saying that the current amount of content is small and there are still some bugs but that they will check back after some time as they think these applications have a huge potential. This supports also the comments of most people giving positive ratings as they often commented about the novel interface and how interesting it is but only a very small number commented on how they made real use of *AR* browsers.

3.2.2.3 Discussion

Overall the data from the distribution platforms show that the existing *AR* browsers perform similarly in term of user ratings. It also shows that there are no strong indicators for country specific or cultural specific effects in respect to the ratings. While the total number of ratings indicate that a large number of users at least tried *AR* browsers once, the real number of permanent users is still hard to estimate. Especially as the ratings suggest that the users opinions are quite different; many gave a low score - and it is likely that they stopped using *AR* browser - while another large group gave a high score. However, it suggests that there is likely a novelty effect affecting the high score of the second group. The comments also raised issues regarding the usefulness of the application, which brings the questions of long term use of *AR* browsers.

The comments from the Apple App Store show that the stability of ARB is one of the major issue that should be solved with a better software quality management. Further problems are caused by the low availability of content and the quality of the implemented interface. Solving all these issues would resolve 75% of the user complaints. A smaller group of users also pointed out problems in regard to content visualization and the rapid battery drain. This should not be underestimated, especially as the amount and density of the content will increase dramatically in the future and end-users may use more permanently of *AR* browser, these problems may become a major issue.

3.3 Information access at event posters

We conducted an online survey about information access at printed information surfaces such as event posters. It is based on situations in which users would interact with (potentially augmented) posters. While our survey is targetd to inform the exploration of *AR* user interfaces for large information surfaces (Chapter 4) it is not targeted at delivering representative results of user behavior at those surfaces.

3.3.1 Survey

Thirty one participants (21 males, 10 females, age: M: 28.5 years, SD: 6.03) participated in the online survey, which was advertised via social network sites and e-mail. Their professional backgrounds were mainly in IT and design professions.

Most of the participants indicated to pay attention to event posters when waiting at public transportation stops (90%), at events like concerts (70%), followed by looking at posters in shops or bars (65%) and while walking through the city (56%). The majority of the participants (85%) stated that the name of the performers of an event should sound interesting (83% for event title) to engage further with the information on the poster. When participants decided to engage with a poster they did so for short durations (5-12 seconds 35%, 15-30 seconds: 48%). Asked about the type of information users try to remember, save or bookmark if they are interested in an event, 58% indicated to almost always remember the names of performing artists followed by the name of the event (48%), the venue (48%), and date (45%). However, 15% also pointed out to never or almost never remember the date. 65% specified that they would rarely remember website links. Habits of saving information for later reference included memorizing it (78%) taking pictures of the poster with their smartphone (33%) or scanning QR codes (13%). Other means for bookmarking were not used by the majority of the participants. While 50% of the participants access the information regularly when back home, 28% also access them through their smartphone on the move. Asked about which digital information they would like to access on an event poster participants mentioned ticket availability and prices (30%) as well as information about the event location (30%). Further information about the performing acts in form of multimedia content was pointed out by 45%. Only 15% explicitly mentioned means to bookmark the event and getting information about related events.

3.3.2 Discussion

Our survey partially confirms previous findings [210] about usage patterns at posters. Users typically engage with posters in opportunistic situations and only for a short time. Regarding access to further information, a third of the participants already used their smartphone to either bookmark an event (by taking a picture) or browsing further information while away from the poster. The observations indicate a current gap between the goal of extending the duration people spent with products or advertisements addressed through rich, interactive augmented print media experiences and the reality in which these interactions take place (namely in mobile contexts). A key insight from this survey - and previous research [238] - is that access of augmented print media in opportunistic situations should allow continuing the experience when moving on. However, this is still not considered widely when designing those type of user experiences. To address this gap we propose a novel type of hybrid interface to support exploration of digital content on augmented posters both on (poster) location and on the go.

3.4 Summary

This chapter introduced a literature survey on context-awareness in *AR* systems, a user survey on the usage of first generation *AR* browsers and a user survey on information access at large printed information surfaces such as event posters.

For our literature survey we started by creating a taxonomy based on three top-level domains: context sources, context targets and context controllers. We further explored possible categories within the top-level domains of sources and targets by specifically focusing on the unique aspects of an *AR* system. Once we identified possible domains, we reviewed existing research in the field of context-aware *AR* following the earlier created taxonomy. Based on this taxonomy, we identified opportunities for future research. Specifically, social context factors have not been considered in depth in existing *AR* systems. Furthermore, there is potential to investigate on varying infrastructure factors (e.g., the availability of situated displays in public space). Similarly, the availability (or lack of) multiple concurrent input and output devices for *AR* interaction has not been investigated so far. Hence, we see a potential to investigate *AR* interaction beyond a single input and output device such as an individual smartphone.

The user survey on first generation *AR* browsers indicated that a significant number of people tried *AR* browsers on their personal mobile devices and mostly noted positively the technology. They also pointed out their interest in this type of application. Interestingly, the used tracking technology, GPS for position, accelerometers and compass for orientation, was not as a limiting factor as we expected, especially as reflected by the feedback from the mobile distribution platforms. Participants also confirmed the high potential of this technology in the future, especially regarding some application areas such as content browsing and navigation. Some of the major issues were the scarcity of content on these platforms, the poor quality of the user interface (and user experience) and issues with battery life, due to the large energy consumption of the variety of sensors involved in a standard *AR* application. From the analysis of the distribution platform, comments indicated the lack of reliability and robustness of *AR* browsers, which is also a common issue for other mobile application. While this survey focused on *AR* user interfaces for location-based data, it is still relevant for interaction with information surfaces. Both the interaction with location-based data and interaction with information surfaces occur in varying mobile contexts, on handheld devices and through similar user interfaces. While the survey showed that there actually are long-term users, it is not yet well understood why and in which contexts these consumers employ mobile *AR* interfaces.

To complement the findings of the previous surveys with insights specifically about interaction with information surfaces, we conducted a online survey about information access at event posters. The survey indicated that information access at posters mainly happens in opportunistic situations and for a short period of time.

Given the nature of the surveys (literature, online survey) presented in this chapter it is advisable to complement them with in-situ observations about the usage of *AR* systems for information surfaces to broaden the understanding about their potential benefits and drawbacks in various usage contexts. This will be done in the subsequent chapter.

4 Interaction with Posters in Public Space

Contents

4.1 ML and SP interfaces for games in a public space 68
4.2 Repeated evaluation of ML and SP interfaces in public space . 83
4.3 The utility of ML interfaces on handheld devices for touristic map navigation 89
4.4 Exploring the design of hybrid interfaces for augmented posters in public spaces 108
4.5 Summary 117

In this chapter we investigate factors that can influence the usage of handheld AR user interfaces for interaction with large printed information surfaces. Specifically, we focus on posters in public space. They are a popular medium in commercial AR applications both for leisure and utility driven use cases. Still, the research community lacks understanding about the merits and drawbacks of mobile AR user interfaces for interaction with posters. The content production pipeline of printed posters often relies on a digital representation being available, which is often created in desktop publishing software. In fact, these digital assets are often used to produce more than one type of physical representation. Besides posters, also smaller form factors such as years are printed and the digital assets can be made available directly on websites or mobile apps.

It is at least partly this dual nature of the printed information surface content, having a physical presentation as well as a digital one, that has spurred a number of previous research focusing on comparisons of user interfaces for interaction with the physical or the digital counterpart (Chapter 2). One of the dominant interaction metaphors on handheld systems for navigating digital information spaces is the *SP* metaphor. It allows to move, rotate and scale a virtual information surface beyond a peephole (the screen of the handheld device), often using touch gestures such as pinch-to-zoom or drag-to-pan. Coming back to the dual nature of content for information surfaces, the *SP* metaphor can also be seen as one way to move a virtual camera through the digital representation of the information surface, just as AR on handheld devices allows to navigate the physical representation through spatial pointing. At the same time, both interaction metaphors are quite different in the way they are operated. *SP* requires mostly finger movements, AR relies on arm and upper-body movements. So far, it has not been understood well under which circumstances these interfaces show their respective merits in terms of performance. Even, when only investigation spatial vs. touch input (without the need to focus on an external physical reference frame such as a poster), studies have come to different conclusions if touch or spatial input is more efficient for navigating an information space. For example, while in 2014 Spindler et al. indicated that spatial input can significantly outperform touch based navigation for atomic navigation tasks [227], Pahud et al. came to the contrary conclusion. They indicated that spatial input was significantly slower than touch input for a virtual map navigation task [180].

Furthermore, both metaphors can be potentially different in the way they are received by bystanders. In terms of visibility of actions and effects (c.f. [190], users' actions with handheld AR user interfaces require spatial gestures. This can potentially result in a high visibility of these actions for spectators (mimicking attention grabbing pointing gestures [242]). At the same time.the effects of those actions remain hidden. In contrast, operating *SP* interfaces on handhelds, while also not revealing the effects of interactions, result in potentially lower visibility of the involved actions (such as finger movements). Consequently, these differences could lead to a different social acceptability for both interfaces (c.f. [196]). In turn, this could influence the acceptance of those interfaces for the users themselves.

These potentially different characteristics in terms of performance and social acceptability motivated the studies presented in this chapter. Hence, in the first part of this chapter, we will investigate the effects of various social settings on the usage of the *ML* metaphor of handheld AR compared to the *SP* metaphor. We do this for a gaming related scenario. Firstly, this is a common scenario for commercial applications and secondly, gaming lends itself to a higher engagement (compared to a solely utility driven task). This could potentially lead to more expressive and visible spatial gestures being used (which we wanted to compare with the private nature of *SP* interaction). In an initial study we compared interaction in a public space with a laboratory setting. This initial study was then expanded to include another public space with different spatial and social characteristics. In the second part of this chapter, we turn our focus to utility driven tasks. We explore potential benefits and drawbacks of *ML* and *SP* for information browsing at tourist maps. Compared to the initial gaming related scenario, which was mostly exploratory in nature, we directly compared *ML* and *SP* with an extended set of user experience and performance measures. We complemented the study at a public space with a laboratory study, to further investigate the role of the information space size on the utility of *ML* and *SP*.

Finally, we will present an interaction concept and prototype which integrates both *ML* and *SP* interfaces into a hybrid interface to allow continuous interaction with information surfaces across multiple usage contexts.

4.1 ML and SP interfaces for games in a public space

Within this part of the thesis our main research interests are to explore *if and how* people would use a *ML* interface for a mobile game in a public location when a *SP* interface is available as alternative, to gauge the *reactions from the general public* and to determine the *impact of location and audience on task performance.*

Therefore, we designed a mobile phone game that could be played at a poster mounted at a public building in a transit area or on the smartphone alone but at the same location. We complemented the observations at the public space with observations of a separate group conducting the same tasks in a controlled laboratory setting.

With this work we add insights about user and audience behavior when using a *ML* interface outside the laboratory and complement existing studies that investigated collaborative use of mobile *AR* systems in the wild.

4.1.1 Game design and implementation

Find-and-select tasks are common in mobile *AR* games. Users are required to physically translate (pan and zoom) and eventually rotate their phones in order to detect targets; selection is typically accomplished through touching the screen. While mobile *AR* games often employ only a *ML* interface to solve the task, mobile *AR* browsers offer alternative list and *SP* views on the data. *SP* interfaces for smartphones allow navigation through dragging (pan) and pinching (zoom).

We wanted to observe how users would adapt to *ML* and *SP* interfaces if they can solve a task with either interface in a public space. We decided on a simple find-and-select task similar to previous performance-centric studies [111]. To engage people over an extended period of time at one location, we designed a game-like experience with background music, audio, graphical effects and challenges. Each level lasted approximately one to two minutes; playing 8 consecutive levels could eventually lead to fatigue. The game could be played with a *ML* and with a *SP* interface (see Figure 4.1) that showed similar views on the game to lower the mental gap when switching between them. The interaction methods to find the targets were different between the interfaces (physical pointing in *ML*, drag-to-pan and pinch-to-zoom in *SP*). Selection was accomplished by clicking in either interface. The poster as reference frame for the game was available in both interfaces (physical for *ML*, virtual for the *SP*). The Field Of View (FoV) of the virtual camera was set to match the one of the physical camera. For the game we did not focus on collaborative activities. Instead, the game tasks required the players to repeatedly find a 'moving worm' that could appear at one of 20 locations (apples on a tree) in two possible sizes. Individual targets had to be selected three times before appearing elsewhere.

To select the targets, users had to be in a minimum distance in front of the target (ca. 30 cm for a small target, ca. 60 cm for a large one) forcing them to physically move back and forth with the *ML* interface or to pinch in and out in the *SP* interface.

Users could explicitly switch between the interfaces by pressing buttons at the bottom of the screen which would show the closest orthogonal view of the virtual poster when switching from *ML* to *SP*. When users pointed their phone down they implicitly switched into a standard view (showing approximately 2/3 of the virtual poster.)

The levels did not increase in difficulty to observe possible learning and fatigue effects; only the positions and sizes of the worms were varied randomly. There were 8 levels in total, each with 15 targets to be played. Through pre-experiments we adopted parameters for dragging and pinching speeds, the default scale for the virtual poster and the minimum

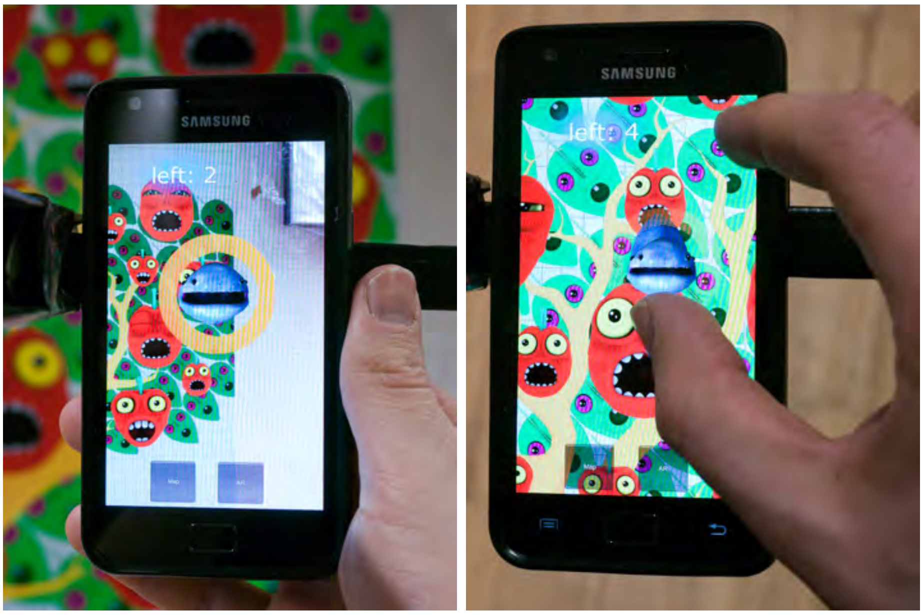


Figure 4.1: A large target within selection distance (indicated by orange ring) in the *ML* view (left). User pinching to zoom in to a small target in the *SP* view (right)

distances for target selection to ensure comparable times in both interfaces for a trained user.

The game was implemented in Unity with Qualcomm's Vuforia toolkit and deployed on a Samsung Galaxy SII smartphone running Android 2.3.

4.1.2 Study design

We designed an outdoor study and replicated a comparative indoor study to act as a control group. The outdoor study took place at a building below a large video wall on a central place in Graz, Austria (see Figure 4.2). The place serves as the main transit zone of the town to change public transportation lines and acts as a waiting area. In addition, musicians or advertisers can often be found here. Participants conducted the study in front of a DIN A0 sized poster that was mounted vertically at a height of 2 m. The control study took place inside a laboratory at Graz University of Technology (see Figure 4.4). Both the laboratory and outdoor studies took approximately one hour per participant and all participants were taken through the sequences by the same one researcher in the interests of consistency.

There were 6 phases: introduction (5 min), training (5-10 min), demographic questionnaire (5 min), main game (15-20 min), interviews and questions (10-15 min) and performance (10-15 min). In the initial training phase the participants were made comfortable with both interfaces to a level where they could explicitly and implicitly switch between the two. They also learned how to easily recover from tracking failures that could appear in the *ML* condition (e.g., due to fast movements or being too close to the poster, see Figure 4.3, left). As it was very cold (at times even down to -10°C, regardless, we wit-

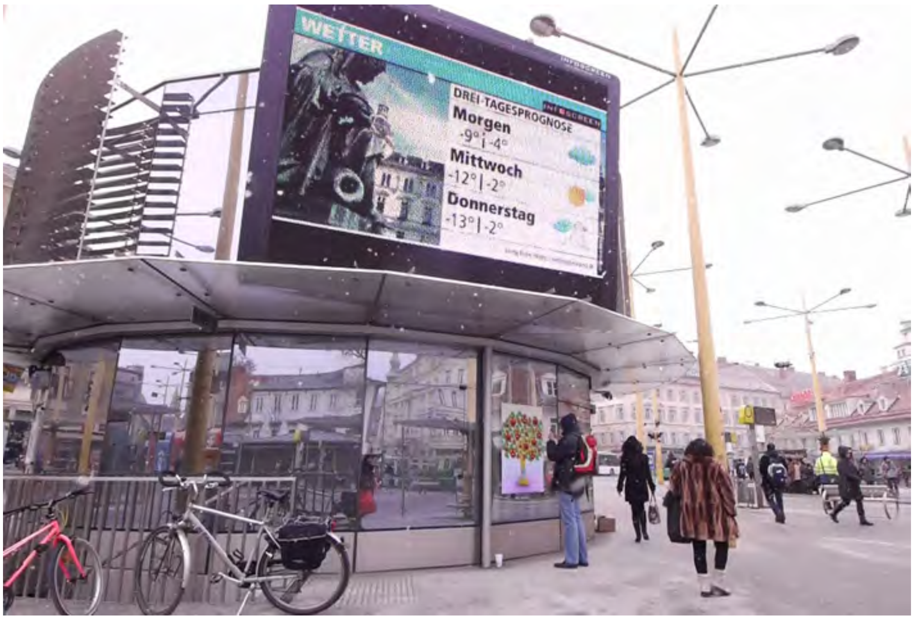


Figure 4.2: A participant playing the game in front of the poster at the public transit place in Graz, Austria

nessed people standing outside waiting for friends) after the training phase, participants filled out a demographic questionnaire in a nearby cafe.

In the main phase they were asked to select fifteen worms in 8 levels each. Participants were free to choose their preferred interaction technique. This was explained clearly in the training phase and again in the transition to the main phase. In addition, it was made clear they could switch interfaces as often as they liked, there were no restrictions on this. Participants were asked to complete the tasks but we clearly emphasized that their target focus was not speed or precision. Participants could set their own pace, taking breaks between the levels as they wished, with warm tea on hand.

The main phase was followed by a questionnaire and interview session in the same café where the demographic questionnaire was filled out. Finally, a performance phase was conducted at the poster similar to the one described by Henze et al. [111]. Participants had to find-and-select the bluest out of 12 boxes ranging from green to blue by panning and touching at a fixed distance (showing approximately 1/4 of the search area) 15 times in 4 repetitions resulting in 480 measurements per group and interface (see Figure 4.3, right). Participants were checked for color blindness before starting this test. This time they could only use *either* the *SP* or the *ML* interface at any one time. This meant that half of the participants started with the *ML* mode and then conducted the task in the *SP* mode, while the other half started with *SP* and then used the *ML* mode to ensure a balanced sample.

Further, a control group of eight participants conducted the exact same procedure from beginning to end, including the initial training and performance phases, but in an indoor laboratory setting. The laboratory setting did not have passersby, only each participant and the experimenter were present. The poster was mounted on the same height as in the

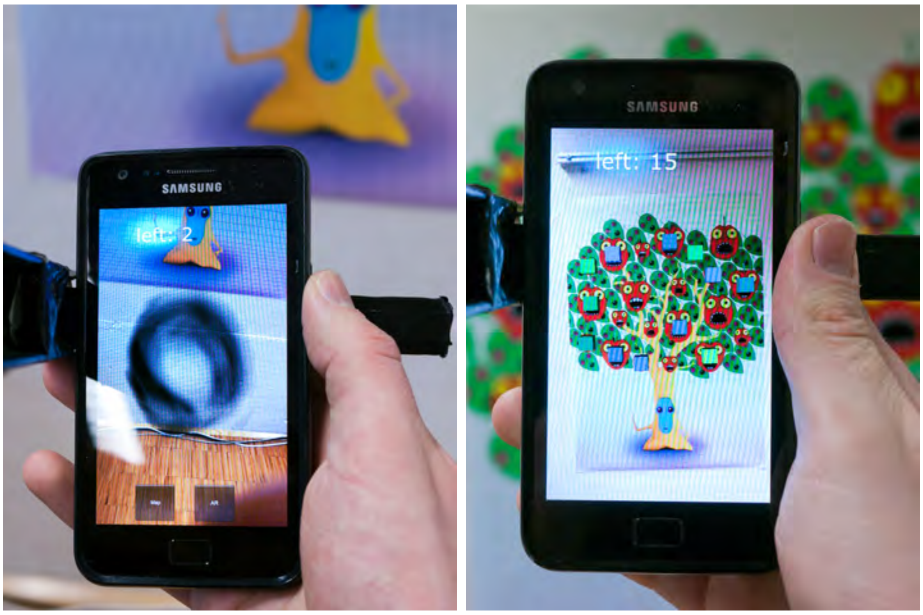


Figure 4.3: Tracking errors indicated by black circle in the middle of the screen (left). Overview of one configuration of colored target boxes in the performance phase (right).

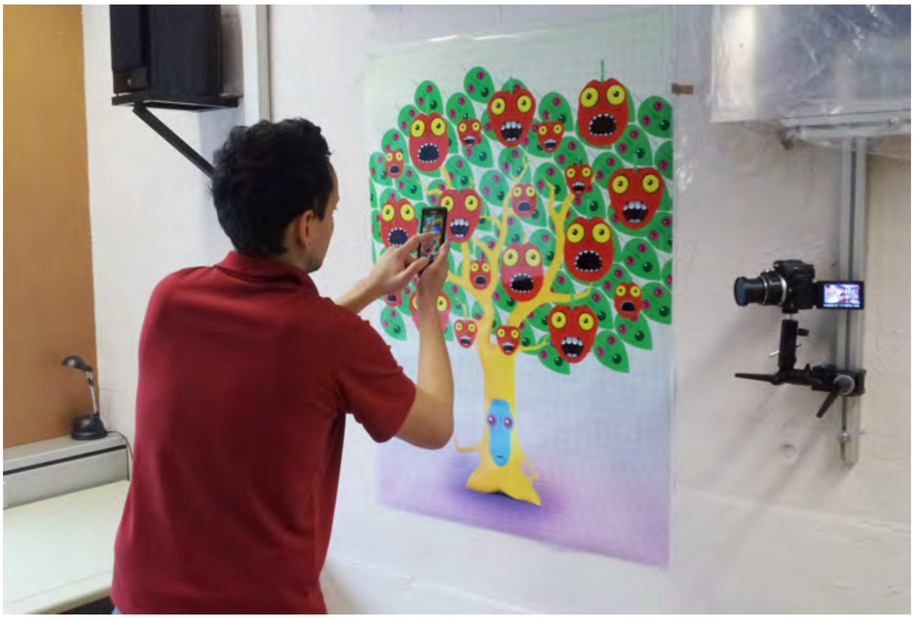


Figure 4.4: Participant playing the game in the laboratory.

public condition.

4.1.2.1 Participants

There were 16 participants in total (8 female, 8 male) evenly distributed between the study at the laboratory and at the outdoor location. In both groups participants were aged between 21 and 30 years. All of them had either a university degree or were studying. Five people in the public location group had a computer science, two a design and one a social science background. In the laboratory group four people had a computer science, three a design, and one a mathematical background. Thirteen of 16 participants were familiar with the idea of *AR*, or had used *AR* at least once, regularly or professionally. All but one participant never to rarely (at most 1 hour per week) played video games and all but one never played video games on mobile devices.

4.1.2.2 Hypotheses

We followed an exploratory approach for the main part of the study to obtain insights into how the participants would employ the system and how the public would react to the interactions of the participants, specifically with the *ML* interface. Nonetheless, we had the following two hypotheses: *H1*: *ML will be used less often in the public setting than in the laboratory.* We suspected that playing the game in the *ML* interface would cause more attention from the public and that participants would feel exposed and watched, eventually switching to the less obtrusive *SP* interface in the public setting. *H2*: *ML will be used less as the game progresses.* As the game levels were repetitive and the main phase was expected to last for 15-20 minutes we suspected that as arm fatigue increases and the novelty of the *ML* interface decreases participants would eventually switch to *SP*.

4.1.2.3 Data collection

We collected video, survey and device logging data, complemented with notes, stills and additional videos taken by one observer. Quantitative data was analyzed with Microsoft Excel and the R statistical package. Null Hypothesis Significance Testing (NHST) was carried out with the 0.05 level.

Video data A small camera with a wide angle lens (100° diagonal *FoV*) was vertically mounted next to the poster (behind a pillar in the public condition), which recorded participants' actions and the reactions from the public during the main task. In addition an observer took notes and additional footage with another camera. In total 2 hours of video footage (only for the main game phase) was collected for the public condition and processed by a single coder.

Survey data We employed questions that are based on Flow [236], Presence [221] and Intrinsic Motivation [51] research and were adapted through a series of studies [13, 17, 18]. We customized them for this study to capture reactions on the system and tasks in the environment using a 5-point Likert scale. A multiple choice questionnaire similar to [197] about location and audience was used and followed by a semi-structured interview focusing on how participants used the system and how they would use it in other settings.

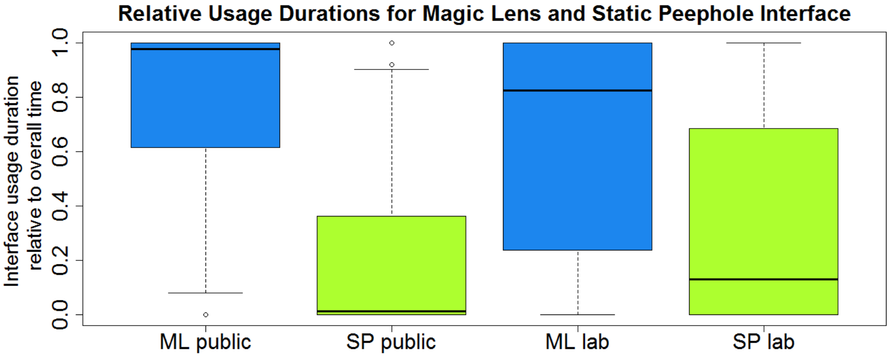


Figure 4.5: Relative usage duration for the *ML* (blue) and *SP* (green) interface in the public and lab condition.

Device data The position of the real camera (in *ML*) or the virtual camera (*SP*) mode was sampled at 10 Hz. Additionally, events such as touches, interface switches, *TCTs* interface were logged on the device. The timing data was not normal distributed so non-parametric *NHST* was applied. One participant in the public location had to abort the main phase after 6 of 8 levels but eventually continued with the performance phase.

Limitations While we employ *NHST*, we stress that with our limited sample size the results are particular to this situated instance. Further exploration with a larger sample in a wider variety of settings is required prior to being able to make any generalizations from our findings. As with many mobile trials conducted in a public space, the setting and tasks are generally somewhat contrived with participants aware that they are taking part in a study where they are accountable to the researcher team while doing tasks designed to test unknown (to them) research-related criteria.

4.1.3 Findings

We report on our observations combining quantitative and qualitative results as well as findings from the public and the laboratory setting where appropriate for our limited sample size.

4.1.3.1 ML was used most of the time

The *ML* interface was used 72% of the time (76% in the public setting, 68% in the lab) as illustrated by Figure 4.5. The *ML* interface was used weak significantly longer in the public setting than in the lab condition as indicated by a Mann-Whitney U test (p = 0.056, Z= -1.59). The significant difference is due to one participant playing solely in *SP* mode in the lab condition. But even with considering this one participant as an outlier (resulting in no significant difference in usage time of *ML* between both locations) our hypothesis *H1* that the *ML* interface would be used less in the public setting is contradicted. Figure 4.6 shows boxplots of the absolute TCTs over all levels.

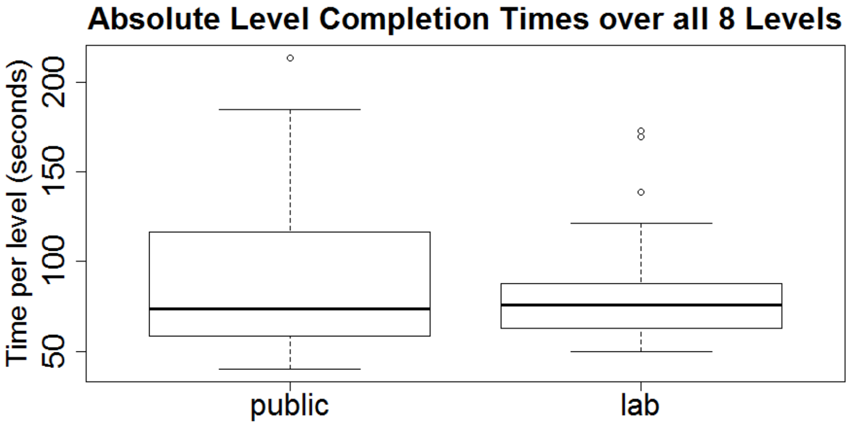


Figure 4.6: Absolute level completion times for the public and lab group.

A Mann-Whitney U test indicated no significant differences for completion times over all levels between the groups. In addition, a Friedman rank sum test did not reveal significant differences for *ML* usage duration between the 8 levels for the public location and for the lab group, thus contradicting hypothesis *H2 that the ML interface would be used less as the game progresses.* Figure 4.7 shows the relative usage duration of the *ML* interface over 8 levels in the public location group.

Generally, participants switched between a position in which they could get an overview of the whole poster to identify the target and then moved in to select the target.

We observed diverse ways of how participants handled the fact that they needed to move back and forth during the game and the holding of the phone itself. All but one participant used a relative fixed arm pose and moved using their feet, stretching their arms only for the last few inches towards the poster.

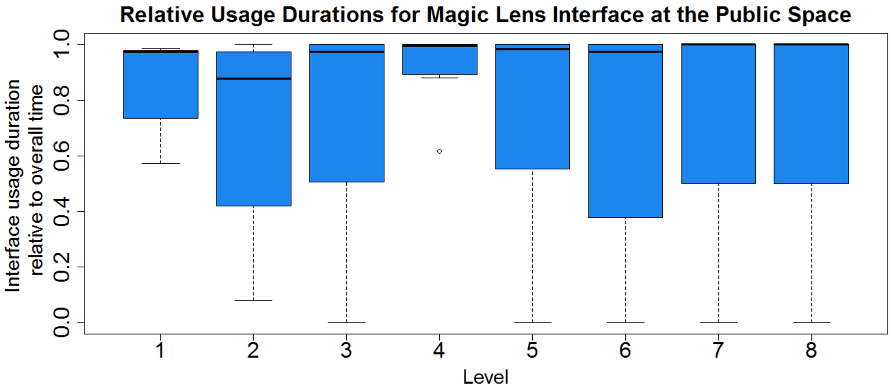


Figure 4.7: Relative usage duration for the *ML* interface over individual levels in the public setting.

As the mounting of the poster should reflect a possible real-world scene its height was not adjusted to match participants' height. Two small participants held the phone above

Questionnaire item	Result	p-value	Z-score
I enjoyed using the *ML* (MD=5) | *SP* (MD=3) view in the environment	*ML*>*SP*	0.036	1.80
I would rather do the task with the *ML* (MD=5) | *SP* (MD=2) view only	*ML*>*SP*	0.029	1.90

Table 4.1: Questionnaire items that were rated significantly higher for the *ML* over the *SP* interface in the public group.

their heads to reach targets at the top of the poster, one of them eventually switched to the *SP* mode after 4 levels. Three participants bent their knees regularly to hit targets at the lower half of the poster (see Figure 4.8).

The phone itself was held in various ways (see Figure 4.9). One participant switched from portrait to landscape mode to get an overview of the scene and stabilize tracking. Two participants held the phone on the long edge as the phone was more stable when touching it and subsequently tracking errors would be reduced; six held it on the short edge. Six participants held the phone mainly one handed, two used both hands. Two participants eventually used their gloves to hold the phone and changed them between levels due to the weather condition. We could not reliably identify fatigue as a single cause for changing hand poses. The tracking system failed regularly and participants adapted to the tracking system throughout the game. Three participants explicitly mentioned they had changed their hand poses to address tracking errors.

4.1.3.2 Reasons for using ML

A Wilcoxon signed rank test indicated significantly higher ratings for the *ML* over the *SP* interface for enjoyment and preference for the public location group (see Table 4.1).

Another participant who used the *ML* mode exclusively said “I would probably not use it if it would be commonly available”. Two participants explicitly mentioned that they felt being faster in the *ML* mode. One felt that the music was too attention grabbing in the environment and distracting, turned it off, and continued to play in the *ML* mode. Another mentioned that with the *ML* interface “you are much more in the game”. One participant said that she had a better overview in the *ML* mode and felt it was easier to step back and forth than to pinch-to-zoom. Similarly, another participant said the *ML* mode was “more intuitive”.

4.1.3.3 Reasons for using SP

While the *ML* interface was used almost exclusively by 6 of 8 participants in the public setting, two female participants eventually switched to the *SP* interface completely after 4 and respectively 5 levels. One of them mentioned “I liked that [*ML*] mode more but switched due to the cold and eventually my hand felt more relaxed”. In the lab condition one participant used the *SP* interface exclusively as it was “more comfortable” and “not as shaky” as the *ML* interface. If tracking recovery did not work as expected or took too long participants tended to switch to the *SP* interface.

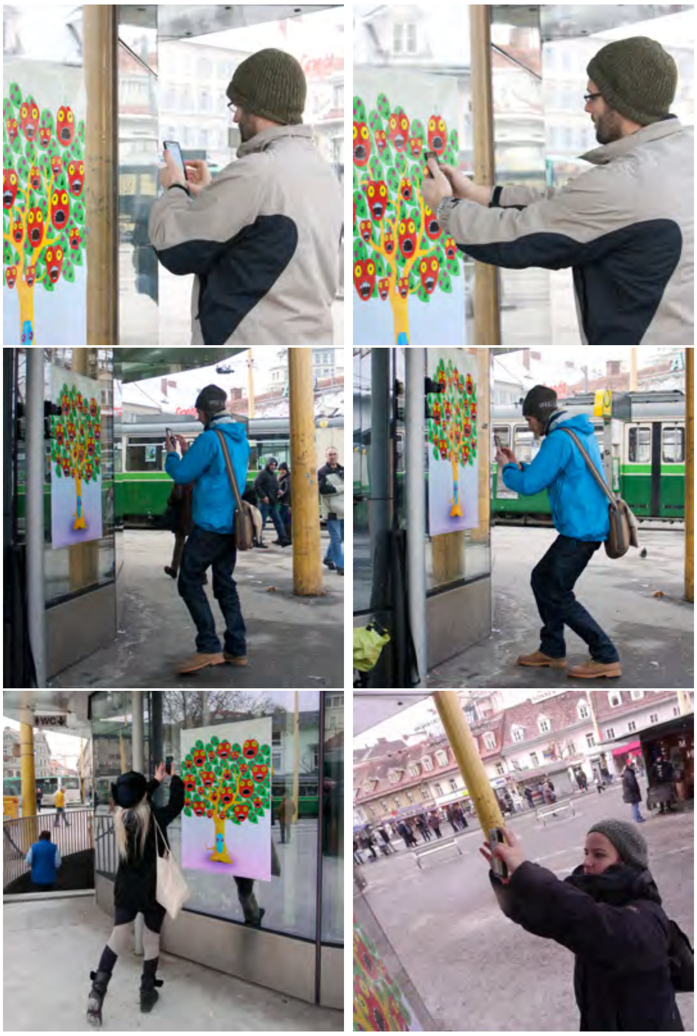


Figure 4.8: Participant using solely his arms to move back and forth (top row), bending knees to hit a target at the lower half of the poster (middle row), holding the phone above the head to reach targets at the top of the poster (bottom row).

One participant who switched back and forth between the interfaces said: “I wanted to use that [*ML*] mode but the system [tracking] did not work so I eventually switched to the other [*SP*] mode and tried again later”. Six participants switched back to the *ML* interface after playing one level of the game in *SP* latest. Two participants used *ML* as overview *SP* for quickly zooming in and two tried the *SP* mode to see whether they could be as fast as in *ML* mode.

4.1.3.4 Reactions from the public

We observed reactions from 691 people, who passed by in a half circle of ca. 10 meters around the poster. Approximately every 5 minutes a larger group of 5-10 people simultaneously passed by to change lines. The majority of the passersby did not notice the

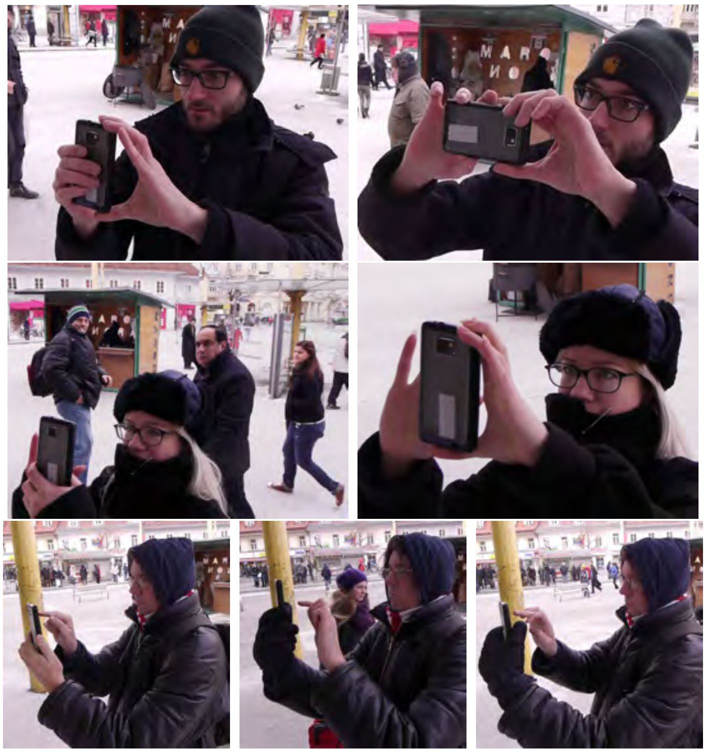


Figure 4.9: Various ways to hold the phone in the *ML* condition: Switching from portrait to landscape mode (top row), holding the phone across the short or long edge (middle row), using gloves to cope with the cold (bottom row).

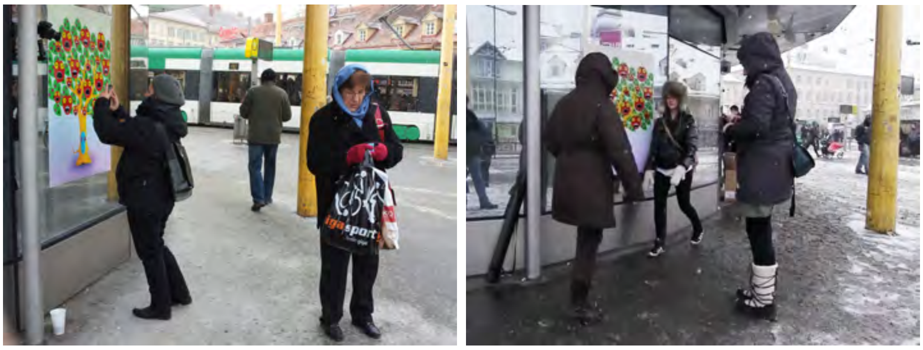


Figure 4.10: Passersby not noticing the participants interacting with *ML* (left) and *SP* interfaces (right).

participants, the poster or the recording equipment at all (68%).

Thirty percent of the passersby had short glimpses of less than a second and kept on walking (Figure 4.11, a). It was not possible to differentiate between the reasons for glimpsing, i.e., whether people looked primarily at the poster, the participant interacting or the wall mounted camera.

Ten people (1.5%) stopped and watched for more than 5 seconds (Figure 4.11, middle

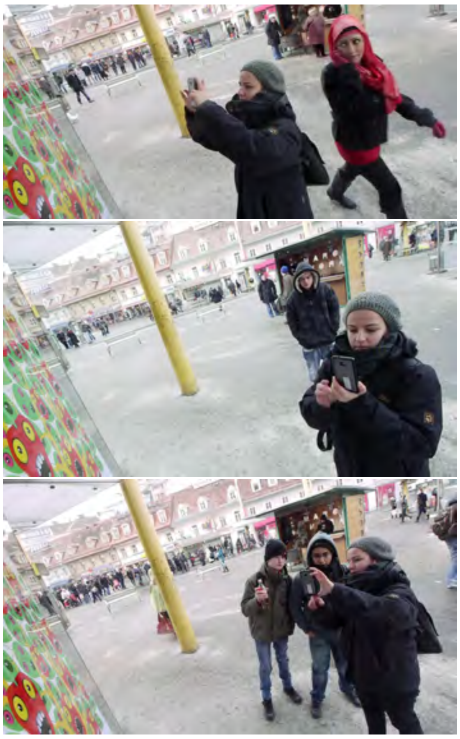


Figure 4.11: Passersby glimpsing (top row), watching from a distance (middle row) and approaching a participant (bottom row).

row). In three occasions (0.5%) participants were approached (by one elderly adult, one young adult, group of two boys) and asked what they were doing at the poster. In one occasion the participant explained the game to the children (Figure 4.111, bottom row).

4.1.3.5 Detachment from the environment

The ratings of following items indicated that participants concentrated on the system and tasks (see Figure 4.12) and did not focus on their environment:

q1: *I concentrated on the system.*

q2: *The tasks took most of my attention.*

Participants also indicated that the environment did not distract them much by rating following items:

q3: *It was hard to concentrate on some targets as I was distracted with the environment.*

q6: *I did not pay attention to the environment when using the ML interface.*

q13: *I felt nervous while using the system.*

In addition, a Mann-Whitney U test indicated significant differences between the public location and lab group for questionnaire items listed in Table 4.2. The ratings to the first two items might indicate that even though participants in the public condition were aware of their different role in the environment they did not care about the actions of the surrounding audience. This is also reflected in participants' comments stating that they

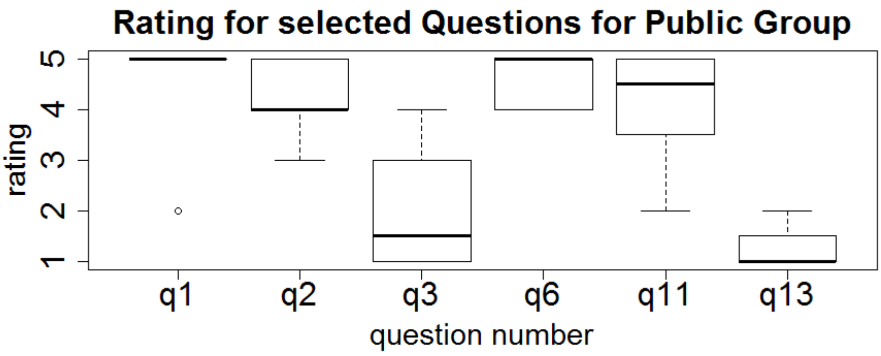


Figure 4.12: Ratings for selected questions concerning concentration on system and task and distraction by environment (5-point Likert scale, 1: totally disagree, 5: totally agree)

Questionnaire item	Result	p-value	Z-score
I did not pay attention to the environment when using the *ML* view. (P: MD=5, L: MD=4)	L<P	0.042	-1.72
I was aware that I had a different role in being there than most people in the environment. (P: MD=5, L: MD=4)	L<P	0.002	-2.91
I would rather do the task with the *ML* view only (P: MD=5, L: MD=3)	L<P	0.039	-1.77
I had to look away from the screen to perform the task (P: MD=1, L: MD=2)	L>P	0.013	2.24
How did you feel using the system in the environment? Cold . . . Warm (P: MD=2, L: MD=4)	L>P	<.0001	3.24
How did you feel using the system in the environment? Insensitive . . . Sensitive (P: MD=4, L: MD=2, 3)	L<P	0.035	-1.81

Table 4.2: Questionnaire items that were rated significantly different between the public location (P) and lab (L) group.

knew people were around but they did not care about it. The significant lower ratings to the social presence questionnaire item in the last two rows might eventually highlight the fact that users in the public condition played the game in a low temperature environment.

During the interviews one participant described the gaming experience as “asocial”. She felt “totally focused on the game” and did not pay attention to passersby at all as she “did not care about anything else”. Another comment was: “The people watch and see that you are doing something – but actually you are completely passive to your environment”

	Public	**Lab**
ML	M=50.2 SD=22.3	M=58.5 SD=22.6
SP	M=43.3 SD=10.3	M=43.0 SD=11.1

Table 4.3: *TCTs* (s) over 4 levels in performance phase.

	Public	**Lab**
ML	M=0.31 SD=0.53	M=0.78 SD=1.18
SP	M=0.38 SD=0.71	M=0.31 SD=0.64

Table 4.4: Selection errors over 4 levels in performance phase.

4.1.3.6 No Significant differences in performance between lab and public group

We included an experimental phase similar to the one described in [111]. We wanted to investigate possible effects of location and audience on task performance. This separate phase was conducted as participants had the free choice on interface usage in the main game phase. The main game phase was not used to analyze task performance. Mann-Whitney U tests indicated no significant differences between the groups for the TCTs or error rates in *ML* and *SP* mode (see Table 4.3 and Table 4.3).

4.1.3.7 Using the interfaces outside of the study setting

Only half of the participants at the public location indicated that they would use the *ML* interface outside of the study setting at a public transportation stop (see Figure 4.13). Figure 4.14 shows in front of which audience the participant would use the interfaces. The questionnaire is similar to the one employed in [197]. According to pairwise Chi-Squared tests of independence there were no significant differences between groups for location or audience ratings. The public group would use the *SP* interface in public transportation significantly more often (χ^2=6.25, p=0.012).

While Yate's continuity correction was applied for addressing the low expected cell count the sample size of 16 items in a 2x2 table should be taken into account when interpreting these results. During the interviews participants further explained their decisions and two mentioned that they would use the interface specifically with friends around.

4.1.4 Discussion

The study demonstrated that, contrary to our expectations, the *ML* interface was used in the field most of the time; with only few significant differences when compared to laboratory usage. The use of the *ML* interface was at least partly driven by curiosity as most participants were already familiar with the *SP* interface and perceived the study as an opportunity to "try out" a new mobile *AR* game. The novelty of the interface was also indicated by the diverse ways participants handled the phone.

The *SP* interface was mainly used when the system could not recover from tracking errors fast enough or when participants did not want to move closer to the poster but rather

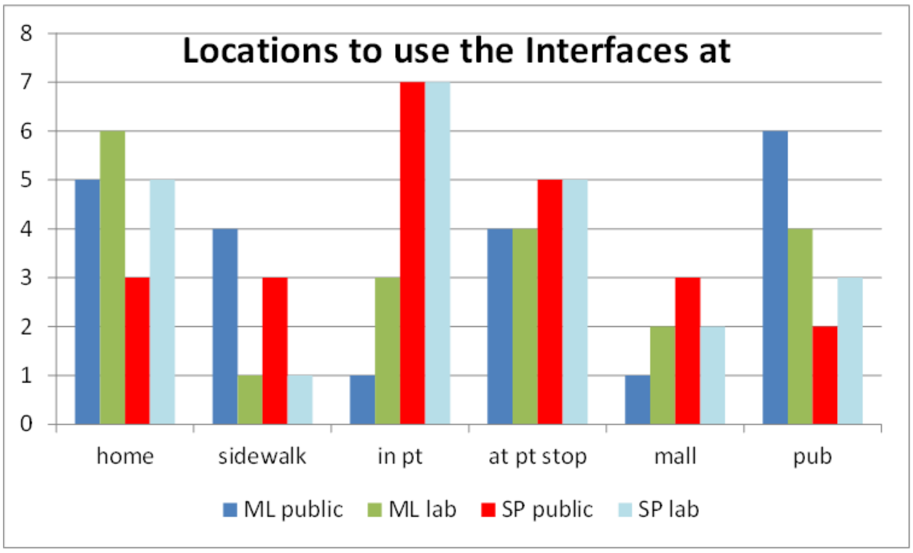


Figure 4.13: Number of participants who would use the interfaces at various locations (pt: public transportation.

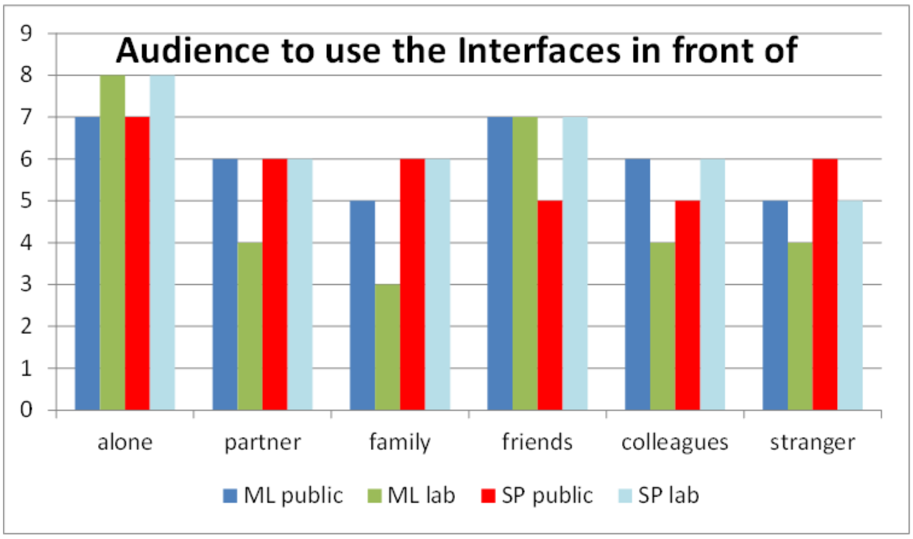


Figure 4.14: Number of participants who would use the interfaces in front of various audiences.

zoomed in to hit the small targets. The levels did not increase in difficulty to ensure we could study fatigue and learning effects. However, we could not uniquely identify individual causes for changing user behavior (especially hand poses). This might be partly due to reoccurring tracking errors being a confounding factor in this study and needs further consideration. Contrary to previous studies about the use of *ML* and *SP* interfaces in handheld systems outside the laboratory [165] we used a game design that demanded the attention of single users and had no collaborative aspects. In this study we found no significant effects of location and audience on user behavior and task performance.

Participants concentrated mostly on the game task and did not pay attention to passersby and activities going on around them in the street. This finding is supported by other studies where for example, mobile *AR* users bump into lampposts while engrossed in the screen interface [166] and is a commonly identified problem with pedestrians using their mobile phones and walking out into traffic.

While the *ML* interface was used by participants most of the time during the study, only half of them indicated they would continue to do so if they had the opportunity to play a game at an augmented poster at a public transport stop in the future. However, the indicated non-game-playing attitudes of the participants need to be taken into account when considering these responses.

Despite confounding factors such as a public space, cold weather and a repetitive task, most users continued to use the *ML* interface. While the two interfaces were designed to balance the achievable performance and ease the mental gap when switching views, participants' comments indicated that the game could just be less engaging in the *SP* interface. Overall, the fact that the *ML* interface was used for an overwhelming percentage of the interaction time requires more exploration.

The majority of passersby did not pay attention to the participants interacting with the poster; if they did then only for a short period of time. As one participant mentioned playing the game in *ML* mode "is comparable of hearing loud music in public transportation. . . . If users do not care about that they might probably also use this [*ML*] interface". Our observations within this study indicate that for a public transit place interacting with a *ML* interface does to a large extent not result in socially conspicuous behavior. The observations complement recent online surveys that indicate a small but growing number of users adapting to novel interactive systems, such as QR code equipped products[1] or mobile *AR* browsers [86, 176] through their smartphones in public spaces.

An open question concerning well-designed augmented posters might be: would people continue to use the *ML* interface once they become familiar and the novelty has worn off? Our study indicates that at least reactions from the public might not inhibit the initiation of *ML* usage. Furthermore, "Playing with friends" was a motivating factor mentioned in the interviews to use a *ML* interface in public when participants would not use that interface alone. Therefore, enabling group activities on augmented posters might lower the barrier for initiating interactions with the *ML* interface further.

4.2 Repeated evaluation of ML and SP interfaces in public space

In the previously presented study we conducted an evaluation on the usage of *ML* and *SP* interfaces for playing a find-and-select game in a public transport area and reported on the reactions of the public audience to participants' interactions [87].

To further the investigate the potential effects of space and place on interaction we repeated the experiment at another public transportation stop in Vienna during two days in July 2012 (see Figure 4.15). The study design, procedure, and evaluation tools were

[1] http://econsultancy.com/uk/blog/8118-19-of-uk-consumers-have-scanned-a-qr-code-survey, last retrieved 20.04.2015.

reproduced (the camera location for recording participants and environment had to be adapted). Ten volunteers (5 females, 5 males, aged 19 to 37) participated in the study. They were acquired through social media channels and a social media company. Participants were locals from Vienna who used this transportation stop before to get to a nearby popular club. They received a small gift for participating.

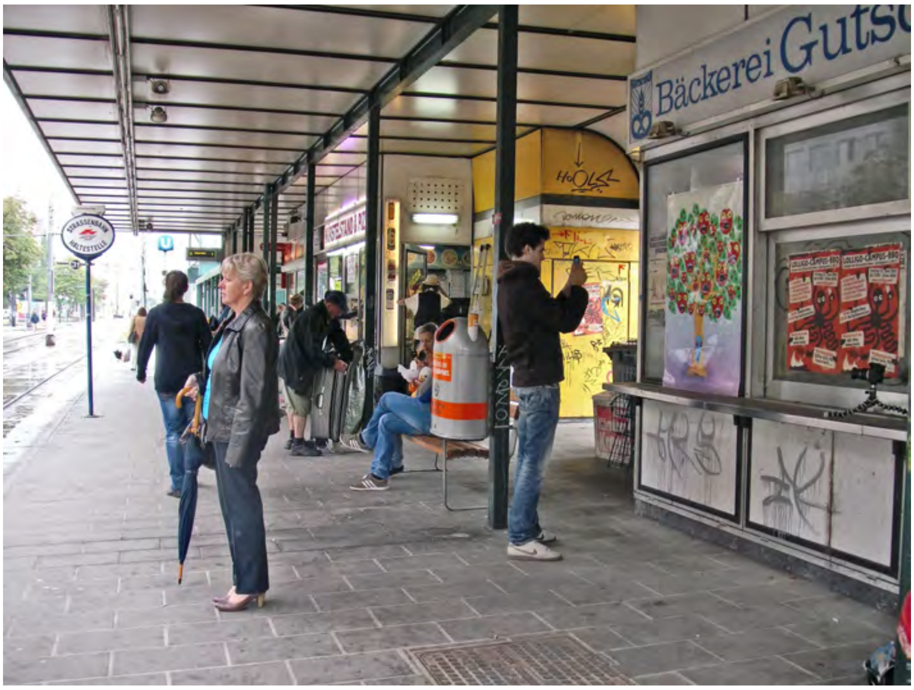


Figure 4.15: The location of the repeated study was a public transportation stop in Vienna, Austria.

4.2.1 Findings

We report on our findings of the public condition in Vienna (PUV) and relate them to previous findings of the laboratory (LAB) and public (PUG) condition in Graz. For the between-subjects design (with location as factor with three levels), the collected data was not normal distributed (and could not be transformed to normal distributed data), thus we employed non-parametric null hypothesis testing. Two participants solely used the *SP* interface in PUV (one in LAB). Note, while we report based on data from all participants, also with those participants treated as outliers there were still main effects of location on the reported dependent variables.

4.2.1.1 SP was used most of the time

In the PUV condition the *SP* interface was used 56% of the time (over all participants) (PUG: 24%, LAB: 32%). A Kruskal-Wallis rank sum test indicated a main effect of location on usage duration (χ^2=26.72, p=1.5e-6). Pairwise comparisons using Wilcoxon

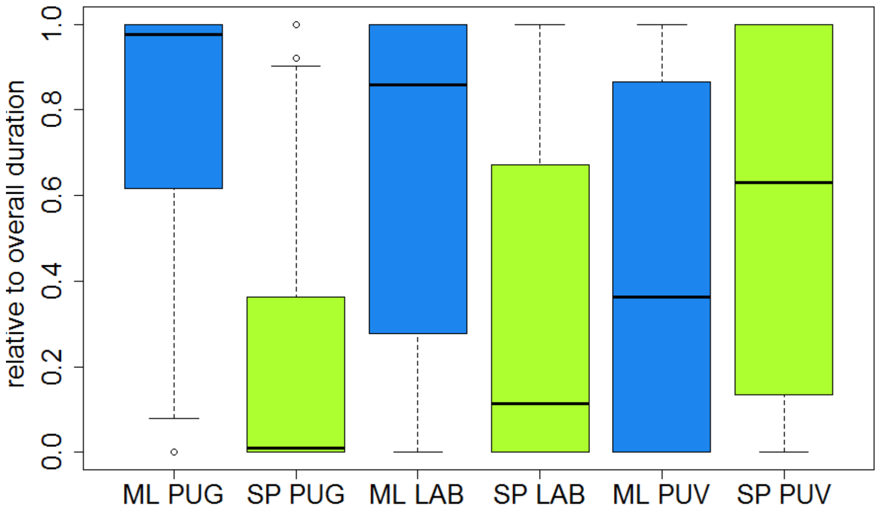


Figure 4.16: Relative usage durations for the *ML* (blue) and *SP* (green) interface for PUG, LAB and PUV

rank sum test with Bonferroni correction indicated that *ML* was used significantly less in PUV (MD=0.36) compared to both PUG (MD=0.98, p=3e-6, r=.41) and LAB (MD=0.86, p=.001, r=.29) (see also Figure 4.16).

4.2.1.2 Preference

A Kruskal-Wallis rank sum test indicated a main effect of location on enjoyment (χ^2=6.24, p=.04, “I enjoyed using the *ML* view in the environment”). Pairwise comparisons using Wilcoxon rank sum test with Bonferroni correction indicated that the rating was significantly lower in PUV (MD=3.5 on a 1 to 5 Likert scale) compared to PUG (MD=5, p=.027, r=.57).

4.2.1.3 Tracking errors

Tracking errors result in a loss of augmentation and do occur even with state of the art tracking systems. In all conditions they occurred throughout the usage of the system. In total 147 (PUG 105, LAB 162) tracking errors occurred (14% of the overall gaming duration in *ML* mode, PUG 9%, LAB 7%). A Kruskal-Wallis rank sum test indicated no main effect of the location on the number of tracking errors per minute or per level but on the duration of tracking errors (χ^2=45.96, p=1e-10). Pairwise comparisons using Wilcoxon rank sum test with Bonferroni correction indicated that the durations of tracking errors were significantly different in PUV (MD=1.6 seconds) compared to PUG (MD=2.3, p=5e-4, r=.88) and LAB (MD=1.3, p=3e-11, r=.79).

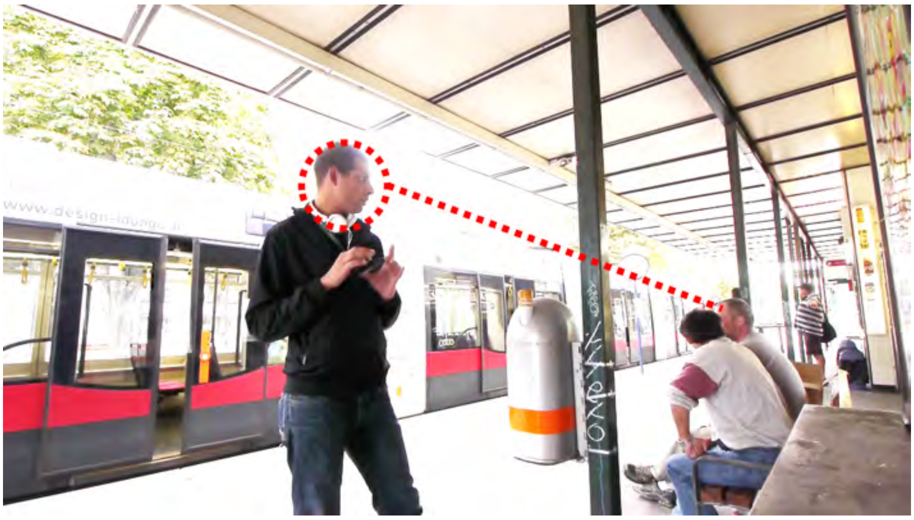


Figure 4.17: Participant looking at two drunken men sitting on a nearby bench, who are chatting and watching the scene.

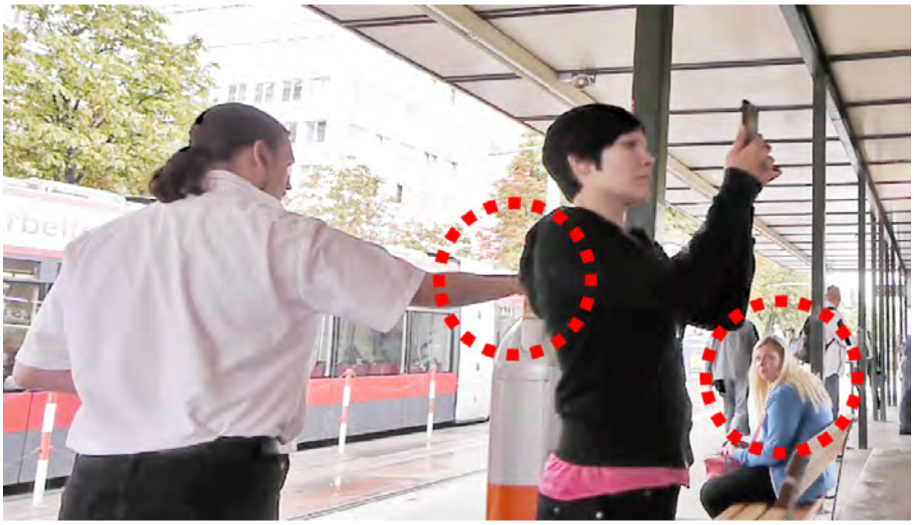


Figure 4.18: Passer-by intruding the personal space of a participant, who is also watched by a woman sitting on a nearby bench.

4.2.1.4 Interactions with passers-by

In the PUG condition only few interactions between passers-by and participants were observed. In PUV 241 (PUG 641) passers-by interactions were identified through video-based interaction analysis (using open and axial coding). From those, 50% (PUG 68%) were related to passers-by not noticing the participants or the recording setup. Twenty-two percent (PUG 30%) and 9% (PUG 2%) were staying and watching the participants actions for more than 5 seconds. However, in contrast to PUG, 15% of the interactions related to intrusions of the *social* space, as proposed by Hall [95] (see Figure 4.17), and

5% intrusion of the *personal space* (see Figure 4.18) of participants in PUV.

4.2.1.5 Reasons for using ML and SP

Two participants used the *SP* interface exclusively in the main gaming phase in PUV (one in LAB). For one of those participants two men were sitting in the social space around him and were talking to each other during the whole duration of the game (see Figure 4.17). During the post-hoc interview he mentioned “this is not the prettiest and calmest environment” but then also mentioned that the *ML* mode is “a bit troublesome” due to the tracking errors. For the other participant a group of 6 men were standing in the public space (˜10m away) and watching her while she was turning her back on them. However, in the post-hoc interview she argued that she found the *ML* interface “a bit more cumbersome . . . perhaps due to my height” [in relation to the poster]) not specifically mentioning the social context. Participants mentioned similar reasons for using the *ML* interface in PUV as in PUG and LAB conditions [87]. In addition, we employed a questionnaire similar to Rico and Brewster [197] to ask participants about locations in which they would use the interfaces.

No significant differences compared to the previous conditions were found. Also no coherent correlations between usage duration of the *ML* interface or preference could be identified.

4.2.2 Discussion

The findings of the original study presented in [87] showed only minor differences in participants’ usage of the *ML* and *SP* interfaces for playing a find-and-select game between a public space and a laboratory setting. Both usage duration and users ratings indicated that participants preferred the use of the *ML* interface. However, this repeated study indicated significant different results both for usage durations and for preference rating compared to previous conditions, also when the participants who used *SP* exclusively were treated as outliers.

One challenge in conducting field studies is the potentially large number of confounding factors which can influence the evaluation outcomes. This impedes reliably identifying cause and effect relations for the outcomes of this study compared to previous conditions. However, while both public locations were transportation areas there were noticeable spatial and social differences between the two public locations which could have effects on the participants’ behavior.

The PUG condition was carried out a location primarily used as *transit area* for changing tram lines with the major waiting areas being more than 20m away (see Figure 4.19). It was located in a wide open space in the city center. People from all social contexts are using this place for changing trams. The general area is under video surveillance and the building at which the study took place was actively operated by the local tram company. It was a place with a high frequency of passers-by coming from several directions but only a few people were standing in the social space of the participants (rather walking behind the participants, see Figure 4.19).

In contrast the location in the PUV condition was primarily used as *waiting* area for

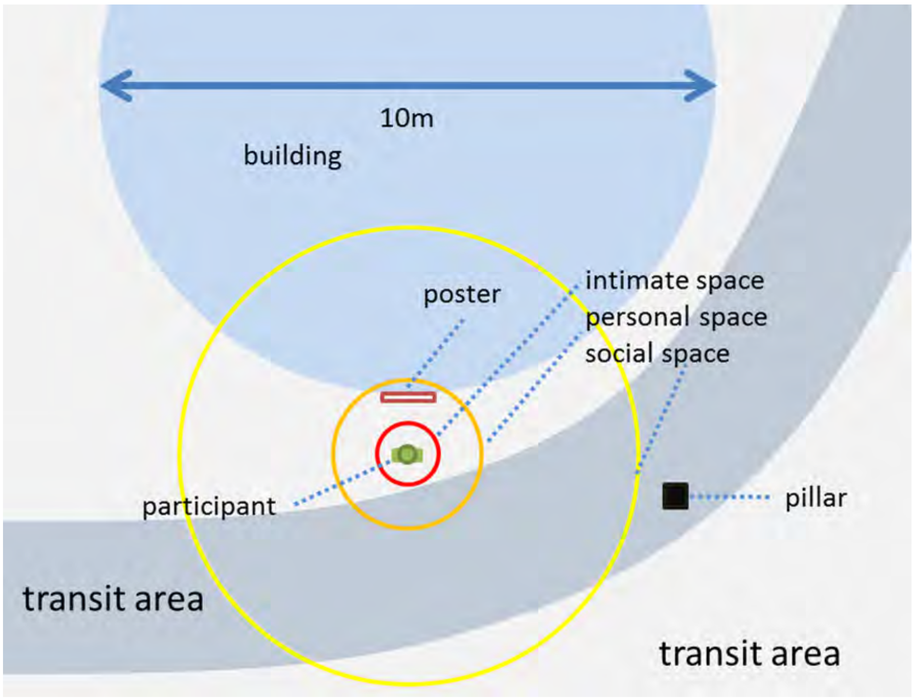


Figure 4.19: Schematic top down view of the space in the PUG condition with Hall's reaction bubbles indicating the intimate (0.5 m), personal (1.2m) and social space (3.6m) of the participants.

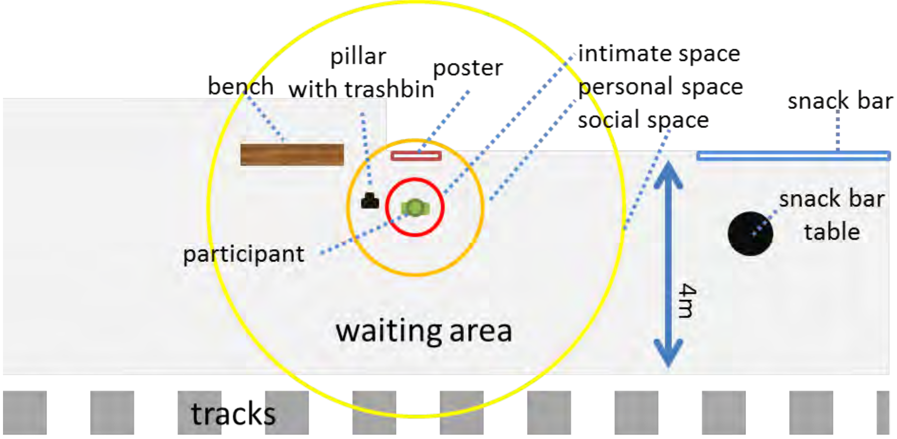


Figure 4.20: Schematic top down view of the space in the PUG condition with Hall's reaction bubbles indicating the intimate (0.5 m), personal (1.2m) and social space (3.6m) of the participants.

people coming from the exit of a near-by metro line (see Figure 4.20). It was located in a disadvantaged area (Vienna Leopoldstadt). Comments of participants about the "shabby" area and experimenter's observations indicate that there might have been a larger *social distance* between participants (mostly middleclass, students) and people with lower socioeconomic status present at the tram stop compared to PUG. In addition, while the number of people identified during the video analysis (in a similar timeframe) was 2.5

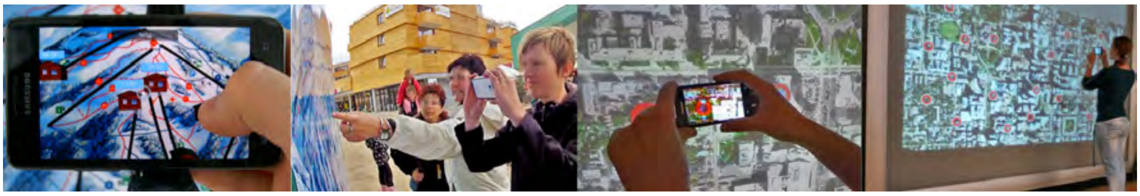


Figure 4.21: *ML* interface used in the semi-controlled field study (column 1), participant interacting with the interface during the semi-controlled field study (column 2), *ML* interface used in the lab study (column 3) and participant interacting with a large map used in the lab study (column 4).

times less than in the PUG condition a larger amount of people were intruding the social space of participants for longer periods of time. Specifically, in the PUV condition 9 of 10 participants could see people sitting on a nearby bench (˜2m away) in their periphery (and those people could watch them - see also Figures 3, 4, and 6).

Those differences between the locations could indicate that the social context in PUV could have inhibited the use of expressive, socially not common spatial gestures used in the *ML* interface, which is supported by the observations in Akpan et al. [2]. Still, there are other potential factors which could have influenced participant's behavior and ratings. They include fatigue, the perceived severity of tracking errors, the role of personality (e.g., intro- and extraversion), intrinsic motivation to use the interfaces, the novelty of the *ML* metaphor and demand characteristics.

We repeated a study on the usage of a *ML* and a *SP* interface for playing a find-and-select game at a public transportation stop. While the study design and procedure were reproduced the spatial characteristics and social context of the location of the study differed from the previous public condition. Significant differences both for the usage duration and the preference were found compared to previous runs of the experiment. Specifically, the *ML* interface was used significantly less and preferred less compared to a previous public condition. Qualitative data analysis indicated that the social context could have influenced the choice of interfaces.

Still, further repetitions of the study should be conducted to better understand individual factors which influence *AR* interaction in the public and interrelations between them. We also need more reliable combinations of evaluation methods targeting the study of expressive interactive systems in-the-wild.

4.3 The utility of ML interfaces on handheld devices for touristic map navigation

Tourists who visit cities or resorts, which they are not familiar with, often use maps as tools to orient, explore and navigate these unknown physical environments. Digital maps on handheld devices, such as smartphones, make location-based services accessible and are popular tools to support touristic needs in these contexts. The dominant technique to interact with these digital maps on devices with touch screens is *SP* navigation using tap-n-drag and pinch-to-zoom as used for example in Google maps. Still, physical maps continue to play an important role in the tourism sector. They address navigational needs of users if there is no data connection, but can also highlight specific *POIs* selected by local tourism associations which might not be easily accessible through general purpose map applications like Google Maps. Furthermore, large physical maps can provide more information at a glance (i.e., a larger information space size) than small screens but lack the advantage of dynamic adaptation and personalization common in digital maps. However, in a touristic place, large physical maps might facilitate the communication between groups of friends or between family members by being a common ground for the discussion [166]. For example, if someone is pointing at a specific location on the physical map, the rest of the group can immediately be aware of the pointed location.

Mobile *AR* applications have the potential to overcome both the static nature of the information on physical maps and the small screen constraints of mobile digital maps through the *ML* metaphor. Recently, *ML* interfaces became popular as an interface for browsing the physical world in location-based applications [140] through Augmented Reality browsers like Junaio, Wikitude or Nokia City Lens. Augmented Reality browsers typically combine a *ML*, a *SP* and a list view for geo-referenced information [76] in the vicinity of the user. Also, *ML* interfaces have become popular with leisure-oriented activities in gaming and advertising [241] often relying on commercially available computer vision-based tracking systems provided by companies like Metaio or Qualcomm.

For non-leisure activities such as information browsing and navigation on (physical) maps the benefits and drawbacks of the individual interfaces are not yet thoroughly understood. Specifically, to the best of our knowledge there are no recent studies investigating the performance of *ML* and *SP* map navigation using state of the art tracking technologies which allow for a wide interaction space and today's popular interaction methods like tap-n-drag for *SP* navigation. A better understanding of the potentials and pitfalls of these interaction methods, given current technology, is critical both for the designers of mobile interfaces as well as business stakeholders considering investments in the provision of novel services for tourists. Specifically, business stakeholders in the tourism domain need assessments on which user interface provides most value to users and, in turn, lead to user "click-through" as well as actual purchase or reservations. For example if *ML* interaction can deliver added-value in terms of performance or user experience in touristic contexts the effort of enabling *ML* interaction with large scale physical posters might be worthwhile. This effort consists among others of authoring 3D models, embedding video streams in a visually compelling way, and designing physical posters and maps specifically with the visual integration of disparate digital content in mind [83]. In contrast, if benefits of *ML* and *SP* interaction are comparable there is no need for added expenses and traditional location-based solutions, making small screen interaction sufficient to address the needs of tourists.

Our work therefore makes following contributions: First, we provide an up to date comparison of *ML* and *SP* navigation for a generic information browsing task on maps. We conducted a semi-controlled field experiment on a public map at the ski slopes of a tourist hotspot in the Austrian Alps. Our comparison revealed that even with state-of-the-art vision-based tracking *ML* is significant slower than a conventional *SP* interface with tap-n-drag for a common map size in public spaces. *ML* also does not perform better in terms of error rate and user experience. In addition, our study did not reveal significant effects of *ML* interaction on the audience or an effect of the setting on the user experience rating of participants. Second, looking deeper into workspace size in a separate laboratory study, we could not see *ML* outperform *SP*, achieving at most equal performance with an increasing size of the map. But, *ML* significantly decreases demand and increases usability compared to *SP*. Third, we reflect on the implications of our findings for *ML* interaction in touristic scenarios.

4.3.1 Semi-controlled field experiment

The municipality of Schladming is a key skiing resort in the Austrian Alps interested in innovating in ubiquitous and mobile technologies to further develop their tourism sector. As a decision aid for selecting technologies potentially driving tourism Schladming wants to get quantitative assessments if mobile *AR* interfaces could add value for visitors. Specifically, they were interested to find out if mobile *AR* could provide added value to existing physical information infrastructure such as panorama posters and provide benefits over pure digital interfaces like Google Maps. Panorama posters of varying sizes showing positions of slopes and local businesses such as restaurants are widely distributed in the area (see Figure 4.22, right) and are a significant cost factor in advertisement budgets in the region. We chose them as reference information space for a study.

4.3.1.1 Study design and task

To address the question if *ML* interfaces could provide benefits over *SP* interfaces for map navigation we designed a within-subjects study comparing the effects those interfaces on user performance and user experience. We chose a semi-controlled field experiment design combining quantitative performance measures (TCT, error rate) with subjective feedback (user experience questionnaires, semi-structured interviews) and observations (video recording and field notes).

One challenge in the design of our experiment (and other field experiments, cf. [37]) is to balance ecological validity (task relevance using an artifact relevant to people in the area) with a sound comparative study design and external validity (being able to transfer results out of the scope of the specific setup). Hence, similar to previous studies we chose a locator task [201]. Locator tasks are high level geoinformation tasks which typically require location search for an object with certain target attributes [191].

The specific task was to find the single lowest priced restaurant among 16 candidate locations on a ski field map. The candidate locations referred to existing restaurants in the region. A tap on the restaurant revealed the price (randomly generated for each location), a long press selected the restaurant (screen size 19x14mm in 60cm distance) as candidate. An icon on the physical map represented the restaurant; a corresponding bitmap was representing the restaurant on the mobile. Interface was the within-participants factor with two levels: *ML* and *SP*. The task had to be conducted 4 times per interface resulting in 2x4x16=128 locations (and 8 final targets) per participant. The starting order of the interface was counterbalanced but the tasks were blocked by interface.

Finding the lowest price could also easily be solved with a list view. However, we decided not to include a list view in our comparison to be able to generalize our findings to other locator tasks. General locator tasks can encompass non quantifiable attributes such as textual opinions of users which cannot automatically be sorted.

4.3.1.2 Apparatus and location

The *ML* and *SP* interfaces were implemented in Unity 3D and deployed on a Samsung Galaxy SII smartphone. In landscape mode the smartphone screen has a physical extend of 9.32x5.6 cm with a resolution of 800x480 px (218 ppi). The physical camera has a vertical angle of view of 35 ° (horizontal 49 °). The physical camera parameters with a vertical *FoV* of 35 ° were matched accordingly with the virtual camera used in *SP*. Also the size of the virtual poster was matched with the size of the physical poster. Thus *ML* and *SP* interaction operated in coordinate systems with the same metric units. The translation of the *SP* camera parallel to the poster was determined by $dx = touchDelta_x * TF_x * pz$ for the x direction (equivalent for y) where dx is the translation delta in world space, $touchDelta_x$ is the drag distance in screen space (px), pz is the current distance of the *SP* camera to the poster in world space, and TF_x is a factor to scale the x (and y) translation dependent on the camera distance. TF_x was empirically set to 0.0076 in order to imitate the panning experience with Google Maps on the test device. Similarly, translation along the principal ray of the camera was implemented. The employed tracking system was the Qualcomm Vuforia *SDK*[2] which is used in many commercial available *AR* applications and provides state of the art computer vision-based tracking performance.

The poster had an extend of 154x84 cm with its center position mounted at a height of 160 cm directly at a facade of the tourism office next to the ski slopes. The text depicting the labels used the Arial font with a text size of 21 pt on the poster (10 pt screen size in 60 cm viewing distance) using white color on light blue background which supports quick scanning of text. The 16 hut objects used as buttons had a size of 4x3 cm on the poster (or 2.5x3.5 % of the whole information space). The average candidate density of the poster (number of huts divided by poster size) was 12.2 items per m^2 and was derived through suggestions of previous studies [201].

The study was conducted outside the central tourism office and ski station, during the ski season in March 2013, shortly after the FIS 2013 ski world championship in Schladming, Austria, see Figure 4.22, left (including proxemic zones defined by Hall [95]) and right. Approximately 20 m away in front of the same building another bigger map was mounted with an extend of 2.3x1.3m but could not be used for the study for technical reasons.

4.3.1.3 Data collection

We collected device, video and survey data complemented with photos and notes. A GoPro camera was mounted at a height of 2 m next to the poster behind a pillar to observe both users' actions as well as reaction of passers-by. The video analysis was conducted by a single trained person. The categories for video analysis were derived from related studies [87][89].

The interaction of the users with the interfaces was logged on the device. Specifically, the motion of the physical camera (*ML*) and virtual camera was sampled at 10 Hz. We em-

[2] http://www.qualcomm.com/solutions/augmented-reality, last retrieved 20.04.2015.

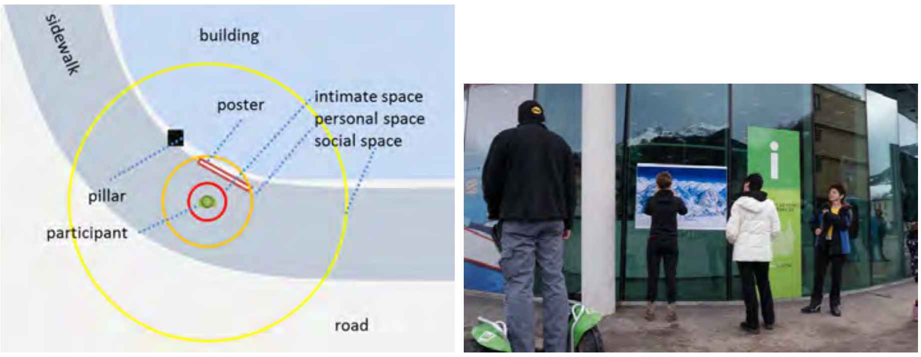


Figure 4.22: Schematic top down view of the study location with Hall's reaction bubbles indicating the intimate (0.5 m), personal (1.2 m) and social space (3.6 m) of the participants (left) and photograph of the location with one participant and three passers-by (right).

ployed the AttrakDiff questionnaire [106] for capturing hedonic (stimulation, identity) and pragmatic user experience dimensions, the Interest/Enjoyment (IE) and Value/Usefulness (VU) sub-scales of the intrinsic motivation inventory [159] and an environmental distraction measure [87]. We analyzed quantitative data with the R statistical package, IBM SPSS and Microsoft Excel. *NHST* was carried out at the 0.05 level. For the device data we excluded samples with more than 3 standard deviations away from the mean. For data which was not normal distributed and could not be transformed to normal distribution we employed non-parametric *NHST*. The reported confidence intervals were computed at a 95% level. With 17 participants, 2 interfaces and 4 trials per method 136 trials were recorded.

4.3.1.4 Hypotheses

Due to the findings of previous studies [201] we suspected that *ML* interaction will outperform *SP* interaction in terms of *TCT*. Hence, our first hypothesis was: *H1: Task completion time of ML will lower than for SP*. Similar, previous studies indicated that users rated *ML* interaction more favorably than alternative interfaces for pointing and navigation tasks. Hence, our second hypothesis was: *H2: Users will prefer ML over SP interaction.*

4.3.1.5 Procedure

At the beginning of the study each participant was introduced to the general schedule of the experiment (after signing an informed consent form) and the setting of the task (i.e., looking for the best price for a meal at lunchtime). They then filled in a background questionnaire (demographics, technical skills). Afterward, the participants were introduced to both interfaces and were free to train both interfaces in a learning phase (with different symbols and locations used than in the main task). The learning phase was typically around 3-5 minutes and was stopped when participants felt comfortable in operating both interfaces. Afterwards, they were informed about the main task. They were specifically asked to conduct the task as fast and accurate as possible. The task was conducted for the first interface in 4 repetitions and participants were allowed to make a break between each repetition but nobody did so. The overall time per interface was around 5 minutes. After the first task block was completed participants filled out intermediate questionnaires (AttrakDiff, intrinsic motivation inventory, environmental distraction) which took around 5 minutes. The main task and rating procedure was then repeated for the second interface (order was counterbalanced). In the end the participant filled out a second background questionnaire and received a gift worth 10 Euros. The overall duration per participant was 30-40 minutes.

4.3.1.6 Participants

Eighteen volunteers from around the area participated in the study (12 males, 6 females) with 10 being between 18 and 34 years old (1<18, 4, 35-54, 3>55). All but one were right handed. The average height of participants was 175 cm high (sd=9 cm). All participants were smartphone users, none had experience with *ML*. On a five point Likert scale (never, . . . , very often) four participants indicated that they use mobile map applications very often, (7: often, 4: occasionally, 3: rarely). Two participants indicated to use physical maps during vacation very often (6: often, 7: occasionally, 3: rarely). We had to exclude data from one participant from the analysis due to logging problems.

4.3.1.7 Findings

In this section we report on the performance-based measures *TCT* and selection errors. We will also investigate motion patterns of the handheld device, gaze switches, user experience measures and audience behavior.

Task completion time The *TCT* of *ML* was significant slower compared to the *TCT* of *SP*. A two-tailed paired t-test indicated a significant effect of interface ($t_{(67)}$=5.34, $p<0.05$, Cohen's d=0.6) on *TCT* with *ML* having a higher *TCT* (M=44.8s, σ=15.8) than *SP* (M=34.9s, σ=11.2). In addition, effects of played levels (1-4th repetition) on *TCT* were investigated. A one-way repeated-measure ANOVA indicated a significant effect of level on *TCT* (*ML*: F(3,48)=10.92, $p<0.01$, η^2=0.14, *SP*: F(3,48)=4.09, p=0.01, η^2=0.07). The achieved power (1-β) for *TCT* (*TCT*) was > 0.99. For *ML* post-hoc pairwise two-tailed t-tests with Bonferroni correction revealed significant differences between level 1 and all other levels (see Table 4.5). For *SP*, post-hoc pairwise two-tailed t-tests with Bonferroni correction did not reveal significant differences between levels 1 and all other levels. However, a pairwise one-tailed t-test indicated significant differences between level 1 (MD: SD:) and 4 MD SD, p = 0.046, see Table 4.6. To investigate if learning effects potentially caused the significant effect of interface on overall *TCT* we re-ran the analysis with only levels 2-4, but still a two-tailed paired t-test indicated a significant effect of interface ($t_{(50)}$=5.34, $p<0.05$, Cohen's d=0.58) on *TCT* with *ML* having a higher *TCT* (M=41.39s, σ=13.6) than *SP* (M=33.4s, σ=11.3).

Errors We investigated the number of selection errors (if a user chose a wrong target) and the candidate coverage (how many candidates have been visited). A two-tailed

TCT **(second)**	**level**	**1**	**2**	**3**
55 (18)	1	-	-	-
44 (13)	2	**.003**	-	-
41 (15)	3	**.0004**	1.0	-
39 (13)	4	**.0007**	.20	1.0

Table 4.5: TCTs (in seconds, mean and σ) for individual levels in *ML* condition and p-values from post-hoc pairwise comparison (using one-tailed t-tests, bold values indicate sign. differences).

TCT **(second)**	**level**	**1**	**2**	**3**
40 (10)	1	-	-	-
34 (13)	2	.06	-	-
33 (9)	3	.07	1.0	-
33 (12)	4	**.046**	1.0	1.0

Table 4.6: TCTs (in seconds, mean and σ) for individual levels in *SP* condition and p-values from post-hoc pairwise comparison (using one-tailed t-tests, bold values indicate sign. differences)

Wilcoxon signed-rank test did not indicate a significant effect of interface (W=7.5, Z=0.36, p=0.72) on selection error. For both interfaces only 3 selection errors happened (with 17*4=68 targets). Similar, a two-tailed Wilcoxon signed-rank test did not indicate a significant effect of interface (W=4, Z=-1.28, p=0.2) on candidate coverage. From 1088 possible candidates (16 candidates x 4 levels x 17 participants) for *ML* 5 candidates (0.46 %) were not visited and for *SP* 9 candidates (0.82 %).

Motion Patterns We explored how participants moved the handheld device relative to the physical and virtual posters. While we designed the system so that certain interaction distances are likely (based on the predefined hut and text sizes) we did not give instructions to participants how they should position themselves towards the poster. While in *ML* participants naturally had to hold the phone towards the poster, in *SP* they generally kept the phone more parallel to the ground. Figure 4.23 depicts the locations of the camera center in the x-y plane of the poster. The camera positions in *ML* refer to the physical location of the device relative to the poster. For *SP* the camera positions refer to the location of the virtual camera and do not correlate to the physical pose of the smartphone. For example, user could actually turn away from the printed poster in *SP* and still complete the task. It indicates that participants more closely followed the spatial layout of the hut candidates (which form a U-like shape on the poster) in the *SP* condition (right) than in the *ML* condition (left). For visualization purposes every 20^{th} camera sample is shown in Figure 4.23 and 4 (0.5Hz). Over all participants (and all pose samples) the effect size of interface on the x position (px) was only small as indicated by two-tailed Wilcoxon rank sum test (W=379134424, Z=9.9, p<0.001, Cohen's d=0.08). However, two-tailed Wilcoxon rank sum tests indicated a large effect size for the y position (py) (U=577568605, Z=119.73, p<0.001, Cohen's d=1.19).

Kruskal-Wallis rank sum tests indicated significant effects of participant on camera movements in px (χ^2 (16) = 10385.58, p<0.01), py (χ^2 (16) = 19254.78, p<0.01) and pz

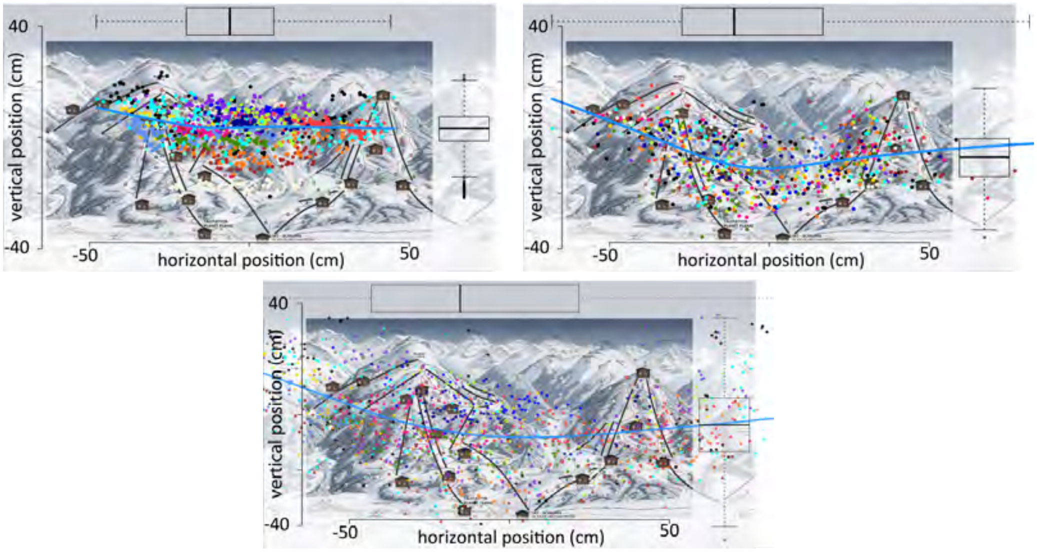


Figure 4.23: Position in the x-y plane. *ML* camera centers: top row, left. *SP* camera center: top row, right. *ML* projected camera centers in poster plane: bottom row. Dots with unique colors represent individual participants. The LOWESS (locally weighted scatterplot smoothing) curve is shown in blue. Every 20^{th} camera sample is shown.

	px (projected px)		**py (projected py)**		**pz**	
	# med. d	**# large d**	**# med. d**	**# large d**	**# med. d**	**# large d**
ML	16 (14)	37 (22)	6 (3)	28 (16)	5	60
SP	7	0	5	0	19	24

Table 4.7: Number of significant ($p < 0.05$) pairwise differences (from 136 possible) with medium (Cohen's d = [0.5, 0.8]) and large (Cohen's d >= 0.8) effect sizes for motion in x, y and z plane of the poster.

(χ^2 (16) = 20325.38, p<0.01) both for *ML* and for *SP* (px: χ^2 (16) = 624.39, p<0.01, py: χ^2 (16) = 971.37, p<0.01, pz: χ^2 (16) = 9889.18, p<0.01) over all camera poses. However, there were more inter-personal differences in the for *ML* than for *SP* with medium and large effect sizes, specifically for motions parallel to the poster (see Table 4.7). Examples for the large variance in the x, y position between participants in *ML* are the blue and black dots in Figure 4.24 (left and middle). The standard deviation of horizontal movements was 6.1 cm for the participant indicated with the blue dots (h=185 cm, male) and 26.1 cm for the participant indicated with the black dots (h=190 cm, male). Figure 4.23 also indicates that in *SP* the variance in the sampled positions between participants is lower than in *ML* resulting in less heterogeneous inter-person motion patterns in *SP*.

In conjunction with the motion patterns in the x-y plane come large variations of px and py dependent on the distance between camera and poster (see , left and middle) and rotations along the x and y axis in *ML* (see , right). Again for *SP* the range of motions along the z axis is smaller per participant.

Also, with *ML* in average the participants moved closer towards the poster (pz, *ML* mean; 61.9 cm σ=17.7, *SP* mean: 67.3 cm σ=18). The variations in z-distance also induced differences in label sizes and hut sizes.

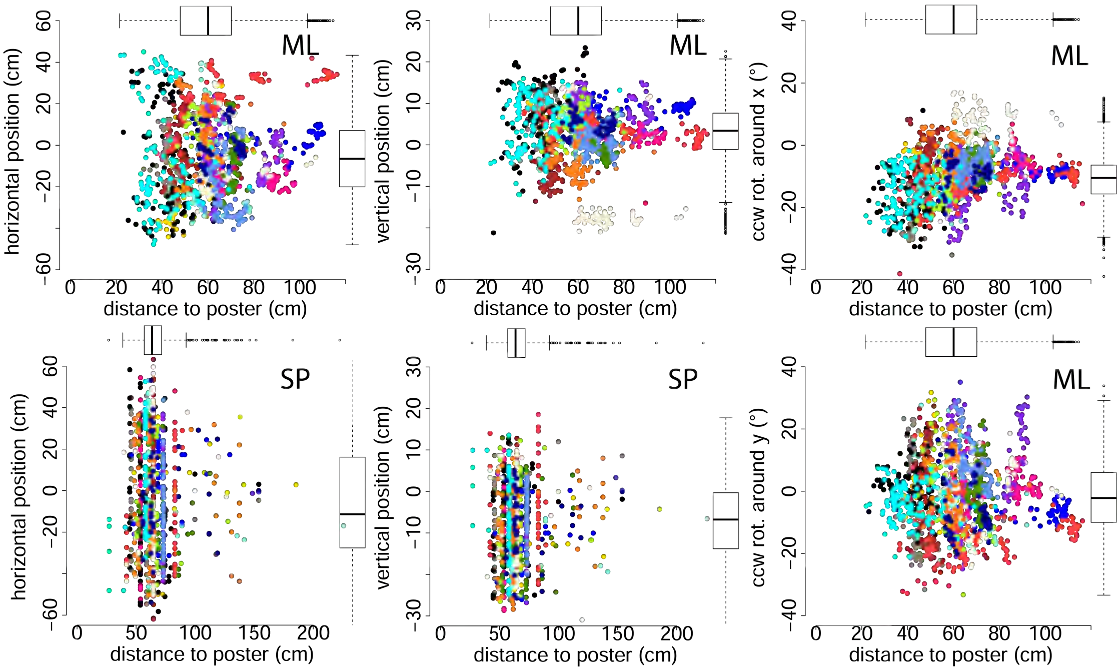


Figure 4.24: Distance to poster in relation to the horizontal (left) and vertical (middle) positions and horizontal and vertical rotations (right). Dots with unique colors represent individual participants. Every 20^{th} camera sample is shown.

Table 4.8 gives an overview of the median and 1^{st} and 3^{rd} quartile values at the moment of candidate selection. Literature reports on recommend text sizes (8-12pt) [50] and soft button sizes [144] for handheld devices. The occurred sizes in our study fall into the previously reported ranges. This is indicates that users adopt their interaction distance to ensure readability and selectability of content.

	ML			***SP***		
	1^{st} qu.	**MD**	**3^{rd} qu.**	**1^{st} qu.**	**MD**	**3^{rd} qu.**
Text size (pt)	8	9.3	11.5	9.1	9.7	10.9
Hut width (mm)	16.2	19	23.4	16.6	17.7	19.8
Hut height (mm)	11.7	13.7	16.9	13.4	14.3	16.1

Table 4.8: Text sizes (pt) and candidate button (hut) sizes at the moment of candidate selection.

User Experience We used the AttrakDiff questionnaire with 5-item scales (strongly disagree . . . strongly agree) to evaluate the effects of interface on Pragmatic Quality (PQ), Hedonic Quality - Identity (HQ-I) and hedonic Hedonic Quality - Stimulation (HQ-S)

[106]. Complementary, the *IE* and *VU* scales of the intrinsic motivation inventory were used [159]. Two-tailed Wilcoxon signed-rank tests did not reveal significant effects of interface on *PQ*, *HQ-I*, *HQ-S*, *IE* and *VU*. The ratings and test statistics are shown in Table 4.9.

UX Dimension	***ML*** **M** (σ)	***SP*** **M** (σ)	**W**	**Z**	**p**	**Cohen's d**
PQ	.32 (.65)	.42 (.31)	47.5	-.83	.41	.28
HQ-I	.64 (.65)	.54 (.47)	75	.85	.39	.30
HQ-S	.93 (.50)	.54 (.81)	94	1.94	.06	.70
IE	.61 (.85)	.71 (.59)	46	-.3	.76	.10
VU	.76 (.92)	1.1 (.56)	29.5	-1.38	.17	.49

Table 4.9: Results of two-tailed Wilcoxon signed-rank tests did not indicate significant effects of interface on the depicted UX dimensions and mean and standard deviation of the ratings (scaled to -2..2).

As the study was conducted in a public setting we were also interested if users would feel potentially more distracted by the audience (passers-by) if they would use expressive spatial interaction methods (as in *ML*) vs. more private and established interaction with *SP*. We employed a set of 3 statements previously used for similar purposes [87] to derive a single score for environmental distraction (by inverting the answers of two items and then averaging the individual answers just as for AttrakDiff and intrinsic motivation inventory). The statements were: 1. “It was hard to concentrate as I was distracted with the environment.”, 2. ”I felt comfortable using the system in the environment.”, 3, “I did not pay attention to the environment when using the system.” Again, a two-tailed Wilcoxon signed-rank test did not reveal a significant effect of interface on environmental distraction (W=62.5, Z=0.62, p=0.54, Cohen's d=0.2) with *ML* having a mean of -.92 (σ=.87) and *SP* a mean of -1.1 (σ=.49). In addition, semi-structured post-hoc interview did not reveal benefits mentioned by the participants for the *ML* interface.

Audience Behavior Using the recorded video data a trained coder identified 370 passers-by in the *ML* condition and 236 in the *SP* condition. Table 4.10 gives an overview of the identified reactions. While only a low number of passers-by interacted with the poster a comparable amount did so in both conditions.

Reaction	***ML***	***SP***
no reaction	74%	79%
glimpse	24%	18%
interaction with the poster	2%	3%

Table 4.10: Reactions of passers-by.

4.3.1.8 Discussion

The physicality of the *ML* interaction did not suffice to engage users more than a traditional screen based interaction and did not show performance benefits. Our results are in contrast to prior findings that user feel more engaged when using *ML* interfaces for leisure tasks [86] or that they can be more efficient compared to *SP* interfaces [201]. While we could not reveal learning effects we cannot exclude that longer-term learning effects can be discarded as a potential source of the performance difference.

The strong differences in motion patterns between *ML* and *SP* can be explained by the smaller number of *DOF* in *SP* (six *DOF*, only translation) vs. *ML* (six *DOF*, orientation and translation) and by the fact that *ML* employs direct pointing with continuous (and imprecise) spatial input. Specifically, *ML* allows panning and zooming simultaneously which can be very useful but also difficult to employ effectively by non-expert users. In contrast *SP* employs indirect pointing by tap-n-drag with non-continuous input; the virtual camera does not change its pose in the absent of user input. Furthermore, the physical extend of the interaction space is much larger in *ML* compared to *SP* (only 9.32x5.6 cm of screen space). This larger interaction space of *ML* paired with the larger number DOFs allows for more expressive interaction, however at the cost of precision. In addition, the larger inter-person differences in *ML* can hardly be explained by the physical characteristics of the participants like height alone. Also, it remains to be investigated if novelty effects of the *ML* interface are a major source of these large inter-person differences.

In previous studies participants mentioned the *ML* interface to be “more fun” and “engaging” [86] compared to an *SP* interface. However, this previous study evolved around a leisure oriented gaming task, whereas in our study we focused on a goal driven task which had no playful elements. It remains to be further investigated if mobile *ML* interfaces are beneficial solely in leisure-oriented tasks such as gaming or if goal-driven and productivity-oriented tasks in mobile settings can also benefit from such an interface metaphor. In touristic environments where users are not in a hurry, more enjoyable goal-driven tasks could be designed by including gaming elements even if it might affect performance.

The relative number of passers-by not noticing or only glimpsing is in line with findings of previous studies [86, 87]. Also, only a low number of participants intruded the personal space of the participants. However, in contrast to previous studies they actively interacted with the poster instead of watching the participant’s actions. The interaction seemed to be driven by an information need of the passers-by (e.g., locating a *POI*) which was addressed by referring to the poster. Participants noticed the intrusions into their personal space but did not interrupt their primary task. Also, in the post-hoc interviews participants did not mention that the social environment was distracting. However, it is not entirely clear to which extend the participants behavior was driven by demand characteristics and the fact they knew they were participating in a study and to which extend similar behavior could occur in a more naturalistic setting.

To summarize, the results of this semi-controlled field study did not reveal benefits of employing *ML* at this specific poster size for our given locator task. Hence, the tourist region of Schladming did not feel encouraged to employ a mobile Augmented Reality solution for this specific poster setup. However, previous work indicated that increasing the workspace size can lead to an improved performance of *ML* compared to interfaces, which do not utilize visual context, such as *DP* [201]. That leads us on to the second study which addressed the effect of workspace size on *ML* and *SP* interaction.

4.3.2 Lab study

We designed a follow-up lab study to investigate how varying workspace sizes would affect the performance of *ML* vs. *SP* navigation. Would the utility of visual context in *ML* help to outperform *SP* navigation as the information space increases? Or would the physical movements required for *ML* interaction mitigate any advantages that the visual context introduces?

We chose a comparable experimental design to the field study and only highlight the main differences. We used interface (*ML*, *SP*) and workspace size (small: 137.5x75.5 cm, medium: 275x75.5 cm, large: 275x149 cm, see Figure 4.25) as within-subjects factors. The physical poster was replaced with a back-projection system (1400x1050 px) and the workspaces were centered at a height of 136 cm. The operational range of the tracking system was 20-200 cm and the update rate 30 Hz (cf. 6-21cm and 10 Hz in [201]).

The vertical extend and mounting height of the large workspace were chosen to maximize the area to be explored while still allowing *ML* usage when standing near it. The medium and small size had half and respectively a fourth of the area of the large workspace and depicted a cropped version of the used background map. While not having exactly the same area, the small size was comparable in dimensions with the poster used in the semi-controlled field experiment. The small and large workspace had the same aspect ratio. The medium workspace was double in width compared to the small to simulate a typical retail window. As mentioned before, our system supported a wider operational range compared to previous work and was targeting larger poster sizes as found in public places. Hence, results reported here are not directly comparable with results obtained in previous studies [201].

For each workspace size we generated 10 maps with the same static background. Restaurant icons (candidate locations) on the reference maps were replaced with prominently visible rings (orange on grey background, inner diameter: 4.5 cm, outer diameter: 9.5 cm). In contrast to the first study, the candidate locations were distributed uniformly randomized (see Figure 4.21, right) over the whole workspace area to facilitate the use of visual context and to minimize learning effects based on memorizing target locations. We chose an item density of 5 items per m^2 (compared 12 items per m^2 in the field study) based on suggestions in related work [201] resulting in 5 (small), 10 (medium) and 20 (large) candidate locations. Only the price was indicated and permanently visible on the device with the same font size as in the field study (no explicit uncover action, no visual feedback on visited candidates). Candidate locations were selected/deselected with a single tap. No restrictions on how to hold the smartphone were imposed.

Also the procedure was similar to the first study. The starting order of interface and workspace size were counterbalanced, and blocked by workspace size. At the beginning of each trial the participant had to go to a starting point in *ML* (150 cm away from the map) or the virtual camera was reset to an equivalent viewpoint in *SP*. The starting order of the maps was randomized between interfaces. The overall study duration per participant was 75-90 minutes. 21 volunteers (11 male, 10 female, average age 30.8 years (σ=7.3), average height 172 cm (σ=8), all right-handed but three, all smartphone users), external to our institution, participated in the study. We had to exclude device data from two participants due to logging problems resulting in performance data from 19 participants.

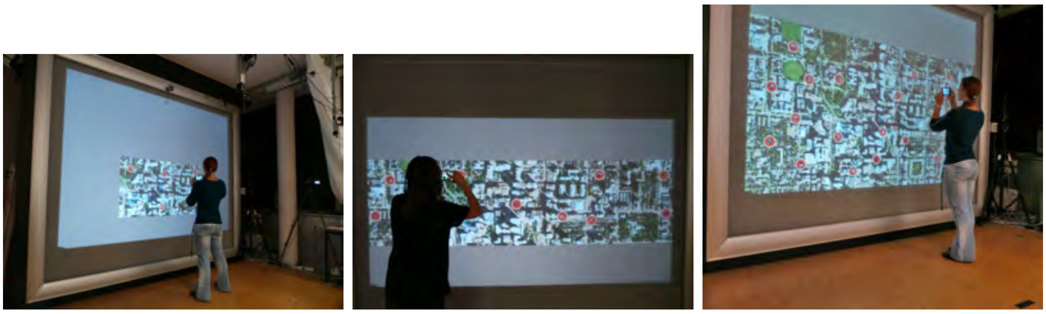


Figure 4.25: The small (left), medium (middle) and large workspace (right) used in the laboratory study.

4.3.2.1 Findings

For our analysis of results we concentrate on the main and simple main effects of interface at individual levels of workspace size. Significant results at a 0.05 level of post-hoc tests are highlighted in bold in Table 4.11 and Table 4.51.

Task completion time *TCT* for *ML* was significantly higher than for *SP* in the small workspace but equivalent for the medium and large workspace. A two-way repeated-measures ANOVA indicated significant main effects of interface ($F_{(1,170)}$=6.6, p<0.02, partial η^2=0.037) and workspace size ($F_{(2,340)}$=1583.7, p<0.01, partial η^2=0.90) on *TCT* as well as significant interaction of interface and size ($F_{(2,340)}$=12.6, p<0.01, partial η^2=0.069).

Post-hoc t-tests showed a significant simple main effect of interface for small workspace ($t_{(170)}$=10.7, Cohen's d=0.82) but not for medium and large workspace. As the mean TCTs were similar for the medium and large workspace we ran a two-one-sided t-test (TOST) analysis with ε being the magnitude of the confidence interval of the *SP* condition. TOST results for medium (dF=170, p= 0.03, ε=2.77) and large workspace (dF=170, p=0.01, ε=3.64) confirmed equivalence of *TCT* between *ML* and *SP*.

Selection errors Selection errors were low for all conditions (see Table 4.11, row 2) and two-tailed Wilcoxon signed-rank tests did not reveal significant differences between individual conditions.

Motion-patterns We observed similar motion patterns for the small poster as in the field study. Figure 4.26 indicates the positions of the camera centers in the x-y plane of the workspace for *ML* (Figure 4.26, right) and *SP* (Figure 4.26. left) for all 3 map sizes (subsampling of 1 pose per 2 seconds for visualization purposes). The middle row of Figure 4.26 indicates the projected camera centers for *ML* in the workspace plane, taking

		Small	Medium	Large
TCT (second)	*ML*	**16.9 (4.9)**	26.5 (7.8)	50.5 (11.7)
	SP	**13.0 (3.4)**	27.9 (9.2)	49.3 (12.1)
Selection errors (%)	*ML*	1.1 (3.5)	4.1 (9.2)	5.3 (5.7)
	SP	1.8 (5.6)	9.9 (11.7)	5.3 (9.3)
Camera path length (meter)	*ML*	4.6 (1.6)	**7.6 (1.9)**	**11.9 (3.4)**
	SP	4.7 (1.4)	**11.8 (4.2)**	**20.8 (6.2)**
Map visibility (%)	*ML*	**30.7 (9.3)**	**8.6 (3.5)**	**15.8 (6.7)**
	SP	**42.5 (13.0)**	**17.9 (4.5)**	**29.9 (8.4)**

Table 4.11: Objective measurements in the laboratory study. Reported are mean and standard deviation (σ).

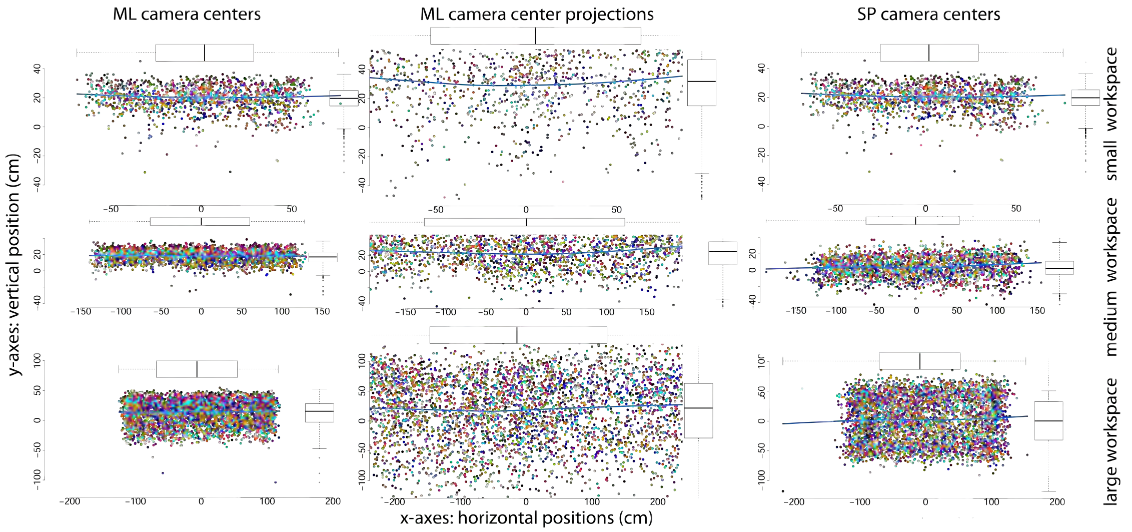


Figure 4.26: Camera position in the x-y plane of the small (row 1), medium (row 2) and large (row 3) workspace for *ML* (left) and *SP* (right), camera positions of *ML* projected on the workspace plane (middle). Dots with unique colors represent individual participants. Every 20^{th} position sample is shown.

the camera orientation into account. Figure 4.27 shows the x and y positions of the camera in relation to the workspace distance.

Over all participants the effect sizes of interface on the position px were only small and medium for py as indicated by two-tailed Wilcoxon Rank Sum tests, showing similar results as in run 1 (for the small workspace), see Table 4.13. In addition the mean distances

interface / size	***ML***	***SP***
small	64.21 (34.8)	89.3 (57.7)
medium	58.7 (28.5)	94.7 (55.7)
large	56.7 (33.8)	89.1 (53.8)

Table 4.12: Mean distances (in cm) between camera and the workspace and standard deviations (σ).

	px (projected px)				**py (projected py)**				**pz**			
size	**d**	**U**	**Z**	**p**	**d**	**U**	**Z**	**p**	**d**	**U**	**Z**	**p**
small	.02 (.01)	>4e8 (>4e8)	.94 (8.04)	.35 (<.001)	1.00 (1.19)	>6e8 (>6e8)	108.3 (122.4)	<.001 (<.001)	.68	>2.5e8	-76.05	<.001
medium	.03 (.05)	>1.3e9 (>1.3e9)	5.07 (7.83)	<.001 (<.001)	1.42 (1.22)	>2e9 (>2e9)	188.8 (167.4)	<.001 (<.001)	1.12	>5.9e8	-158.15	<.001
large	.01 (-1.12)	>4.5e9 (>4.3e9)	3.10 (-11.7)	.002 (<.001)	.28 (.32)	>5e9 (>5e9)	61.5 (71.51)	<.001 (<.001)	1.12	>1.9e9	-212.5	<.001

Table 4.13: Test statistics and effect sizes (Cohen's d) of interface on the positions of the camera based on all position samples. For px and py the values based on camera positions projected on the workspace plane are listed in parentheses.

of the camera to the poster (pz columns in Table 4.13) were significantly smaller for *ML* compared to *SP* (see Table 4.12).

In the *ML* conditions over all workspace sizes the physical camera was moved significantly closer ($t_{(281180)}$= -210.4933, p<0.01, Cohen's d=.71) to the workspace (mean z-distance 58.6cm, σ=32.7) than with *SP* (mean z-distance 90.9cm, σ=55.0). Hence, with *SP* individual map locations (1x1 cm sampling) were visible on the smartphone screen significantly longer for the small ($t_{(10349)}$=-262.5, p<0.01, Cohen's d=2.5), medium ($t_{(20624)}$=-661.9, p<0.01, Cohen's d=4.6), and large ($t_{(41249)}$=-1103.2, p<0.01, Cohen's d=5.4) workspace (see last row of Table 4.11). also indicates how long each 1x1cm cell of the large workspace was visible on the smartphone screen relative to the overall *TCT* (over all trials). For the *SP* interface in the large and medium workspaces most parts were visible on the device screen 20-30% of the time.

Looking at inter-personal motion differences depicted Table 4.14 similar results as in the semi-controlled field experiment can be found for the vertical movements (py) for the small and medium workspace. For the large workspace size the there are few differences between interfaces. In addition, it is noteworthy that the projected camera positions of *ML* in the workspace plane indicate a similar low number of interpersonal differences as *SP*. This indicates that the constraint physical translations of *ML* were compensated with orientation changes. Finally, there are less differences for the horizontal movements (both

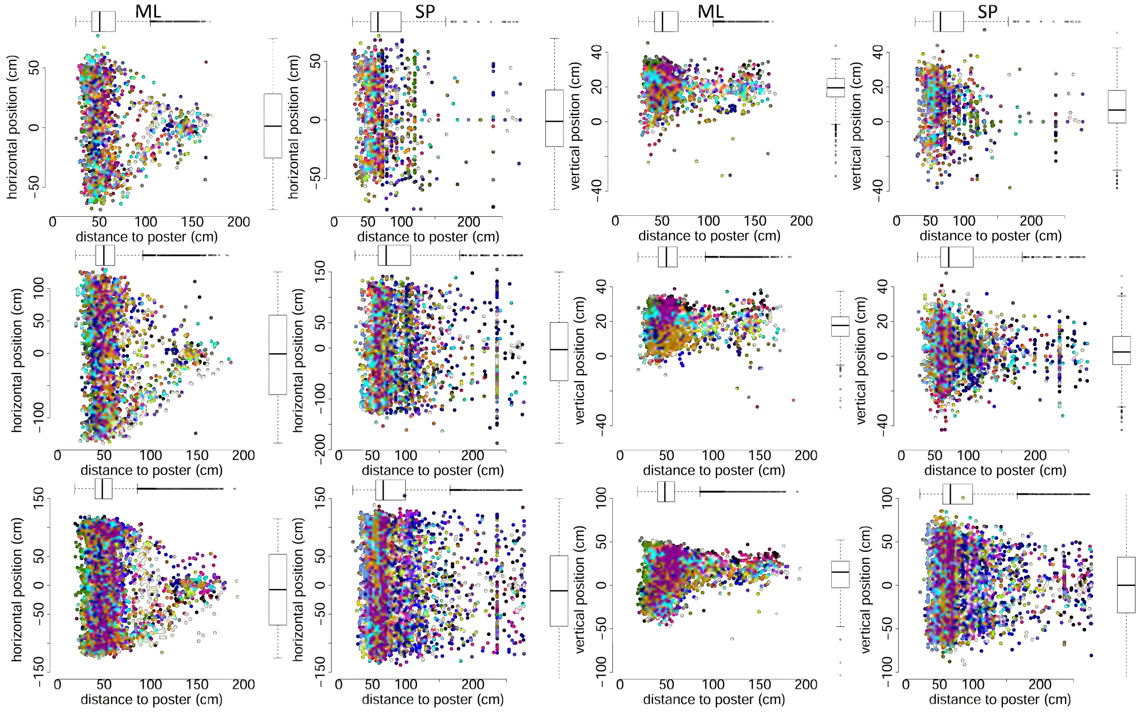


Figure 4.27: Camera distance in relation to the horizontal camera positions (columns 1-2) and vertical camera positions (columns 3-4) for the small (row 1), medium (row 2) and large workspace (row 3). Dots with unique colors represent individual participants. Every 20^{th} position sample is shown.

for *ML* and *SP*) and in the forward-backward movements (pz) for *ML*.

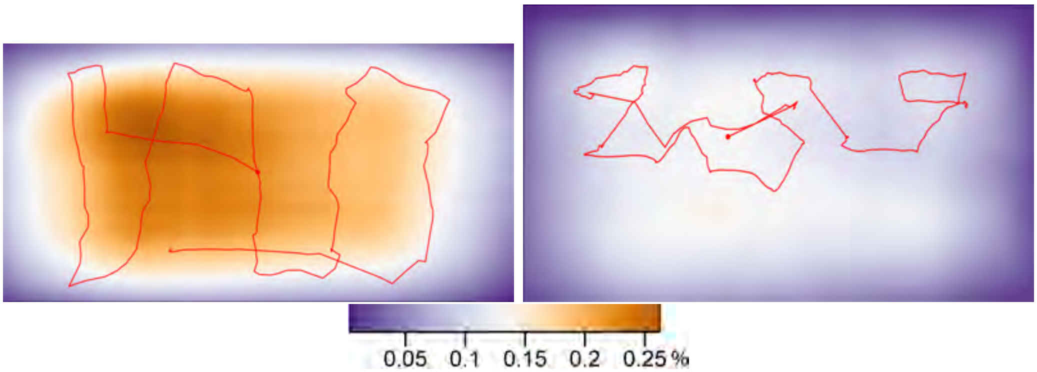


Figure 4.28: Map visibility in the device screen relative to *TCT* over all trials on the large map. Typical camera motion paths of a single participant overlaid in red (*ML*: left, *SP*: right).

Demand and usability Demand and usability were investigated with the NASA TLX and the *PQ* dimension of AttrakDiff [106]. Two-way repeated-measures ANOVA indicated significant main effects of interface and workspace size on the TLX and *PQ* dimensions but

		px (projected px)		**py (projected py)**		**pz**	
size	**interface**	**# med. d**	**# large d**	**# med. d**	**# large d**	**# med. d**	**# large d**
small	***ML***	2 (1)	0 (0)	15 (5)	21 (2)	5	2
	SP	0	0	2	1	9	12
medium	***ML***	0 (0)	0 (0)	12 (5)	32 (0)	6	3
	SP	0	0	4	0	16	10
large	***ML***	0 (0)	0 (0)	3 (0)	0 (0)	9	2
	SP	0	0	0	0	12	14

Table 4.14: Number of significant ($p < 0.05$) pairwise differences (from 171 possible) with medium (Cohen's d = [0.5, 0.8]) and large (Cohen's d >= 0.8) effect sizes for motion in x, y and z plane of the small, medium and large workspaces.

		Small	Medium	Large
Mental Demand	*ML*	2.4 (1.6)	**3.5 (2.1)**	**4.5 (2.2)**
	SP	2.4 (1.7)	**4.2 (2.2)**	**5.6 (2.8)**
Physical Demand	*ML*	**3.8 (2.4)**	4.0 (2.3)	**4.5 (2.2)**
	SP	**2.1 (1.7)**	3.3 (2.3)	**5.6 (2.8)**
Temporal Demand	*ML*	**3.81 (2.6)**	**3.7 (2.1)**	4.5 (2.6)
	SP	**3.19 (2.4)**	**5.0 (2.2)**	5.2 (2.8)
Performance	*ML*	2 (2.3)	**1.4 (1.4)**	2.1 (1.8)
	SP	1.6 (2.2)	**2.5 (2.2)**	2.9 (2.2)
Effort	*ML*	3.5 (2.3)	3.5 (2.1)	4.9 (2.5)
	SP	2.6 (1.9)	4.3 (2.4)	5.3 (2.5)
Frustration	*ML*	2.6 (2.3)	2.6 (1.9)	**3.2 (2.4)**
	SP	2.0 (2.0)	3.3 (2.5)	**4.5 (2.8)**
PQ	*ML*	0.5 (0.8)	0.6 (0.6)	**0.4 (.08)**
	SP	0.7 (0.7)	0.3 (0.7)	**-0.2 (0.7)**

Table 4.15: TLX and *PQ* dimensions. Reported are mean and standard deviation (σ).

are not reported due to space constraints. Similar, post-hoc t-tests indicated significant differences for the main effect of interface for the TLX dimensions highlighted in bold in Table 4.15 (ten point scale low to high) but are not reported here. For pragmatic quality (five point bipolar scale,-2..2), post-hoc two-tailed t-tests did not indicate significant differences for small and medium but for the large workspace ($t_{(20)}$=2.6, p=0.01, Cohen's d=0.6).

Preference and map knowledge Preference for the interfaces was dependent on the workspace size, too. For the small workspace five participants preferred *ML*, 16 preferred *SP*. For medium and large, *ML* was largely preferred (medium: 14 *ML*, 7 *SP*, large: 19 *ML*, 2 *SP*). Reasons mentioned by participants in post-hoc interviews for preferring *ML* for the large map included having a "better overview", "seeing both the overview and detail at once" which "avoids strenuous zooming gestures", "feeling more confident that no [candidate] location was missed". This is also highlighted by 17 participants reporting that *ML* supported them better in building map knowledge (4 for *SP*). These advantages diminished as the workspace size and number of items decreased making the *SP* interface "sufficient". In contrast, the *SP* interface was reported to be "more stable".

4.3.2.2 Discussion

Subjective ratings and feedback in the post-hoc interviews indicated potential usability benefits for *ML* for the medium and large workspace. Participants reported less mental demand (medium and large), physical demand (large) or frustration (large) for *ML*. For the large workspace *ML* had a significant higher pragmatic quality rating. "Better overview" and a "better sense where near-by candidates are located" were commonly reported by participants. That the physical map played a subtle role became apparent when participants reported to have relied on their "peripheral vision to sense where I am on the map" while focusing on the handheld screen. Peripheral vision but also active gaze switches might have been used to correct the upcoming navigation decisions, e.g., by "avoiding empty areas". We assume that the presence of the physical map supports acquisition of survey knowledge [201]. However, we did not observe many participants actively planning a route prior to navigating the workspace, but this might be due to the nature of our task. Specifically, it remains to be investigated if results are consistent if the semantic value of the visual context increases; i.e., if the map structure supports the user not only on a perceptual level. How building and usage of survey knowledge using split attention between a handheld screen and a physical map at a different focal plane works remains to be further explored.

Significant shorter camera motion paths in *ML* for the medium and large workspaces seem to confirm the subjective feedback. The added *DOF* of the *ML* camera (translation and orientation) over *SP* (translation only) can explain the shorter motion paths as around 30° of change in pitch were common for navigating the medium and large workspace in *ML*. However, the increased degrees of freedom of *ML* come at the cost of text readability due to perspective foreshortening (which we did explicitly not correct for with billboarding). This is also apparent in the smaller mean distances between camera and workspaces. In *SP* participants were able to read the text from further away as the camera was always perpendicular to the text plane. Participants stated: "labels were better readable [in *SP*]". In fact, participants reported strategies to find the right z-distance that would allow them to "barely identify the numbers" to avoid pinch and spread gestures for switching between overview and detail. The differences in motion patterns between interfaces were similar as in the semi-controlled field experiment. However, we did not observe as many interpersonal differences in *ML* for the horizontal motion parallel to the workspace. This could be explained by the fact that, in contrast to the poster of the semi-controlled field experiment, in the lab study the candidate locations were uniformly (and randomly) distributed across the whole workspace. Compared to the laboratory study in the semi-controlled field experiment the participants could more conveniently browse the candidate locations by rotating instead of translating once they found a suitable position. This highlights the fact that the attributes of the physical artifacts (in our case the poster design) have to be carefully taken into consideration when designing *ML* interfaces. For the vertical motion parallel to the poster inter-personal differences were similar high between the field and the laboratory runs for the small and medium workspace size. Their vertical extend was similar to the one of the field run. However, for the large workspace size in the laboratory study there were only few differences between *ML* and *SP*. This could be explained by the observation that participants could not easily visit the candidate locations by pure rotation but had to physically translate to be able to read the labels. The lower number of inter-personal differences in the z-distance to the poster for *ML* between lab and field run can be explained by the fact that in the laboratory setting participants had to start each trial from a constant z-distance and then moved towards the poster. In the semi-controlled field experiment participants could move freely during the whole experiment.

Regarding the small workspace size it appeared that *SP* outperforms *ML*, confirming the results of our first study. No benefits could be found for *TCT* and subjective measures. In the interviews participants reported to see no advantages of *ML* for the small workspace as they could "quickly navigate" the digital workspace with *SP* without losing overview. This is in contrast to previous findings [201]. This performance difference could be explained by the effect of current input modalities for *SP* (touch-based screen interaction) and the lack of prior knowledge of *ML* navigation. For the medium and large workspaces we found equivalent TCTs. It remains an open question if with further advancements in tracking technologies *ML* can outperform *SP* or if human motor skills play a more dominant role. Despite the results we found about TCTs one should carefully consider the importance of performance relative to hedonic and affective user experience dimensions in touristic contexts. If performance is not a priority, e.g., if tourists are not in a hurry *ML* should be considered has a design candidate for supporting map navigation tasks.

4.3.2.3 Limitations

Reflecting on our approach to investigate the benefits and drawbacks of *ML* interaction for map navigation we identified a number of limitations. Depending on the viewpoint semi-controlled field experiments can be seen to either combine the best of two worlds – internal validity from laboratory experiments with ecological validity of field studies – or to make too many compromises. Specifically, one can argue that collecting qualitative feedback in a task embedded in the natural routines of users should be preferred over performance-based measures in a field situation as the later can diminish the ecological validity of results. On the other hand it is worthwhile to observe that both performance measures and subjective feedback can be consistent between laboratory and field situations as is the case for our studies. Also one can argue that in a tourism context affective and hedonic aspects of user experience should be addressed more in the interaction design and that our application is too task focused. While we agree that these aspects should be considered for the design of actual products and services in our comparative study settings we wanted to investigate if the interface metaphor alone leads to differences in user experience.

4.3.3 Discussion

Several implications can be drawn for *ML* interaction with maps in tourism. First, supporting relatively simple locator tasks on maps through *ML*, which are solely goal-oriented, are likely to only add value to an experience if the map size is larger than DIN A0. The gained overview of *ML* interaction for large maps cannot be easily compensated with focus and context techniques on small screen devices due to screen space constraints. Designers of such experiences should carefully consider spatial affordances of the physical posters which they want to augment beyond following basic guidelines like a proper mounting height of the poster. Specifically, the spatial constraints between the individual locations on the poster which should be augmented and the user can have an impact on the usability of the whole system.

Second, results do not implicate that mobile *AR* applications for smaller map sizes such as foldable pocket maps are of no use. They do highlight the fact that such applications focusing on locator tasks should go beyond the support of pure utilitarian value and try to address hedonic and emotional aspects in their interaction design. Also, there is a whole class of further tasks which could benefit from *ML* interaction also on smaller maps. Specifically, it would be interesting to purse this work with tasks that are going beyond a predefined goal, e.g., tasks where survey knowledge is critical to success.

Furthermore, *ML* and *DP* can support one handed spatial navigation which is not practically possible with *SP*. This can be relevant when touch interaction is not an option, e.g., when carrying gloves on a ski slope. Another opportunity for investigating benefits of *ML* in tourism are social experiences where multiple tourists (friends or family) are collaborating together on a large physical map to perform a task such as planning their visit of a place. Even though there are other ways to search for *POIs*, like restaurants, or to plan routes without *ML* (such as one person searching on her phone and the others looking over the shoulder, or everybody searching on their phone at the same time) there is potential that *ML* based experiences might be more enjoyable for the tourists [166]. It would be interesting to study if there is more collaboration and enthusiasm on such experience with a physical map and participants with their own *ML* versus a virtual map and participants with their own *SP* in a tourism context. *ML* can also allow tourists to personalize tourists' experiences when interacting with a physical map. For example, seeing only the *POIs* on the map relevant to a particular user, have different font sizes for better accessibility, etc.

Finally, business stakeholders should carefully consider alternatives to creating *ML* experiences for large printed maps. Specifically, interactive public displays can be a viable alternative for engaging people with local product and services, however come at a cost of maintenance of content and restrictions in personalization. Also, interactive public displays might not be practical in some remote locations (e.g., mountain maps where it would be very expensive to bring electricity and where the weather could damage the display). Interactive public displays can also be combined with mobile phone interaction to address needs of different users groups [3] and mitigate challenges with noticing the interactivity of printed posters [83].

4.4 Exploring the design of hybrid interfaces for augmented posters in public spaces

The previous sections indicated that *AR* can have benefits over alternative interfaces when interacting with large printed information surfaces. However, our investigations also indicated that these utilitarian and hedonic benefits depend on various context factors. In contrast to applications using the poster solely as physical hyperlink (such as Google Goggles[3]), *AR* applications require the users to explicitly stay in the vicinity of the physical object during the whole time of interaction. By not doing so, the experience vanishes. Indeed, most current *AR* applications neglect that interaction with printed information surfaces in public spaces often arises in opportunistic situations e.g., waiting at a public transportation stop [238].

Hence, we argue that similar to the creation of mobile multimodal systems [146] consideration of mobile contexts and the specific characteristics of handheld *AR* should guide the creation of interfaces for augmented print media deployed in public spaces. Within this section of the thesis we focus on those user interfaces, looking especially at printed posters: eye-level mounted printed papers that can be attached to planar (most often vertical) surfaces and consisting of graphics and text.

We analyzed augmented print media experiences and structured the design space for hybrid interfaces for these experiences. Furthermore, we derived recommendations to guide the instantiation a specific hybrid interface design and applied it in two case studies. In these case studies the augmented experiences exhibit varying degrees of visual integration between real and virtual elements. Our findings have relevance for the design of a broad variety of augmented print media experience, beyond posters, such as magazines or flyers that are consumed in mobile contexts.

4.4.1 Hybrid interfaces for augmented print media

Interfaces for physical hyperlinks use real-world objects (or attached tags) to retrieve media through URIs (or search queries). For example, a growing number of newspapers equipped with QR codes allow downloading further media (such as videos) related to articles. As the initial retrieval of media typically only takes a few seconds (get out the phone, start app, scan) these interfaces can be employed in various mobile contexts that are characterized by multitasking, frequent changes of secondary user goals and chances of interruptions [203][238]. Information browsing on the go is enabled by displaying solely the retrieved content (e.g., on a mobile website). The physical object used as gateway to the information is not represented in the interface after the initial information access phase. In fact, no assumption is made about the spatial relatedness between the physical object and the linked content.

However, this spatial relatedness between physical and digital parts is a core characteristic of *AR* which can enable rich and engaging user experiences. In handheld *AR* systems this relationship manifests itself both in the realistic *representation of the physical surrounding* in the interface (rendering of the camera view) and the *physical navigation* (spatial gestures and postures) in that space. In contrast to geolocation based *AR* (using GPS and orientation sensors) the spatial extend of the reference frame provided by print media is very limited. The user experience of systems using only *AR* interfaces is tightly bound to this physical reference frame in which the user can be localized. The main question that we therefore try to address through the exploration of the design space of augmented posters is:

[3] http://www.google.com/mobile/goggles, last retrieved 20.04.2015.

How to maintain the user experience initiated at an augmented poster if users move away from it?

To address this question we structured the design space as follows:

1. Frame of reference of the poster

2. Navigation of the information space

3. Transition between interaction spaces

4.4.1.1 Frame of reference

In common Video See-Through (VST) *AR* systems (such as employed on handheld devices) users perceive the physical environment in two ways. First, users are embedded physically in the environment and perceive it through all their senses. Second, they perceive a visual but *digital representation* of that environment on their handheld device. The camera stream is capturing the environment with a different *FoV* and resolution than the human eye and is rendered as background in the interface. It is this digital representation to which additional content is added. Through these two components users perceive the augmented "reality". In fact other interfaces such as digital maps or even list views as employed in *AR* browsers also use (more and more abstract) representations of the physical environment (satellite views, distance and heading information).

While the physical context is lost when moving away from the print medium one can utilize its digital representation to keep up the parts of the original reference frame in a virtual space that are also used to display the added content (i.e., the poster itself without the environment in which it was mounted) (see Figure 4.29). Mapping the extend of the real medium on its virtual representation and resembling the real camera characteristics (foremost the *FoV*) in a virtual allows to keep the spatial relationship (position and sizes) consistent between the reference frame and the augmented items.

To create a digital representation of arbitrary physical environments that can be both viewed and interacted with can become a very challenging task as the scene has to be represented from arbitrary viewpoints. This involves creating a (concrete or abstract) 3D model of the physical scene or at least capturing it from representative viewpoints [74]. However, in contrast to arbitrarily augmented objects, print media consists (in many cases) of one or more layers of graphical and textual content and has nowadays most often already a digital representation (pictorial or vectorial) that finally gets physically instantiated during the printing process.

Our first design recommendation is therefore: *Preserve the frame of reference through a digital representation that encompasses the core parts of the physical environment.*

4.4.1.2 Navigation of information space

AR navigation in the reference frame is generally done in six *DOF* (translation + rotation around x, y, z axis) and accomplished through physical interaction (moving around), It's thus mainly constrained by the users' movement relative to the physical print medium.

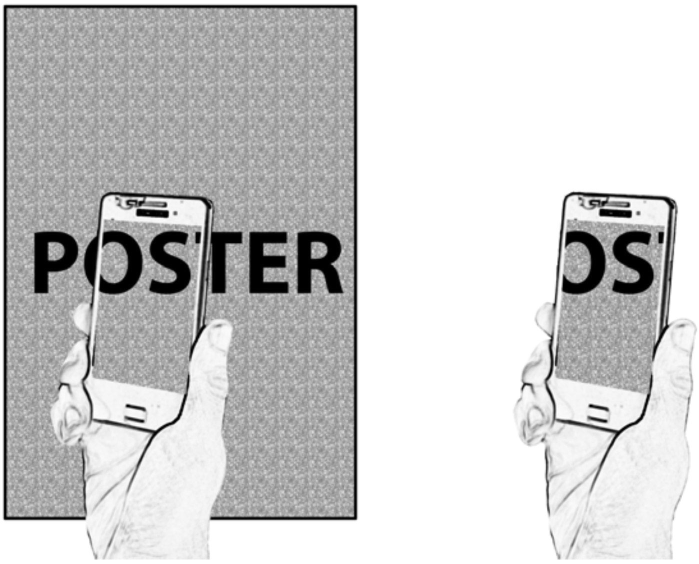


Figure 4.29: The representation of a physical print medium can be preserved by turning it into a digital surface.

To support the continuous interaction on the-go, we must ensure that users can navigate and manipulate the digital representation of the physical object while they are standing away from the printed media.

On mobile devices 3D navigation and object manipulation is a non-trivial task. While there are approaches for both navigation (e.g., [265]) and manipulation (e.g., [97]) many assume some kind of constraint, like application scenario [179] or physical plausible behavior of the virtual environment [97]. Others are not validated in mobile contexts [265]. In contrast multi-touch rotate-scale-translate techniques have become ubiquitous on small (and large) screen mobile devices and are at the hands of millions of users.

Which *DOF* can be constrained in the virtual view is dependent on the spatial complexity of the digital content and the supported tasks [233].

If the digital content integrated on augmented print media is of two-dimensional form (as is the case for the majority of available content in the internet available in mobile *AR* Browsers [86]) or can be interacted with from a frontal point of view the virtual camera position can be constrained to be perpendicular to the digital representation of the print medium. In turn, the navigation in the virtual space can be constrained to 4 DOFs (translation along x, y, and z axes). Now rotate-scale-translate techniques can easily be used. Furthermore, when exploring multipage print media one can utilize surface gestures like flicking to resemble the turn over of physical magazine pages, as done with content aggregation services like Flipboard[4] for handheld devices.

If the digital content is spatially more complex, specific navigation techniques like the ones described in [122] can be integrated. For example, instead of supporting full six *DOF* navigation of elaborate 3D city scenes in virtual space constrained navigation

[4] http://flipboard.com/, last retrieved 20.04.2015.

techniques can be applied to ease the navigation task [145][179]. However, to the best of our knowledge there are no widely adopted techniques for navigating and manipulating in general 3D environments for mobile devices, yet.

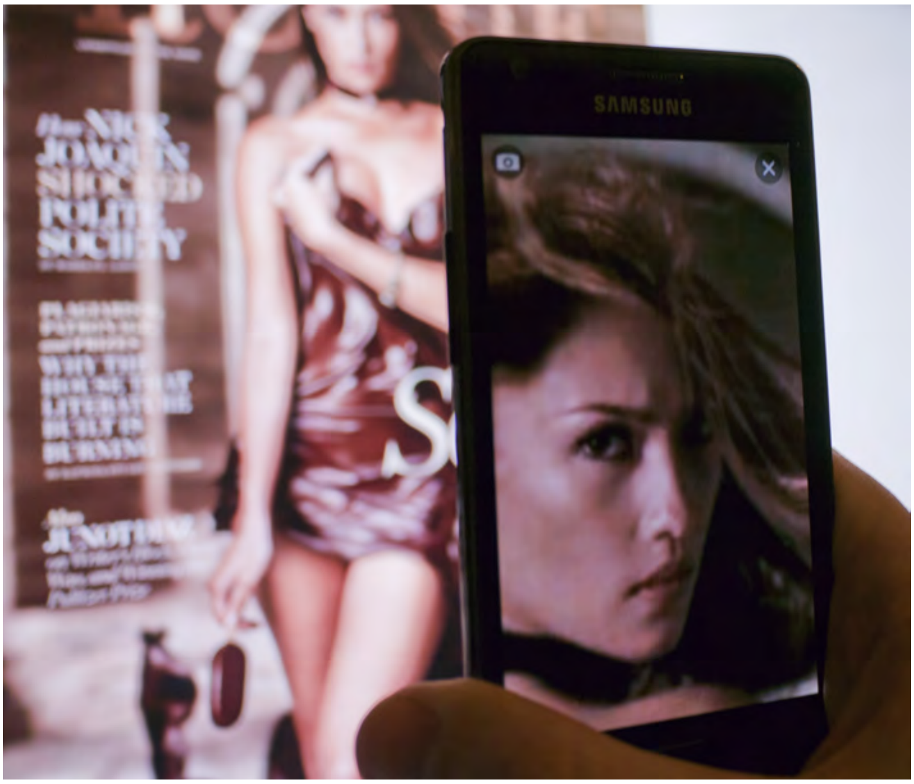


Figure 4.30: To watch the augmented video users are forced to keep the physical magazine in view.

Furthermore, most current augmented poster experiences only employ simple object selection and manipulation techniques. We therefore suggest employing rotate-scale-translate techniques and simple touching for object interaction.

So our second design recommendation is: *favor ease of navigation over complete navigability of the virtual space.*

4.4.1.3 Transition between interaction spaces

An alternative interface for an augmented print medium is not only beneficial if the user moves away but it can also support the exploration of media items when still in the physically vicinity of the printed object. For example, as the viewpoint in a handheld *AR* interface is directly controlled via the movement of the users' arms and hands in a metric space fatigue might result over time. This fatigue can be pronounced if the user is forced to hold the phone in one position over an extended period of time. A common real-world example (as of April 2012) is watching an embedded video on a poster or magazine (see

Figure 4.30, application created by Zappar[5]) which only appears if the print medium is in the view of the camera. Enabling the consumption of the video while still providing the reference to the print medium in virtual space could result to significant fatigue.

However, offering different interaction spaces such as *AR* and a pure virtual representation can be challenging as users need to map the context of interaction when transitioning between them [78] (foremost the viewpoint and the control mapping for navigation). For example, if one restricts the possible viewpoints in the virtual space to be perpendicular to the print medium the user has to mentally overcome the gap between the position of physical camera relative to the real print medium and the position of the virtual camera relative to the virtual representation of the print medium. To lower the cognitive effort of mapping of points from one space to the other view transitioning techniques can be employed [167, 249].

Consequently, our third design recommendation is: *Minimize the cognitive effort when transitioning between interaction spaces.*

4.4.2 Case studies

Using our design space we propose a hybrid interface for augmented posters. We introduce the hybrid interface for two case studies, featuring differing degrees of visual integration between real and virtual poster elements. While both case studies employ vertically mounted posters, similar hybrid interfaces can also be used in other physical configuration of print media.

4.4.2.1 Hybrid interface: AR + zoomable view

Our hybrid interface consists of an *AR* component and a zoomable view component of the print media.

The *AR view* represents the environment in the interface through a live camera view and allows navigation of the information space through movements of the handheld device.

In the *zoomable view* the physical reference frame is represented by a 2D rendering of the print media. In contrast to previous approaches [93][143] the whole information space is interactive and navigable through pan and zoom capabilities (drag and pinch gestures) following the *SP* metaphor [264].

As the relative positions of the digital media elements w.r.t. the printed medium do not change (whether it is now represented physically or digitally) the mental effort needed to map points from the augmented physical space into the virtual space is eased.

The *transition between both components* can be initiated in two ways. Firstly, users can point their phone away from the printed media and will get an orthogonal overview of the whole poster (or page of a magazine). If the user has to unexpectedly leave the implicit change to the virtual view allows her to continue the content exploration later on. Figure 4.31 depicts the example of a vertically mounted poster but the transition could easily be applied to horizontal media like magazines as well. Furthermore, this behavior can be triggered if the tracking of the print medium is failing (see [261] for further examples of dealing with tracking failures). Secondly, users can trigger a transition from the current

[5] http://zappar.com/zaps/rogue/, last retrieved 20.04.2015.

physical camera viewpoint to the closest orthogonal virtual camera viewpoint explicitly. This behavior is useful when an information item is accessed in the Augmented Reality view and should be further explored in a less straining pose (e.g., triggering the playback of a video which is then consumed in a relaxing pose). We allow to transit back and forth between the spaces and, in contrast to previous approaches [132] [44], do not impose a virtual interface when the user is still at the augmented object.

In the following we introduce our two case studies using this interface: Event Poster and Game Poster.

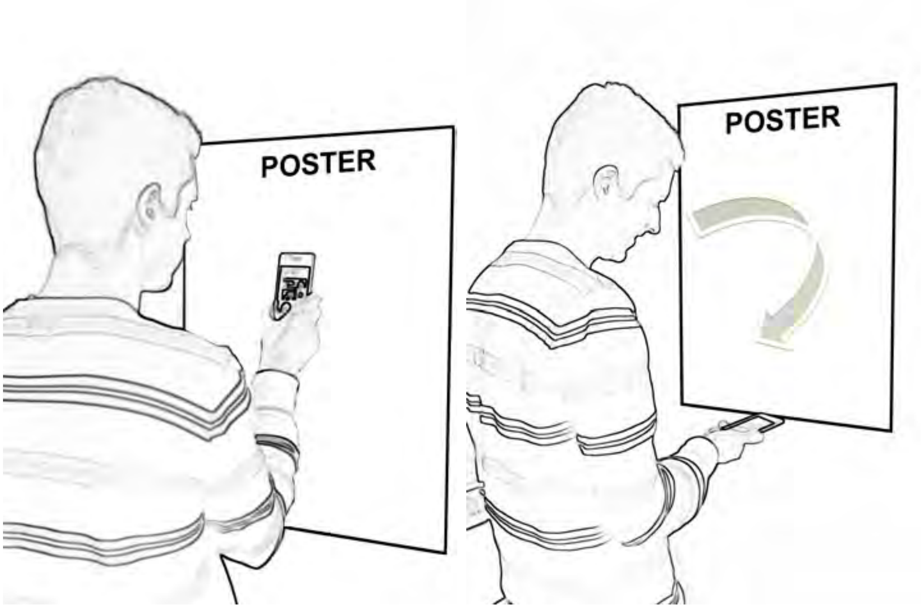


Figure 4.31: Transition from *AR* into zoomable view by pointing the phone away from a poster.

Both prototypes were developed for android smartphones (we used a Samsung Galaxy SII) employing a *NFT* system and running at approximately 30 frames per second.

4.4.2.2 Case study I: event poster

In a first case study we describe a design of an augmented poster in an information browsing task on street posters. It encompasses an augmented poster that could today be created by publicly available tools such as Layar Creator or semantic authoring tools [84].

The Prototype The integration between real and virtual elements on *AR* poster reflects what a designer can achieve without the need for collaboration with a 2D graphics designer, a 3D artist or a programmer. Specifically, we did not include any 3D models but rather traditional media items that are readily available on Internet. In our prototype, we added simple widgets (Facebook, calendar, regions to toggle the visibility of other media items) that could be included in manifold print media independent of the actual content. For the poster theme we chose a local rock band of Graz, Austria. The digital content was retrieved over YouTube and from the website of the band. The augmentation poster (size DIN A0) is illustrated in Figure 4.32 (left). The transition between the *AR* and the zoomable 2D view could be achieved through two techniques. First, an explicit transition from *AR* to zoomable view can be initiated by a vertical bezel swipe [205] thus "freezing" the current view. This technique can be used when still at the poster location when users want to explore information items and not necessarily need pointing at the poster. An implicit transition between the two views is initiated by pointing the phone upwards (enabling the *AR* view) or downwards (enabling the zoomable view). When switching from *AR* to zoomable view by pointing downwards an animation is transforming the last known position of the camera to a default overview showing the whole virtual poster. The video and the Facebook widget only appeared after touching circular trigger regions that could help to temporarily reduce visual clutter on posters with high item density.

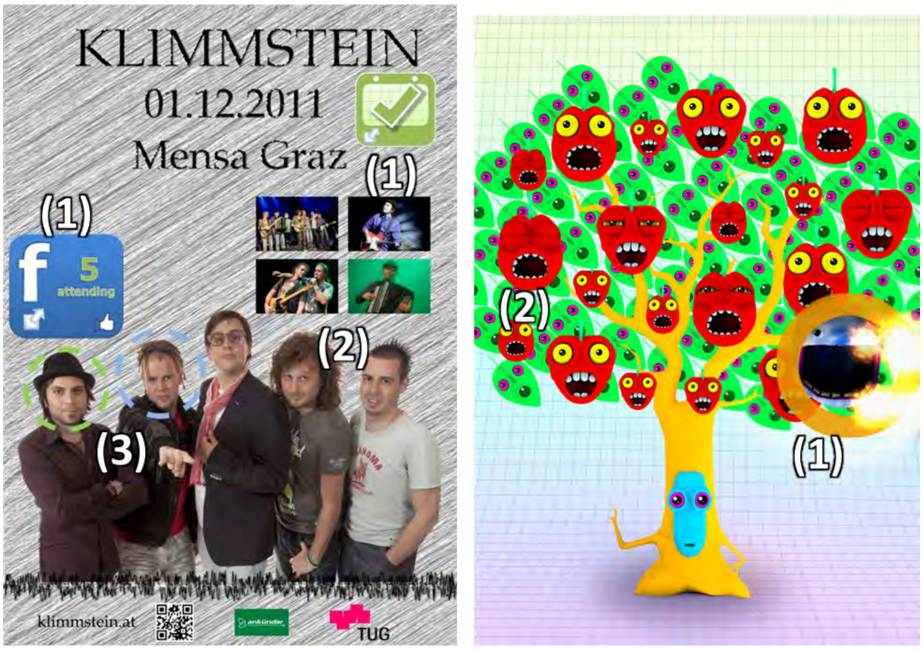


Figure 4.32: Posters with depicted digital content. left: event poster with 2D media items like widgets (1), image collection (2), trigger regions for showing / hiding content (3) and videos (not visible). right: game poster with 3D (1) and 2D (2) animations.

Initial User Feedback A formative evaluation gave us first insights into how users would handle the hybrid interface in the presence of low visual integration between real and virtual objects. We deliberately evaluated this prototype in the lab without considering the mobile context to concentrate on usability aspects of the hybrid interface [130].

To test our prototype we conducted the evaluation with 9 participants (age: M: 25.5 years, SD: 5.6, 4 female, 5 male, 2 with background in *AR*, the others having not interacted with *AR* interfaces before) recruited from on campus. The experiment was conducted inside an office of our institute in which the poster was mounted at eye level. Participants were video-recorded and asked to think aloud. They were informed about the scenario of exploring related digital media connected to an event. A learning phase at a different poster should make participants familiar with the interaction techniques. In the running phase, users were asked to explore the poster (shown in Figure 4.32 left) as long as they wish and think aloud. This part was of exploratory nature to observe how participants would interact with the different information items on the poster. Participants received no instructions on how or which information items they should access.

Interaction in AR and zoomable view: While users were asked to start exploring the information in the mode of their liking, seven out of nine participants started to explore the poster with the *AR* view. They preferred walking up to the widgets and triggering the action ("add entry", "like") by tapping as one participant stated "it's nice to get the information at the touch of a button". These seven participants also initiated the video playback in the *AR* view but eventually switched to the zoomable view with one saying "If I want to watch things for an extended period of time in this (*AR*) mode it feels just a little uncomfortable". Others also mentioned that they would prefer the zoomable view for detailed information exploration with one participant mentioning: "Just walking up and clicking on things is really easy, but if I want to watch something in more detail I like go to this (zoomable) mode". While walking up to the display and pointing on items was found easy one participant also preferred the zoomable view "as I do not want to walk up to the poster all the time". One user explicitly disliked the *AR* view mentioning privacy issues and stating "I am afraid that others might see what I see when holding up the phone. I like to watch the information on my own".

Transition between views: Being able to switch between *AR* and zoomable view was generally appreciated by the participants. The surface swipe for changing from *AR* into zoomable interface was most often used when viewing the video and initiated with the index finger of the dominant hand. A user stated that "it feels nice to capture it (the video item) right where you are". However, not all participants made use of this technique, rather zooming in from the default 2D overview after moving the phone downwards with one mentioning difficulties in conducting the swipe gesture. The animation transforming the egocentric into an exocentric view in this interface was not considered to be necessary by participants. We also observed that three participants tried to pinch and drag on the phone's display while in *AR* mode.

Discussion The evaluation focused on gathering initial feedback about the combination of *AR* and zoomable interface when accessing digital media embedded on a physical poster. Explorative in nature it indicates that participants made use of both views depending on the media type. They preferred to watch the video in a comfortable pose enabled by the zoomable view and exploring information items like widgets with simple functionality directly at the poster.

However, animated transitions from the *AR* to the closest orthogonal 2D view or the overview was not considered necessary. We thus omitted the animated transitions in the later prototype as this study indicated,. On the contrary, visual similarities between *AR* and virtual representation might even be too high. This might be explained by the visually similar representation of the printed and the virtual poster in contrast to the visually demanding changes when switching from a live into a map view [167].

4.4.2.3 Case study II: game poster

The second case study employs the same hybrid interface in a game poster setting. In contrast to the low visual integration between real and virtual elements of the previously described prototype we focused here on achieving a tight coupling between printed poster and digital game elements. For this prototype, we collaborated with programmers, 3D artists and graphic designers.

The conducted user evaluations are described in detail in sections 4.2 and 4.2.

4.4.3 Discussion and recommendations

The usually limited timeframe for accessing information at posters in mobile contexts requires interfaces that do not end the user experience when users leave the physical vicinity of the object. They should allow retrieving and initially exploring the most relevant information quickly directly at the physical print medium and enable further interaction with the information also if that print medium is absent. We initially addressed the missing support for continuing augmented poster experiences in mobile contexts through the exploration of the design space of hybrid interfaces. While we implemented and initially tested our design concepts they have to be further formally evaluated under real-world conditions.

We presented two instances of augmented posters that employed a frontal view and did not allow for complex 3D scene navigation. As described before offering full six *DOF* interaction in the virtual space is neither trivial nor always desirable [233]. Depending on the complexity and type of 3D content that is augmented on the poster one could integrate recent mobile device specific navigation techniques such as [122].

The recommendations we can give from our explorations are: 1. Allow users to explore information while away from the augmented media. To support this preserve the frame of reference of the printed media. 2. If you employ complex 3D scenes think carefully what kind of interactions you want to support in an alternative view. Favor ease of navigation over complete navigability 3. Minimize cognitive effort when transitioning between interaction spaces. While in our case studies transitioning between the augmented poster view and the virtual view was possible without view transitioning techniques they might be beneficial for more complex 3D scenes. Similar think about when to support implicit interface changes. For example one could automatically initiate a view transition for time consuming media items such as videos.

4.5 Summary

Within this chapter we presented a series of studies on the influence of contextual factors on utilitarian and hedonic values of mobile AR user interfaces for interaction with large printed information surfaces.

In the first gaming-oriented study, we investigated the use of *ML* and *SP* interfaces at a public transit place. The *ML* interface was used significantly longer and was preferred by participants over the *SP* interface. The audience on the public space mainly did not pay attention to the participants interacting. Participants themselves did feel isolated from their environment. A comparison to a control group in a laboratory setting did reveal only few differences, despite extenuating circumstances such as weather conditions and the transit nature of the space itself.

We then repeated the study at a public transportation stop with different spatial and social characteristics. Significant differences both for the usage duration and the preference were found, compared to the previous runs of the experiment. Specifically, the *ML* interface was used significantly less and preferred less compared to the previous public condition. Qualitative data analysis indicated that the social context could have influenced the choice of interfaces. However, due to the setup of the study, it can not be reliably stated that social characteristics of the setting were the main factor leading to different user behavior. Further study designs should be employed to investigate this in more detail. Specifically, remote evaluations, where the experimenter is not presented, should be considered in order to minimize participant bias and support a higher intrinsic motivation for using the system. However, this would require creating and distributing a system at near production quality and hence is left for future work. In future work, we also want to conduct studies at further public locations, particularly those that afford social interactions (e.g., mall, train station).

We then turned our focus on the utility of the *ML* metaphor on small screen handheld devices for map navigation. We investigated both performance and user experience aspects. In contrast to previous studies a semi-controlled field experiment (n=18) in a ski resort indicated significantly longer *TCTs* for a *ML* compared to a *SP* user interface in an information browsing task. The follow-up controlled laboratory study (n=21) investigated the impact of the workspace size on the performance and usability of both interfaces. We found *ML* interaction can add value over *SP* interaction for goal-oriented information browsing tasks on maps under specific circumstances (middle and large map sizes). If these conditions are not met, designers of printed information surfaces should consider other factors to increase the user experience with *ML* interaction. Alternatively, they could consider further user interfaces, like interactive public screens, to overcome the limitations of digital maps on small screens and the static nature of printed maps.

Finally, we explored the design space of hybrid interfaces for augmented posters considering mobile users' contexts and the characteristics of AR. Instead of providing solutions that will work for every augmented information surface, we combined AR and zoomable interfaces that will work for many, specifically, vertically mounted posters.

5
Interaction with Security Documents

Contents

5.1 Introduction to document inspection 120

5.2 Technical background 120

5.3 Feasability of interactive user guidance for hologram verification 123

5.4 Towards efficient AR user interfaces for hologram verification 134

5.5 Summary 146

The previous chapter presented studies about large printed information surfaces. In this chapter, we turn our focus to small surfaces. Within the scope of this thesis, we see small information surfaces as ones that can be handheld. For example, flyers, money bills, passports or many books fall into this category. Also, in the previous chapter we looked at interactions which encouraged large bodily movements, e.g., movements involving the whole or the upper body. In contrast, in this chapter, we investigate the utility of handheld devices for tasks that require fine grained motions like subtle hand, arm and head movements.

As with large information surfaces, investigating all possible small surfaces is beyond the scope of this thesis. Instead we focus on documents with security features such as passports and money bills. They are a widespread real-world example of small information surfaces and lend themselves well for studying those fine grained motions.

We start this chapter with an introduction to document inspection and holograms. Then we provide the technical background that is fundamental to be able to build mobile user interfaces for document verification. Afterwards, we introduce our first iteration to build such an interactive user interface and compare it against a common approach using a printed manual. Finally, we present a set of refined interaction approaches aiming at reducing the verification time further.

5.1 Introduction to document inspection

Document inspection is an important part of many security protocols and administrative procedures. This process requires investigation of some or all security features present on a document to be able to decide on its validity. Dependent on the target audience and available tools, document inspection is divided into three classes [247]: First-line inspection (e.g., watermarks, security threads holograms or optically variable ink) is generally done by the public and does not require tools. Second-line inspection (e.g magnetic ink, bar codes, luminescent printing) is carried out by trained personnel and involves special tools. Third-level inspection is usually done by forensic experts, requires sophisticated equipment or knowledge and may even be destructive to the document in question.

Knowing about all the relevant security features requires in-depth training and re-training as new documents are created. For example, a police officer in the field may encounter passports of many different nations in different versions and issue dates. It is very difficult to keep up-to-date with the exact details of changing security features of such documents.

Here, we concentrate on specific security features, namely view-dependent elements and in particular holograms. Such elements require special printing techniques and are therefore hard to forge. They display strong changes in appearance under different viewing directions and dominant light directions. These elements appear on various id cards, passports and most notably banknotes. Consequently, the task of checking such elements is of general interest.

While trained professionals can often reliably identify fake holograms in less than 30 seconds[1], most lay people inspect holograms on security documents just by looking for changes in appearance or the pure presence of rainbow colors, which has no particular value w.r.t. security [247]. First level inspection of holograms is currently based on printed guides which are often issued by public authorities. They usually show distinct patterns visible within the hologram area. However, they often lack an indication on the viewing direction and do not specify requirements on the lighting conditions. Consequently, the inspection may be tedious for the untrained user. Also, in real-world situations manuals are likely not always at hand, so users fall back to solely looking for appearance changes.

Hence, the target audience for our investigations are untrained laymen, who do not have prior training on document inspection and in particular hologram verification.

5.2 Technical background

While this chapter mainly focuses on user interfaces for hologram verification, we first present the necessary technical background that allows to realize these interfaces.

The basic idea of our approach is that, given the assumption of a dominant light source and the known pose of a camera w.r.t. the security document, we can retrieve a (previously recorded) reference image of the hologram. This reference image can then be used either in a user-led or automatic verification step, where the current view of the hologram is compared with the indexed reference image.

[1] according to a domain expert consulted

5.2.1 Capturing view-dependent elements with mobiles

The view-dependent security elements show high-detail images that change drastically depending both on the viewing direction and the dominant light direction. Therefore, a single image cannot capture the full appearance of such elements. We chose to represent the elements using a Spatially Varying Bidirectional Reflectance Distribution Function (SVBRDF) representation (see Haindl and Filip [94] for a complete overview) that allows us to both preserve the dependence on viewing and lighting angles as well as the spatial variation of the images. Furthermore, we are only interested in planar, thin surfaces - printed documents - and therefore we do not require accurate models of self-shadowing or subsurface scattering effects.

However, because we are targeting a handheld mobile application where the device and the document are both moving, we require a full Bidirectional Reflectance Distribution Function (BRDF) representation as opposed to a surface light field as captured by Jachnik et al. [124]. Thus we are effectively using a 6D appearance model per color channel, where the radiance I is a function of both location (x, y) on the document, as well as incoming light direction l and viewing direction d:

$$I = I(x, y, l, d). \tag{5.1}$$

The direction vectors l and d are unit length and therefore have only two DOF.

For our use case we make several simplifying assumptions. We assume that the total radiance from a point on the element is dominated by a single major light source direction. Thus we do not integrate over all incoming light directions, but a single snapshot is enough given the dominant direction. Furthermore, we do not require a fully radiometric calibration and do not control for automatic exposure and white balancing in the camera.

We simply sample the appearance as a set of images indexed by viewing direction d and light direction l. We do not attempt to estimate a smooth $BRDF$ model covering all points on the element, but rather keep the individual images as the final representation. This preserves the sharp changes in appearance when the element flips from one view to another, as well as the necessary detail in the spatial domain.

In practice, the dominant light direction poses a challenge in a mobile setup. Without any prior knowledge, we cannot reliably index into the list of appearance images. Therefore, we re-use the LED light source on a mobile as a constant source of illumination in the scene. As this is usually close to the camera, it easily dominates other light source in the environment. Because the LED is fixed with an offset vector o with respect to the camera, the light direction l is now a function of the camera pose with respect to the document (see Figure 5.1). The light direction is now proportional to the camera position P plus offset vector o rotated by the camera rotation in world coordinates.

$$l \propto P + R \cdot o \tag{5.2}$$

For a fixed distance to the surface, P is just a rotated vector as well, and we obtain a similar equation for the viewing direction

$$d \propto R \cdot \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} \tag{5.3}$$

Thus our representation is reduced to a 5D model, indexed by the full 3D camera rotation and the location (x, y) on the document.

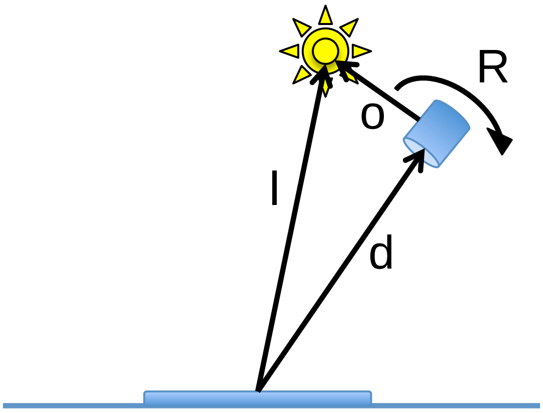


Figure 5.1: With the LED light source in a fixed configuration to the camera, there are only three *DOF* in the input to the *SVBRDF* function.

When operated at a small distance to the document, the built-in LED flashlight of a mobile phone dominates other light sources in typical indoor scenarios. However, this assumption is invalidated with strong artificial light sources or when operating outside (e.g., direct sunlight). In such cases, the workspace must be carefully shielded (e.g., manually).

Furthermore, the flashlight may introduce severe specular highlights, even directly on the hologram. These highlights usually appear around the orthogonal view of the target, but do not affect the application much, because the more interesting views for verification are often at an angle away from the normal. Moreover, the verification of most holograms does not require dense sampling but relies on a rather limited number of specific views.

5.2.2 Retrieving reference views on a mobile phone

To retrieve a reference view of a hologram we need both the dominant light source assumption and the pose of the phone w.r.t. the security document.

By visually tracking the known document our system estimates the current viewing direction and camera pose. The camera rotation indexes into the stack of appearance images of a reference element.

The above described approach only works for single security documents. This approach can be simply extended to multiple documents by first recognizing the overall security document. In our approach we adapted a mobile visual search pipeline for this task, running standalone on the mobile device. We compute SURF features [18], cluster them in a hierarchical k-means tree [172] and perform geometric verification by robust homography estimation to re-introduce spatial information. This provides reasonable recognition performance and scales up to a large number of documents. We then configure a natural feature tracker with a representative example of the recognized document class.

5.2.3 Automatic verification

As mentioned previously, once current view and reference view are associated users can either manually compare both images or an automatic matching process can be used.

Verification of selected reference data at runtime demands a suitable similarity measure. We explored two approaches. First Normalized Cross Correlation (NCC) was evaluated for matching, giving reasonable results for the majority of views. However, certain holograms (e.g., rainbow) can show a large amount of different colors, which leads to noisy measurements. From our experience, pre-filtering (Gauss, Median) with a reasonable kernel size improves robustness for this type of holograms. On the other hand, stereograms tend to produce more distinct patches, which lead to strong edge responses. Consequently, it seems reasonable to use both intensity and edge information in the process. Hence, we propose shape-matching using the modified Hausdorff distance [59] and weight the contributions according to Eq. 5.4,

$$\begin{aligned} score = (s_{NCC} * k_{NCC} + d_{NCC})f \\ +(s_{MHD} * k_{MHD} + d_{NHD})(1 - f) \end{aligned} \tag{5.4}$$

where s_{NCC} and s_{MHD} denote the corresponding individual results, k and d denote individual scaling coefficients, and f is a weighting factor. The individual scaling coefficients are computed using all the recorded reference data and are used to perform appropriate scaling at runtime . We use integral images and a sliding window to avoid a costly registration step. With these modifications, matching runs in real-time on mobile devices.

5.3 Feasability of interactive user guidance for hologram verification

As stated before current printed guides for hologram verification show reference views but do not specify at which pose (and under which lighting conditions) exactly a user could find this view on the actual document.

Hence, in a first approach we wanted to improve the accuracy with which a user can be guided to see a given reference image. The idea is that the mobile system guides the user to capture a frame from the same viewing direction and under the same light direction as captured in a reference image set. Using the LED light of the mobile phone as a dominant light source, the task is simplified to aligning the current pose of the mobile phone operated by the user with several reference views having six DOF. An overview of our first system can be seen in 5.2.

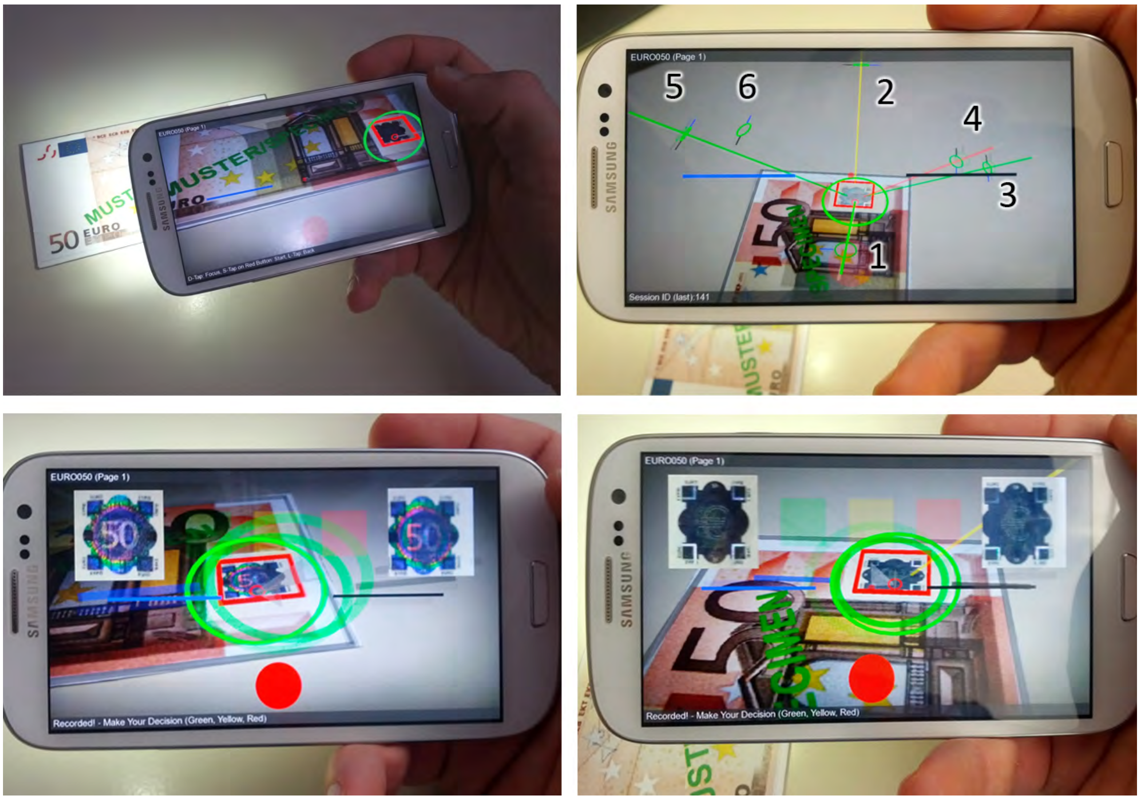


Figure 5.2: First iteration of our interactive system for verification of view-dependent elements: It performs *SVBRDF* capture using the built-in LED on the mobile device (top-left). The user gets an overview of relevant views for verification, which are color-coded w.r.t. the decision of the user (right, note the number attached to each view). The system allows the user to accurately match given referene views and to compare the changes of holographic or similar security elements with the corresponding reference appearances (bottom).

5.3.1 User interface concept

We propose a visual guidance approach inspired by two widely known metaphors, namely iron sights and virtual horizon. Iron sights are used to align the viewing direction of the operator with the direction of the device. In general, shaped alignment markers are used for this task, which are positioned at a given distance on the device. Accounting for distance or scale depends on the task and requires a calibration procedure. This is often applied in sighting mechanisms.

The virtual horizon is an indicator of level, which is often used when a device needs to be aligned relative to the ground. At any time the instrument shows the level of the object relative to earth gravity. Implementations range from a simple water level for mechanical tasks to advanced electronic devices used in aircrafts.

Based on these techniques, we subdivide view alignment into three steps: We match the direction of the viewing ray (iron sights), the position along ray and also the in-plane rotation (virtual horizon). It is crucial to guide the user through these steps, so that accurate alignment can take place (see Figure 5.3 for a conceptual overview).

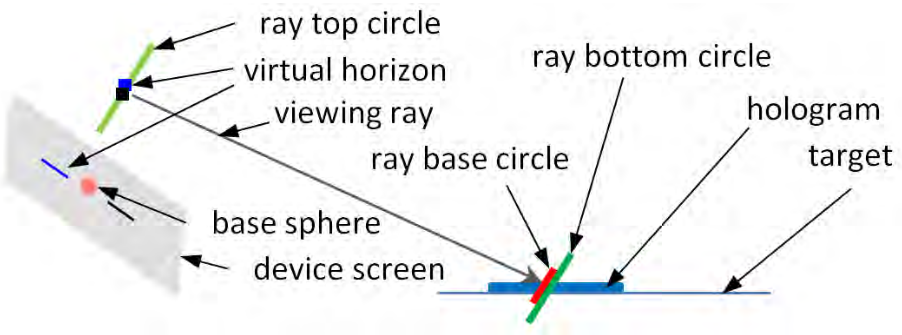


Figure 5.3: Geometry of the proposed alignment approach. Matching the current view with a given reference view takes place by aligning the viewing ray direction, position (base sphere on the device screen with the ray base circle, ray top circle with the ray bottom circle) and orientation (virtual horizon on top of ray with the virtual horizon on the device screen).

5.3.2 Implementation

We implemented the proposed guidance approach in an interactive prototype for mobile devices. The iron sights setup is realized by using two big circles which mark start- and end-point of the viewing ray. By using the intrinsic parameters of the camera, we scale the lower ray circle so that it overlaps entirely with the top circle once direction and distance match. For easier alignment, we additionally use a smaller ray base circle, which is intended to overlap with a small sphere fixed on the device screen. Their scale is also adapted with the intrinsic parameters. The virtual horizon setup consists of two lines placed at the top of the ray and two similar lines fixed on the screen. By using two different colors for each line, we account for a possible ambiguity in rotation around the optical axis (see Figure 5.4).

We used the following color scheme to support the three-step alignment approach: red for on-screen sphere, small ray base circle, green for big ray base circle, top ray circle and blue/yellow for the virtual horizon. Depending on the most similar (w.r.t. orientation) reference pose, an automatic pre-selection is carried out by the system, drawing the full iron sights and virtual horizon setup for the selected reference pose only. Whenever the user makes a selection, the color of the reference ray is adapted. So the user gets a short summary of her decisions when viewing the setup from farther away and also knows where no decision was recorded up to that point. We also draw the last captured ray associated with the current reference pose so that the user can get an impression of how well the captured views fit (see Figure 5.2 and Figure 5.4).

5.3.2.1 SVBRDF capture

The choice of reference poses obviously depends on the hologram (e.g., number of transitions) and is constrained by the particular setup being used. For each view, we require stable tracking and reproducible appearance. While the first mainly excludes low angles and extreme close-up views from being recorded (tracking failure), the latter limits the maximum viewing distance and avoids orthogonal angles which produce specular highlights due to the placement of the LED light. In practice, it seems reasonable to operate

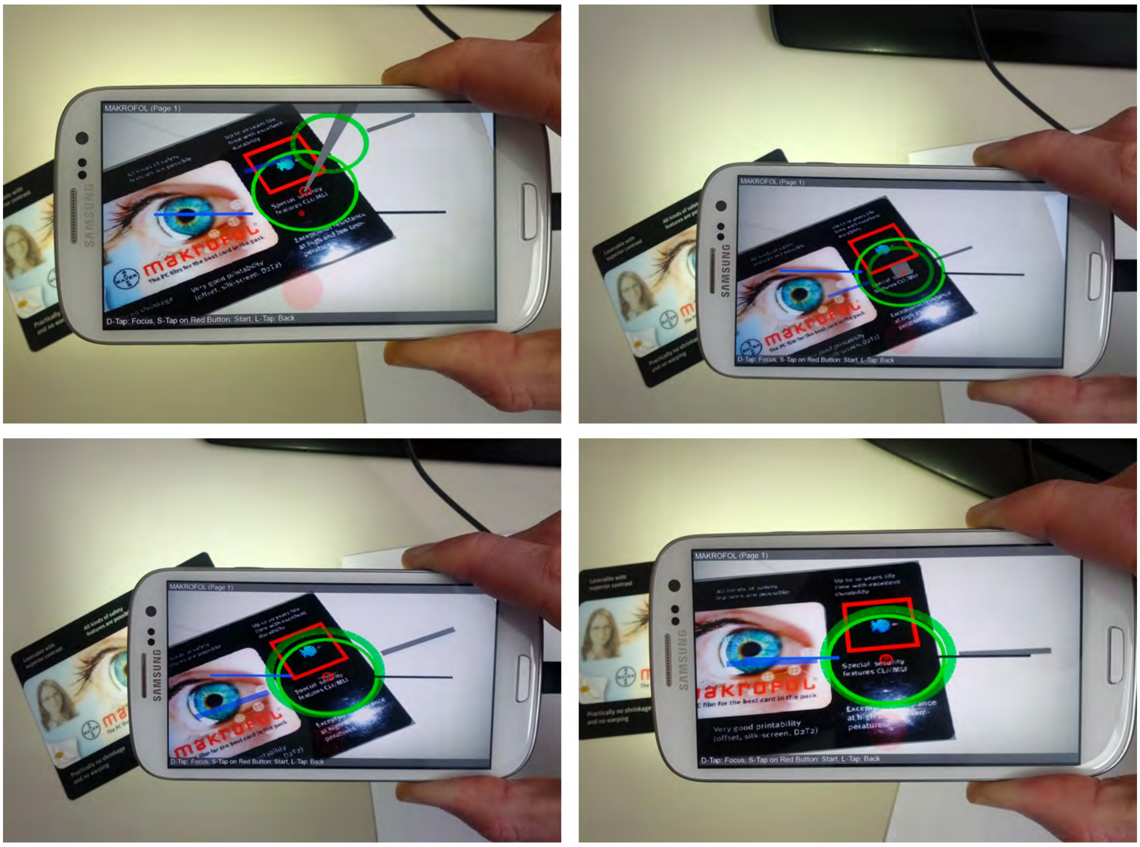


Figure 5.4: Exemplary alignment sequence: Not aligned (top left). Aligning direction using iron sights (top right). Adjusting distance (bottom left). Aligning rotation using the virtual horizon (bottom right).

roughly at constant distance from the hologram, giving a hemispherical capture space. For the holograms described in this paper we recorded two to six views with stable appearance of the patches at a distance of approximately 10 cm.

During verification, image capturing is triggered by the user when the alignment of a reference pose and the current pose is deemed close enough for accurate visual feedback. In this case, an auto-focus operation is triggered and the tracking pose is checked for stability before the current frame and corresponding pose are recorded. This is to avoid recording of pose jitter or blurry patches. We assume the hologram to be planar and project the bounding box of the hologram into the image by using the current pose. We then estimate an image transformation with respect to the hologram region on the undistorted template and subsequently warp the sub-image containing the hologram. Consequently the appearance of the warped patch corresponds to the selected viewing direction. This allows for an efficient comparison. We display this patch side-by-side with a reference patch. This similarity must be rated by the user to express consent, uncertainty or rejection.

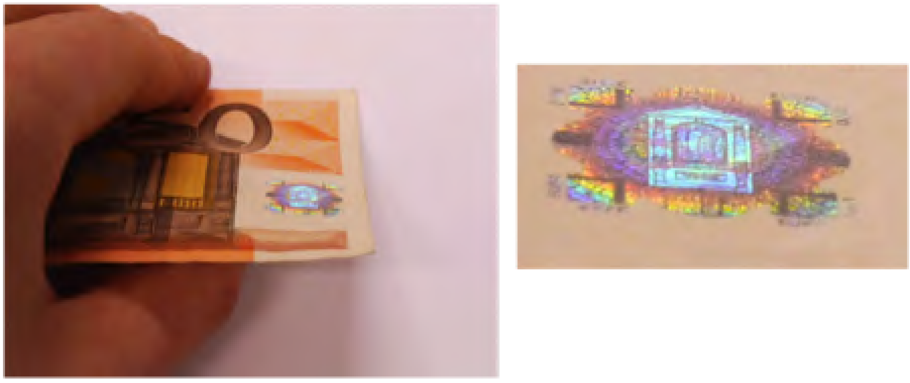


Figure 5.5: Exemplary view used in the *DM*. Overall image indicating the viewpoint (left). Zoomed image of the hologram patch (right).

5.3.3 Evaluation

To test the feasibility of the proposed approach for mobile hologram verification, we determined several performance parameters with users in a pilot study. This study had two aims:

- Record the performance of users in target acquisition.
- Provide a first comparison to a simple paper based method.

For system performance, we wanted to know, how accurate users can acquire the necessary viewing directions, given our guidance systems. Understanding the potential accuracy limits is important for determining minimal angles and distances between views for verification and learning what differences the system has to tolerate. Moreover, we wanted to see, if users can correctly verify a hologram using the current system. It is not clear up front, if the representation on the screen under real lighting conditions is comparable and looks similar enough to users.

An additional goal was to analyze the potential for automatization of the process, which includes automatic capture and matching of hologram patches.

5.3.3.1 Study design and apparatus

We followed a controlled, within-subjects study design recording view alignment and matching performance, but also comparing the effects of the *AR* interface and a Digital Manual (DM) (providing visual step-by-step instructions, see Figure 5.5) on several aspects in a hologram verification task. We investigated both performance based measures (alignment error, *TCT*, error rates in matching and for the main task) and user experience dimensions (instrumental dimensions like usability, non-instrumental dimensions like hedonic stimulation and identity and emotional dimensions like intrinsic motivation).

The experiment took place under controlled laboratory conditions. Specifically the lighting was fixed to allow for comparable results in the *DM* condition. Both interfaces were deployed on Samsung Galaxy (SIII) smartphones. The tasks were carried out on while seating at a round table but users were free to move around at any time (see Figure 5.6).

5.3.3.2 Task and procedure

As main task we chose the verification of the hologram present on a 50 Euro banknote, which is one of the most often counterfeited banknotes in the Euro zone[2]. It must be noted that the holograms on banknotes with higher values (100, 200, 500 Euro) behave in a very similar way. The participants should inspect 4 holograms with each interface. Specifically, they were asked to view the hologram from 6 different viewpoints (depicting 3 different pictures the banknote value, a window, and a doorway - see Figure 5.2 for view locations) but were free to stop the hologram verification before completing all views if they already came up with a decision. They were instructed to compare the reference close-up view of the hologram with the view of the hologram that they were inspecting and decide if they were similar. We pointed out that the holograms do not have to match on a pixel-by-pixel view but did not give any further hints on what similar meant, leaving this decision up to the participants. After inspecting the hologram from all 6 views participants should come up with an overall decision if the hologram was a real one or a counterfeited one. We did not inform the participants at any time before, during or after the experiment if counterfeited (or real) holograms were among the ones they inspected. We used 8 printed specimen notes in total (4 per interface) and only left a hole for showing the underlying hologram of a real banknote (see Figure 5.6) to avoid that the checking of further security features of the banknote could influence the participants judgments. All employed holograms were real (no counterfeited hologram was used).

At the beginning of the experiment users filled in a background questionnaire. They proceeded with a learning phase of the starting interface (*AR* or *DM*, counterbalanced) inspecting a hologram not related to banknotes followed by the main task. After inspecting each banknote participants briefly indicated their confidence in following aspects in an online questionnaire: Is the current hologram real or fake? Did the depicted reference

[2] https://www.ecb.europa.eu/press/pr/date/2013/html/pr130719.en.html, last retrieved 20.04.2015.

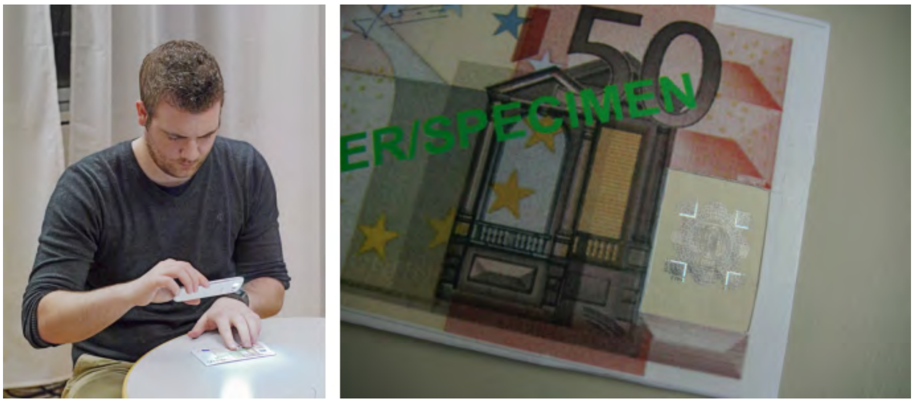


Figure 5.6: Image showing table setup used during the study (left). Specimen banknote with window showing hologram to be checked by participants of the study (right).

viewpoints match with the ones of the participants? Did the depicted reference close-up views match with the ones the participants saw?

After checking the holograms on all 4 banknotes participants completed intermediate questionnaires regarding workload and UX qualities of the interaction. They then repeated the procedure (training, main task, questionnaires) with the second interface. At the end of the study a short semi-structured interview was conducted focusing on aspects observed during the participants' interactions with the interfaces. The overall duration of the experiment was around 60 minutes.

5.3.3.3 Participants

We conducted the study with 17 volunteers (1 female). Most participants reported to have considerable experience with computers and a high interest in technical matters. Only two volunteers reported not to own a smartphone or tablet. However, the majority (13 participants) had never checked a hologram before. Three of the participants were English speaking but all instructions and questionnaires were given to the participants in either German or English (as they preferred).

5.3.3.4 Data collection

Within the experiment we collected device-, video- and survey data complemented with photos and notes. For the *AR* system we recorded camera poses and user interactions and captured hologram patch data along with *TCT*. In case of the *DM* we measured the *TCT* with a separate clock. In addition the interactions of the users were video-taped. Besides quantitative analysis of data we employed several subjective scales to capture both general UX dimensions as well as task-specific aspects. Specifically, we employed the Nasa TLX for workload assessment [101], AttrakDiff [106] for capturing hedonic (stimulation, identity) and pragmatic UX dimensions and the *IE* and *VU* sub-scales of the intrinsic motivation inventory [159]. We analyzed quantitative data with the R statistical package and Microsoft Excel. *NHST* was carried out with the 0.05 level. For the positional and orientation data we treated all data outside the 2.5% and 97.5% percentiles as outlier. The percentiles were computed on the aggregated data over all views.

5.3.3.5 Results

We first analyze user performance in view navigation by comparison with the 6 given reference views at all relevant events. The subsequent analysis of patch similarity using image-based measures gives an impression on the performance of the proposed approach for mobile *SVBRDF* capture. Then, we provide results on task-level performance (hologram verification) for the *AR* system and the *DM*, which attributes to patch similarity rated by the user and the ability to come up with a final decision. Finally we provide results related to the user's subjective assessment.

One participant took significantly longer for the proposed tasks then it was suggested. As this behavior was limited to a single person, we consider the associated runs to be outliers and do not use the associated data in the evaluation.

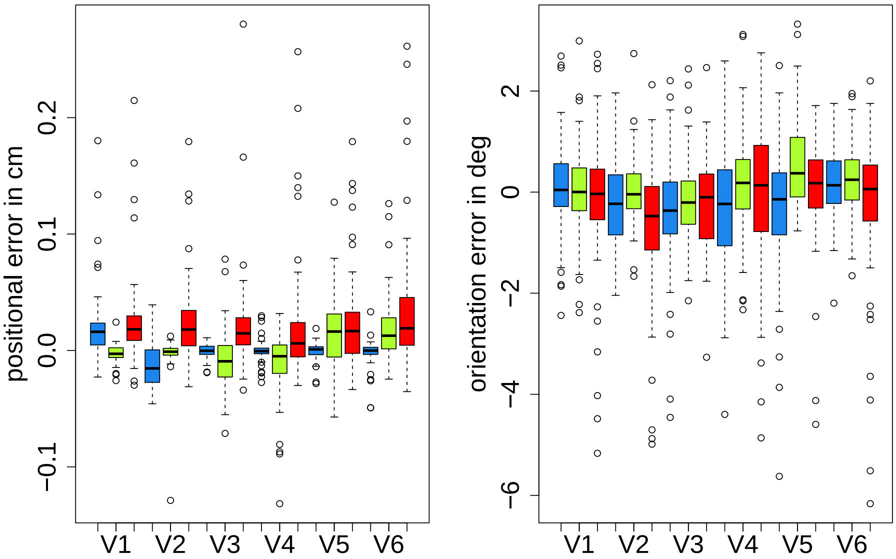


Figure 5.7: Alignment errors for different views of the hologram captured in the user study. Translation (left). Rotation (right). Axis color-coded: x: red, y: green, z: blue

Maneuvering to Target Poses We analyzed data corresponding to all selected views during the study. Ranges of alignment errors in translation and rotation give a hint on the level of accuracy attainable with our guidance approach (see Figure 5.7). For translation the range of translation error is -8mm to 10mm. The range of rotation error is -8 to 8 degrees. Overall, the largest error is encountered with view number 4. This was the first view typically selected by most of the participants, when they were still gaining familiarity with the system.

Another way to assess the performance of users using the guidance system, is to compare the captured patches with a suitable image similarity metric. We register reference and captured patches using optical flow [185] and use NCC as our measure for patch similarity. The optical flow correction is to account for inaccuracies due to unstable tracking (see Figure 5.8). In this setup 4 out of the 6 views obtain average NCC scores above 0.75. This suggests that the proposed setup for SVBRF capture w.r.t lighting conditions and pose accuracy allows acquisition of hologram patches for non-expert users. Two views have very low NCC scores, however. Again, one of them is the view most users approached first, when they cannot be considered entirely familiar with the system.

Task performance Regarding the TCT, the medians of the AR and the DM interface were 188 and 103 seconds, respectively (see Figure 5.9, left). As the data was not normal distributed a two-tailed Wilcoxon signed-rank test was employed and showed that there is a significant effect of interface ($W = 1687, Z = 4.48, p < 0.05, r = 0.48$) on TCT.

Participants rated how sure they were that the banknotes are real and fake for each banknote(see Figure 5.9, right). In addition they rated how confident they were that the individual hologram views corresponded to the reference close up views and how

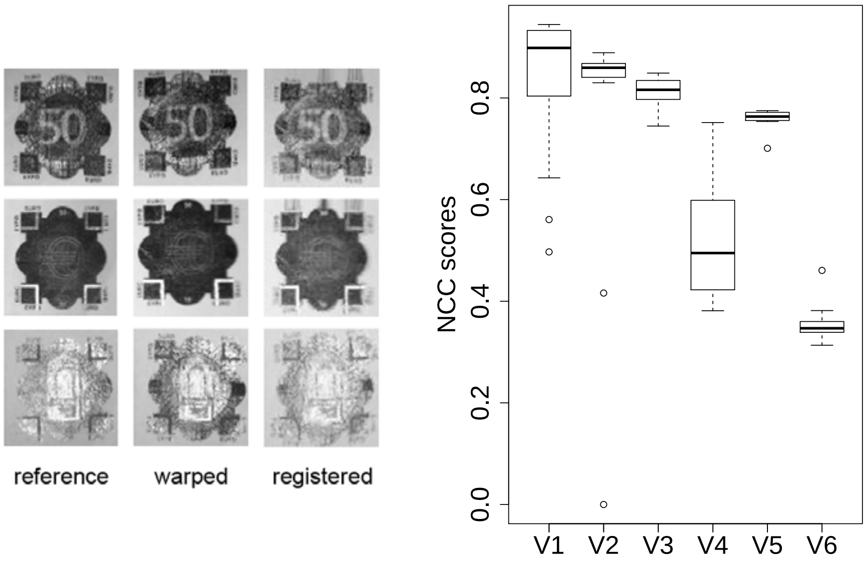


Figure 5.8: Matching registered patches: reference, warped image, registered image (left). *NCC* scores with registered images for different views (right).

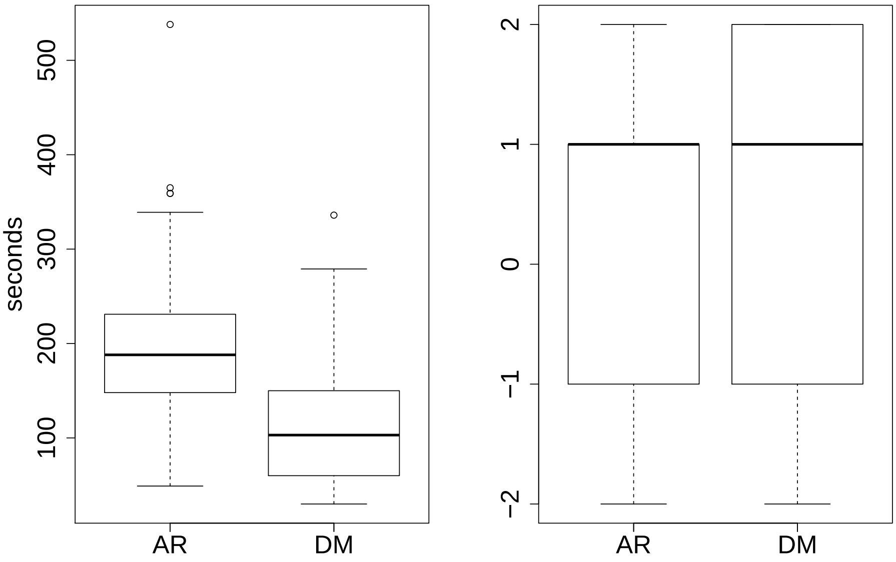


Figure 5.9: *TCTs* (s) for the *AR* and *DM* interfaces (left) and agreement to 'I think the hologram is real' (right).

confident they were that their viewpoints corresponded to the reference viewpoints (camera poses). For the pooled-results (over all 4 banknotes) a two-tailed Wilcoxon signed-rank tests showed no significant effect of interface on any of those ratings.

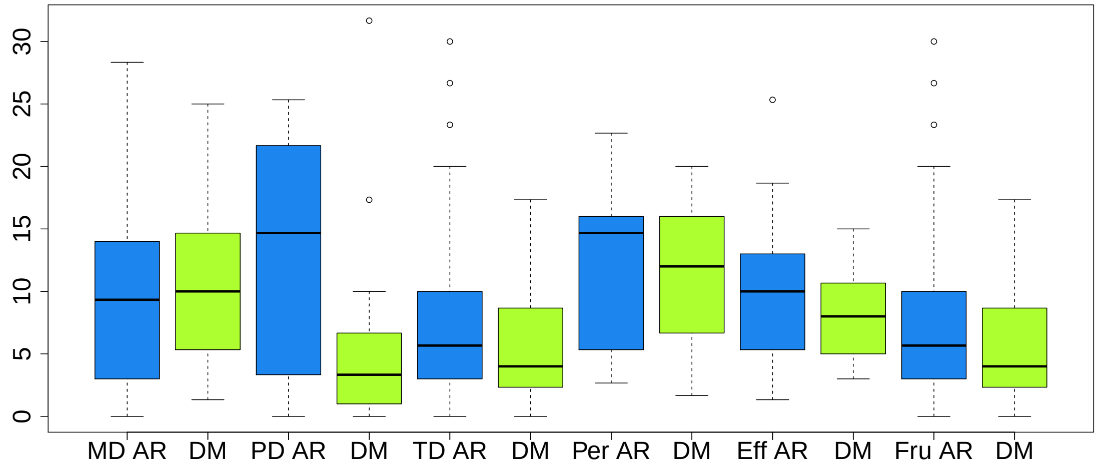


Figure 5.10: Weighted NASA TLX dimensions for demands imposed on subject and for task interaction (MD: Mental Demand, PD: Physical Demand, TD: Temporal Demand, per: Performance, Eff: Effort, Fru: Frustration.

Subjective assessment We used the NASA TLX weighted scores scheme to assess subjective demands. For computation of the scores we used both the magnitude of load (ratings) and sources of load (weights) which evaluate the contribution of each factor. The ratings for demands on subject and for task interaction are shown in Figure 5.10. Two-tailed Wilcoxon signed-rank tests indicated a significant effect of interface ($W = 98, Z = 2.13, p < 0.05, r = 0.37$) on physical demand (MD for AR: 14.67, MD for DM: 3.33) and a significant effect of interface ($W = 111, Z = 2.32, p < 0.05, r = 0.40$) on temporal demand (MD for AR: 5.67, MD for DM: 4.00). There were no significant differences in NASA TLX weighted scores for the other dimensions.

The AttrakDiff questionnaire is an instrument for measuring the attractiveness of an interactive system along pragmatic and hedonic user experience qualities. Paired, two-tailed t-tests were conducted to compare the effects of the interfaces on the *PQ*, *HQ-I*, and *HQ-S*. Each subscale consists of 7 items with a bipolar rating scale. We used 5 item scales and averaged the ratings of all 7 items for each subscale. Group differences for UX qualities *PQ*, *HQ-I* and *HQ-S* between the AR and DM interface condition are reported in Table 5.1 and Figure 5.11. The interface had a significant effect on all dimensions, with the AR interface leading to a significant lower score for the pragmatic (usability) dimension (with a medium effect size) but significantly higher scores for the hedonic dimensions (with large effect sizes).

We also assessed the participant's intrinsic motivation through the intrinsic motivation inventory [159]. Specifically, we employed the *IE* and *VU* subscales (5-point Likert scale). A two-tailed Wilcoxon signed-rank test indicated significant effect for AR ($MD : 0.86$) and DM ($MD : -0.29$) on IE ($W = 123, p < .05, r = .38$). There was no effect on VU (see also Figure 5.12).

	AR		***DM***				
Quality	M	SD	M	SD	t(13)	p	Cohens's d
PQ	-.08	.37	.28	.37	-2.58	.02	.37
HQ-I	.42	.37	-.15	.60	3.20	.005	.78
HQ-S	.6	.39	-.54	.67	7.58	6e-7	1.92

Table 5.1: Group differences for UX Qualities *PQ*, *HQ-I* and *HQ-S* between the *AR* and *DM* interface condition.

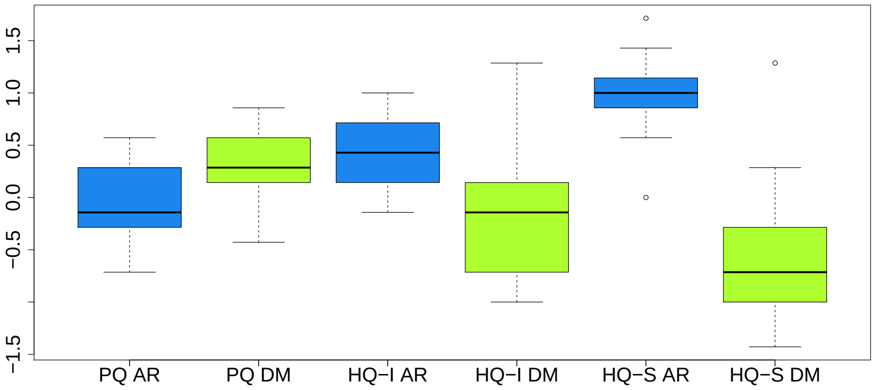


Figure 5.11: AttrakDiff scores for *PQ* (*PQ*), Hedonic Identity (*HQ-I*), and Hedonic Stimulation (*HQ-S*) on a 5-item bipolar scale.

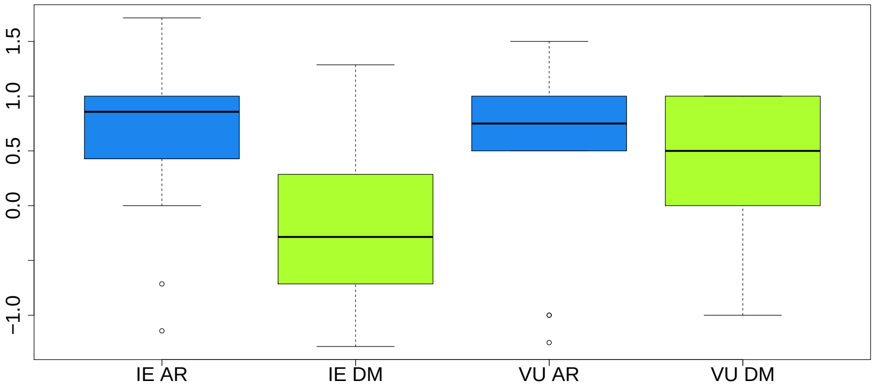


Figure 5.12: intrinsic motivation inventory scores for *IE* and *VU*.

5.3.4 Discussion

The results obtained with the proposed approaches for *SVBRDF* capture and user navigation demonstrate that it is possible to record different appearances of hologram patches with consumer hardware.

More specifically users were able to reach the six views used in the study with reasonable accuracy (maximum range of translation error from -8 to 10 mm, maximum range of rotation error from -8 to 8 degrees; see Figure 5.7). It must be noted that the used specimen banknote did not remain entirely planar during the study. Although potentially leading to larger errors, real banknotes often suffer from similar deformations. Consequently several users commented that final alignment was tedious and should be automated. This could be achieved by capturing several frames and selection of a reasonable trade-off between stability and alignment accuracy.

Patch similarity computed after registration gave *NCC* scores greater than 0.75 for four of the six views (see Figure 5.8). While the pixel-wise registration improved *NCC* scores noticeably, the obtained pose accuracy was close enough to the reference view that the appearance of the view-dependent elements was correct. Thus, while we need to automatically correct for small pose variances, the correct sector for the view-dependent appearance was usually selected.

On the task level, the *AR* system shows similar verification performance compared with a *DM*, but longer *TCTs* (see Figure 5.9) and higher physical and temporal demand (see Figure 5.10). This is reasonable, because users are forced to move to the right pose, which ensures repeatable conditions and reasonable matching of patches. However, none of the evaluated interfaces was able to provide clear evidence whether a hologram is real or not (see Figure 5.9). While the *AR* system was also not rated better in terms of usability (*PQ*), both the AttraktDiff (see Figure 5.11) and intrinsic motivation inventory (see Figure 5.12) questionnaires indicated significant higher ratings for hedonic dimensions and *IE*. This could indicate a higher motivational value for non-professional end users to employ the *AR* system for hologram verification.

Not being able to work with actual fakes in the study certainly limits insights concerning practical usability. However, credible fake documents and in particular holograms are difficult to produce or acquire. The challenge is to get hold of samples which are not immediately identified as fake but strongly resemble genuine items. This also means that simple photocopies are of limited value. However, the approach could be evaluated with holograms from different documents embedded in a generic looking surround or even with rotated or possibly thermally treated holograms. The latter would allow to gain more insights w.r.t. practical usability.

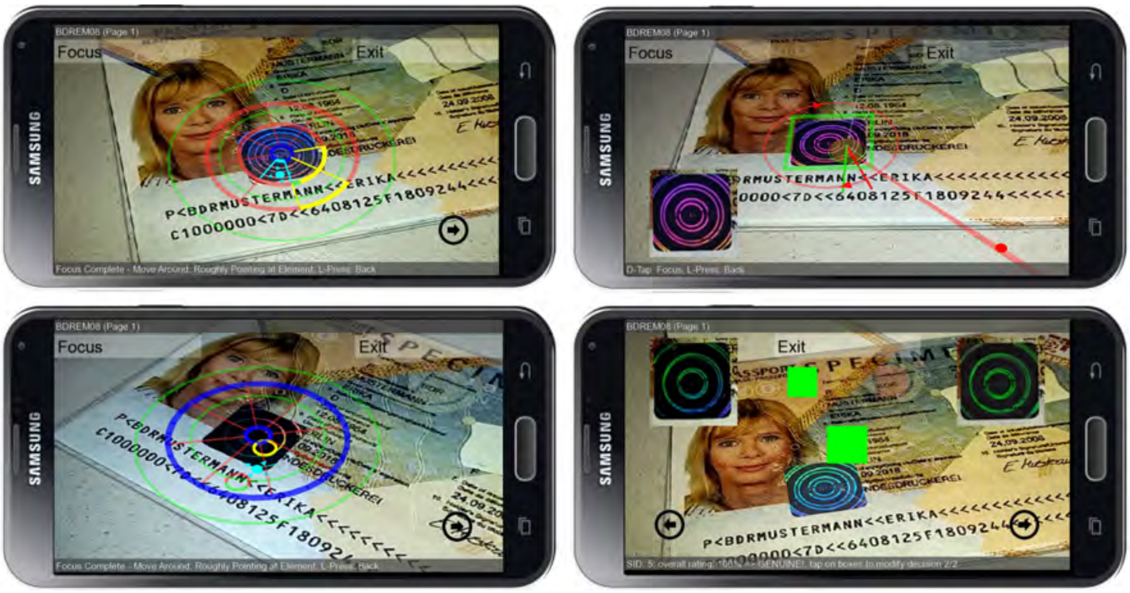
5.4 Towards efficient AR user interfaces for hologram verification

In the previous section, we investigated the feasibility of hologram verification with a user interface that requires users to align as accurate as possible with a given six *DOF* pose.

Within this section we investigate if and how the verification time can be improved. Specifically, we report on an improved alignment interface, introduce two new user interfaces which relax the accuracy demand of the alignment interface and finally compare those three interfaces (all three interfaces are depicted in Figure 5.13).

5.4.1 Revisting user guidance for hologram verification

In order to get reasonable input data for verification, the user should be supported throughout the image capture process. In general, manual interaction such as tapping on the screen



Focus Complete - Move Around: Roughly Pointing at Element. L-Press: Back

D-Tap: Focus. L-Press: Back

SID: 5; overall rating: 100% -> GENUINE!, tap on boxes to modify decision 2/2

Figure 5.13: User interfaces for hologram verification: Constrained navigation (top-left), alignment (top-right) and hybrid user interfaces (bottom-left) are designed, implemented and evaluated within a user study. They allow to reliably capture image data suitable for automatic verification. Results are presented to the user in a summary (bottom-right).

is not desirable for reasons of accuracy. Consequently, image capture should be triggered automatically, when the user is in a suitable position.

We observed that many holograms feature similar appearances in very different locations. Consequently, one could think of rejecting the entire pose information during matching and just taking care that the user is pointing towards the hologram. However, this does not seem feasible, since users cannot be expected to sample the hologram space (hemisphere) without guidance. This is backed up by an informal user study we conducted, in which a given hologram had to be sampled as deemed appropriate by the user. We obtained reasonable matching scores, but a very low spatial coverage. Users did not know when to stop the process. Although in case of originals, *TCT* could be reduced by an early-exit based on the matching score, this is not feasible with fakes. It is mandatory to consider the viewing direction in order to get a good coverage of the pose space and provide a reliable exit mechanism. Using information about the viewing direction when matching provides additional verification security.

An obvious approach is to guide the user to align the mobile device with exactly those view points, which are associated with the selected reference data. Alternatively, a portion of space can be visualized for sampling by the user, which requires coverage of a larger region instead of given positions. Combining both approaches leads to a hybrid variant, which uses a comparatively small region for sampling relevant data. In the following, we cover the design of these approaches in more detail. In favor of usability, we decided to omit an explicit check of the in-plane rotation of reference views during matching. This is motivated by the fact that when views are placed on a hemisphere, reasonable results can be achieved by just rotating the target.

5.4.2 Alignment interface

Sampling holograms can be treated as an alignment task, where users have to point at the center of the element, align with the viewing direction using iron sights, match the rotation along the viewing ray using a virtual horizon and take care of the recording distance [103]. Although this causes a lot of strain, we believe that a careful design in conjunction with automatic recording and matching, can lead to considerable gains in efficiency. This could make the alignment approach a strong competitor for constrained approaches.

We propose an improved alignment interface, which was designed in an iterative process involving continuous user feedback. A sketch of the elements involved can be seen in Figure 5.14. We observed that users often had trouble matching the overall orientation/rotation of the document and the device with the original approach. Not being able to do so, makes the overall alignment process more tedious. Consequently, the revised approach starts with coarse alignment of document-device orientation. We project the camera center and the top point of a reference pose down on the target surface and compute the relative angle as a rough indicator of initial alignment. This can be visualized as a color-coded indicator within a circle around the element. Depending on the sign of the computed error, arrows are placed on the circle to indicate the required movement of the target. Upon successful alignment (within a certain range), we proceed with more accurate indicators for the viewing direction. We use animated rubber bands as indicators for pointing at the element, but also for the vertical angle on the hemisphere. In both cases, the goal is to follow the animated arrows in order to shrink down the rubber band into a point (see Figure 5.15). Finally, a focus indicator is realized as a scaled sphere placed at the base point of the current viewing direction on the target. Animated, directed arrows indicate the required direction of movement. Note that we perform an initial focus operation at the first view to be aligned and keep this setting throughout the process.

Views are captured sequentially, with feedback on the overall progress of the operation. This aims to reduce visual clutter for the user. Upon successful alignment, several frames are recorded from the live-video stream and automatically matched against prerecorded reference data. From these measurements, the one having the highest matching score is selected as the result patch for the user. During the process, we provide guidance towards the desired direction, but also feedback regarding the quality of alignment. Similar to the previous approaches, we aim to minimize the required movements for the user by automatic selection of the nearest view. A live-view of the rectified hologram patch is constantly displayed during spatial interaction in order to provide visual feedback of the changes in appearance with varying recording position (see Figure 5.15 for an exemplary alignment sequence).

After recording each of the views, a summary including the current overall decision (genuine/fake) is presented to the user (see Figure 5.13). The user may skim through the captured views and compare them side-by-side with the expected reference data. If the system suggestion is revised by the user, an overall similarity score is recomputed, which eventually changes the final decision. The user may also re-record certain views in order to get a better basis for the final decision. This can be done in the summary for the current view and works for all the approaches described in this paper.

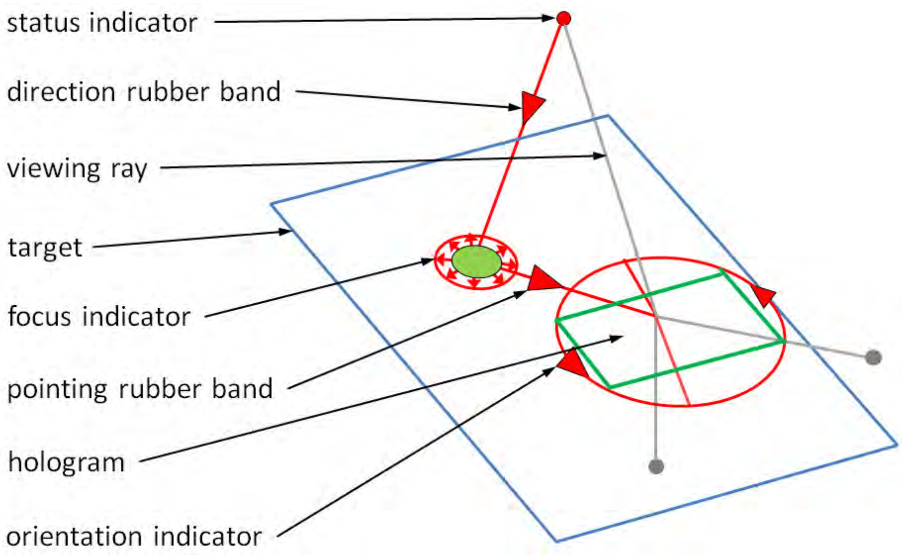


Figure 5.14: Geometry of the revised alignment approach. Matching takes place by alignment of target rotation and pointing with the indicator at the element. Finally the viewing direction is refined using the direction rubber band at an acceptable viewing distance.

5.4.3 Constrained navigation interface

The task can also be treated within a constrained navigation framework. The idea is to guide the user to sample larger portions of space instead of aligning with single views. By giving more freedom to the user, this can reduce workload and TCT.

The initial step is to guide the user to point at the hologram as required by the recording setup. We provide guidance using an animated rubber band, which shows a moving arrow, once outside a given radius away from the element (see Figures 5.16, 5.17). Then, the capture distance needs to be adjusted as a starting point for an auto-focus operation, so that the assumption about the flashlight being the dominant light-source holds. For this purpose, we scale the entire widget and require the user to adjust the distance, so that the outer ring of the widget stays within the given distance bounds.

Although the robot recording operates on a hemisphere, it does not seem reasonable to apply this concept directly. An augmented hemisphere would certainly lead to coverage of the entire space, but not necessarily in the shortest possible time. With an augmented hemisphere, the most obvious movement is to scan hull slices and then rotate the document for the next slide. We empirically verified that changing orientation from an orthogonal starting point (conic) is much faster than target rotation with slice-scanning.

In favor of efficiently treating both originals and fakes, the user should be guided towards different viewing directions or ranges. We propose a 2D orientation map (projection of the conic space) [107] for this task. It is divided into slices that are aligned on one or more tracks. The current position on the map is visualized by a cursor, and the current slice is also highlighted. The cursor position is corrected by the target orientation, so that the movement direction always corresponds to the orientation of the device (see Figures 5.16, 5.17). In general, it is not sufficient to just capture a single shot inside each

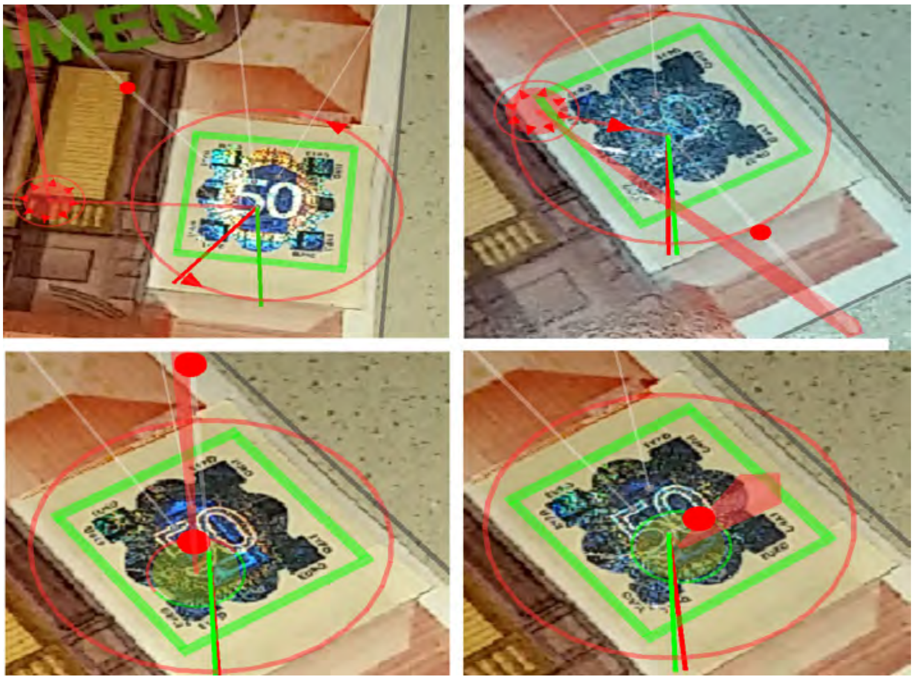


Figure 5.15: Exemplary alignment sequence: Not aligned (top left). Aligning target rotation (top right). Pointing at target (bottom left). Aligning viewing direction along hemisphere arc (bottom right).

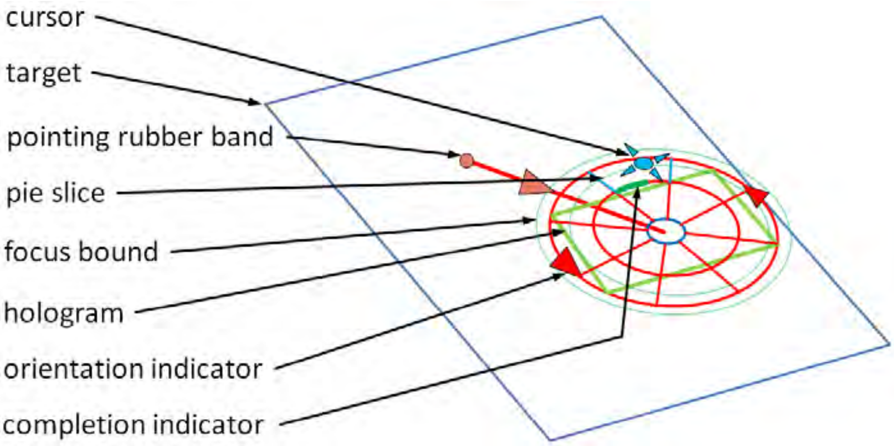


Figure 5.16: Geometry of the proposed constrained navigation approach for sampling the hologram. The user is guided to point at the element and a cursor is controlled by the 2D orientation on an augmented pie, divided into slices and tracks.

slice. We record several shots per slice, that differ at least by a given angle threshold. The exact amount is automatically calculated, taking into account the area of the slice. Consequently, the user can move freely inside the pie slices during the process. The tiny arrows around the cursor serve as movement indicators. Whenever the user remains static inside a non-completed slice, flashing arrows remind to move on. The upper arc defined

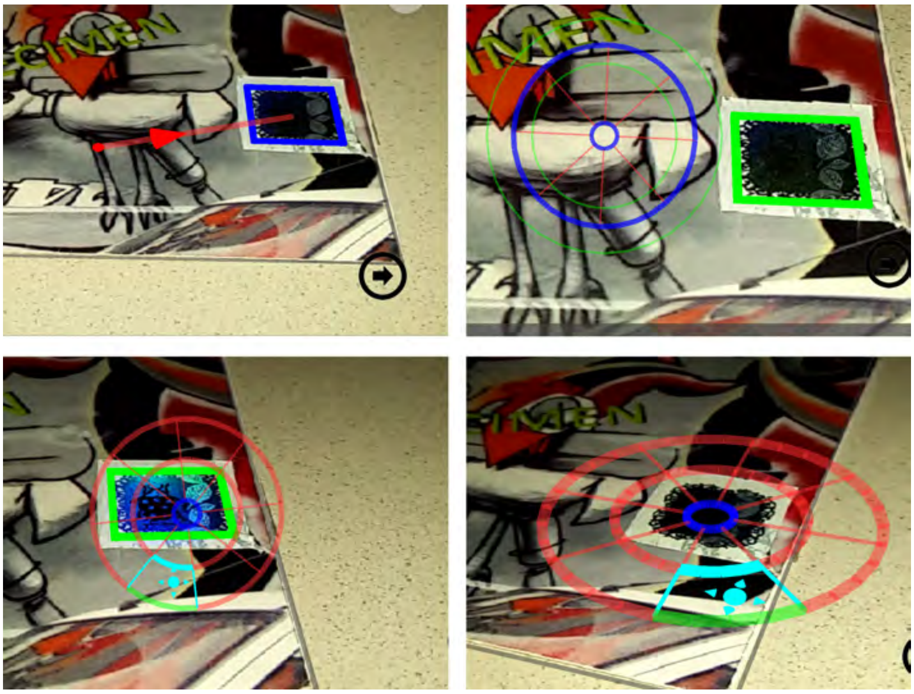


Figure 5.17: We guide the user to point at the element using an animated rubber band (top-left). Focus adjustment showing the layout of the orientation map and green distance bounds (top-right). Constrained navigation UI with pie slices (bottom-left). Augmentation directly onto the document/element (bottom-right).

by a (sub-)slice is used as a completion indicator, which switches from red to green with increasing slice coverage. The orientation map is realized as a widget placed in the screen plane (2D-CON) or augmented onto the target (AR-CON).

In a pilot study, we tried using either no visual information on the capturing procedure or a progress bar without any orientation information. Using no visualization at all gave the best completion time, but also the lowest spatial coverage. In the following, we dropped the interface without guidance and the progress bar. It must be noted that even with the AR-CON interface, not all participants sampled the entire hologram. Consequently, we went to incorporate slightly more guidance with the goal to only check pie slices containing a reference view (see Figure 5.18).

5.4.4 Hybrid interface

The location of reference views cannot be mapped straightforward to pie slices. It may be necessary to associate several pie slices with a single reference view, increasing the amount of slices to be checked. Since the number is generally much lower than the total number of pie slices, we use small regions on the augmented map around reference locations, which also serve as local completion indicators (AR-HYB, Figure 5.18).

These two UIs were evaluated in another pre-study, this time involving a demonstration phase. According to the results obtained, AR-HYB had a much lower TCT compared with

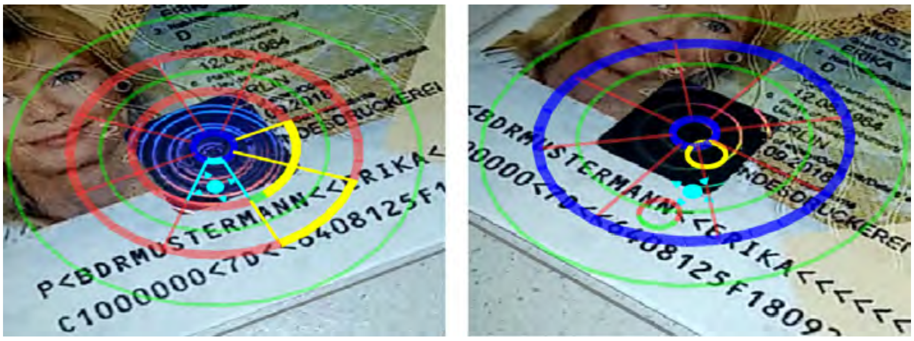


Figure 5.18: *AR* UIs with guidance for interesting subspaces. Either pie-slices (*AR*-CON, left) or circular regions (*AR*-HYB, right) are indicated for sampling by the user.

AR-CON. Users were able to complete the task using both approaches (perfect coverage of interesting slices/regions) and obtained reasonable patch-matching scores. Users generally gave very positive ratings concerning the type of guidance and overall usefulness of the application, with a clear preference for *AR*-HYB. Motivated by user demand and our own reasoning, this clearly moved the approach more in the direction of an alignment task. As we consider our informal studies only suitable for guiding the design process, we conducted a more detailed evaluation.

5.4.5 Evaluation

We evaluated the most promising candidate for constrained navigation (CON) and the hybrid approach (HYB, see Figure 5.18) against the alignment UI (ALI, see Figure 5.15). After image capture, a summary is presented to the user (see Figure 5.13) independent of the UI used for capture. The global system decision is communicated via a colored square (green...valid, yellow...unsure, red...invalid) to the user. Each reference has its own page, showing the reference data on the left side of the screen and the best recorded match on the right side, along with a local rating by the system, which can be changed by the user in case of doubt. It must be noted that we also monitored distance as capture condition, so that the users had to stay within the allowed distance range for the CON and HYB interfaces. We manually selected two reference views per hologram with a visually equal spatial distribution. We consider two views to be the minimum for verification of view-dependent elements.

5.4.5.1 Study design and tasks

According to a domain expert we consulted, professionals can identify most fake documents or holograms within a few seconds. The focus of the following study is on laymen without advanced domain knowledge or experience, using an off-the-shelf smartphone for hologram inspection. In contrast to our previous work [103], we do not compare a printed manual to an *AR*-System, but we seek to improve upon the long inspection time of *AR*-based hologram verification.

We designed a within-subjects study to compare both the performance and user experience aspects of the three aforementioned user interfaces for hologram verification.

The study had two independent factors: interface and hologram. The independent variable of main interest was interface (with three levels: ALI, CON, HYB). We modeled hologram as fixed effect (four level), since the holograms were deliberately selected (and not randomly sampled from a population) in order to represent intensity-dominated and shape-dominated samples including common mixtures.

For each of the four holograms, we selected the corresponding reference views with the goal of minimizing the variance an individual hologram could have on the results. Dependent variables of interest were TCT (both capture and decision time), system performance (how well the system could verify the validity of the hologram), user performance (how well the user could verify the validity of the hologram), and user experience measures (usability, workload, hedonic and motivational aspects).

For each interface, the actual verification procedure started upon pointing the center of the screen at the element and tapping on it. For the ALI interface, the user had to align the rotation of the document with the current reference view (azimuthal angle), point at the center of the hologram and adjust the viewing direction (polar angle) along with the capture distance. In case of the CON interface, the user had to point at the element, following the base rubber band. Then, the orientation cursor had to be moved inside the indicated (connected) pie slices by changing the azimuth and inclination angles through device movement and monitoring the operating distance. The HYB interface had to be operated in a similar way. However, the cursor had to be aligned and moved inside small circular regions. Upon successful sampling, the system summary/system decision was presented to the user.

5.4.5.2 Apparatus and data collection

We conducted the study in a lab with illumination from the ceiling enabled (fluorescent lamps). In order to minimize variations induced by daylight changes, we kept the blinds of the room closed throughout the entire study.

All user interfaces were integrated into a single Android application running on the Samsung Galaxy S5 mobile phone (Android 4.4.2) and using the built-in camera with LED flashlight enabled. Reference data for verification was recorded with a robot using the same device.

We used four holograms as shown in Figure 5.19, each on a different base document. With our choice of samples and reference data, we aimed to address the non-trivial case of hologram substitution, since that is rather common according to a document expert we consulted for our study. Although some of the views we selected (i.e., black patches) may not resemble the typical appearance of holograms for the public, we believe that the large visual difference w.r.t. the other image in the pair justifies their use.

We collected data for evaluation through automatic logging on the test device itself, questionnaires and interviews. For data analysis, we used Matlab, R, and SPSS. Null hypothesis significance test were carried out at a 0.05 significance level, if not otherwise noted.

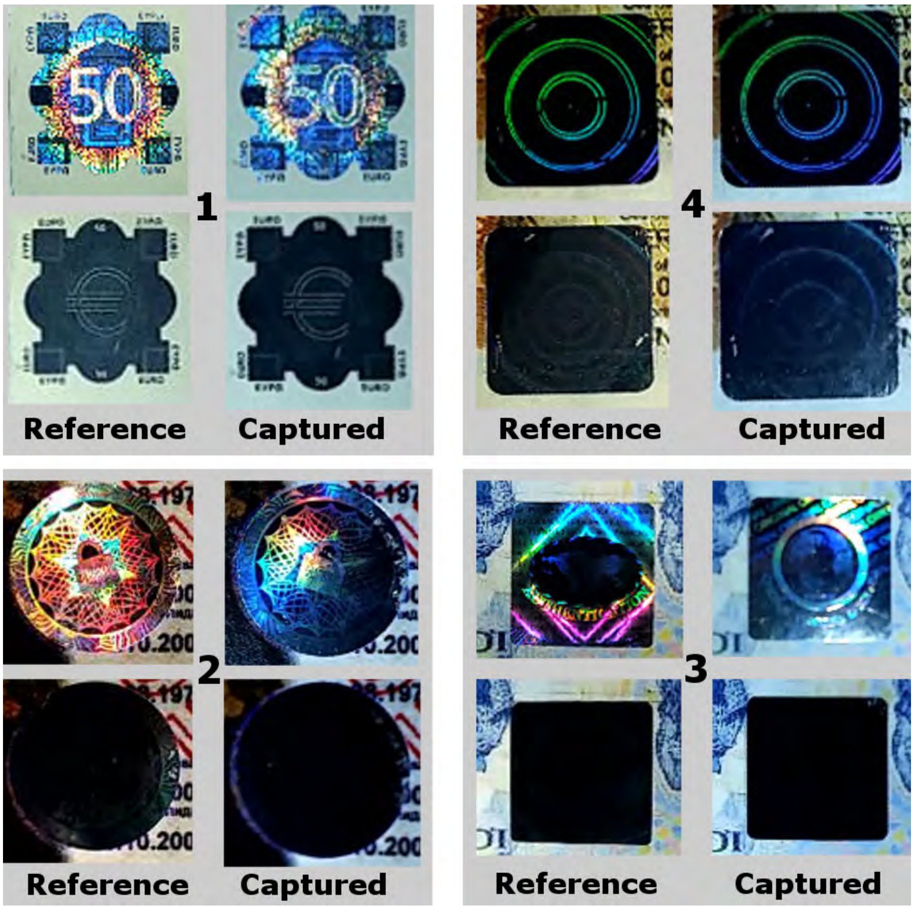


Figure 5.19: Samples used in our study. We evaluated all user interfaces with two original (no. 1, 4 - top row) and two fake (no. 2, 3 - bottom row) holograms, where each was placed on a different document template. Reference information recorded with the robot setup is used by the system for matching, while the other images are exemplary recordings during verification by the user.

5.4.5.3 Procedure

Each participant was informed about the study purpose and the approximate length prior to the start of the study. The participants filled out a demographic questionnaire and then conducted the Vandenberg and Kuse mental rotation test [248]. They were informed that they would test a total of 12 holograms with three user interfaces (four holograms per interface). Although 12 holograms were shown to the participants as a stack, only a subset of four holograms was used for all interfaces (see Figure 5.19).

The following procedure was repeated for all three user interfaces. A training phase with both a correct and a fake document (not appearing in the actual study) was conducted. This also included an explanation of application controls along with document classification and tracking. Participants could test the interface as long as they liked (on average less than five minutes). After feeling comfortable with the interface, participants were asked to use the current interface to capture four holograms, one at a time. After capturing a single hologram, the system presented its decision on the validity of the single views and an overall decision (valid, unsure, invalid). After seeing the system decision, the participants were asked to fill out a post-task questionnaire, in which they were asked to assess the validity of the hologram on their own (5-item bipolar scale: I am totally sure that the hologram is fake ... neutral ... I am totally sure that the hologram is valid). After validating four holograms with the current interface, the users filled out a post-interface questionnaire (5-item Likert scale, ease-of-use and time items of the After Scenario Questionnaire [149]), the NASA TLX questionnaire (with weighting of items) [101], the AttrakDiff [106] and Intrinsic Motivation Inventory questionnaires [159].

After having conducted this procedure for all three interfaces, the participants filled out a final questionnaire, in which they should choose their preferred interface (overall preference, which interface was fastest to use, which interface was easiest to use). Finally, they were asked about the reasons for their choices. Participants received a voucher worth 10 EUR for their time.

The starting order of both interface and hologram was counterbalanced. The tasks where blocked by interface. While each participant was exposed to each hologram three times we took care to make them believe it was a separate hologram (by showing a staple of several documents and hiding from them which document was drawn out of the staple). Also each participant was exposed to individual interface-hologram combinations exactly once during the study. The whole procedure took on average 90 minutes. Participants could take a break anytime they wanted.

5.4.5.4 Participants

19 volunteers (2 female, age $M = 26.8, SD = 4.46$) participated in the study. All except one participant owned at least one smartphone or tablet, where the majority (16) had been using it for at least one year. In general, participants reported to be interested in technology. Thirteen participants had already used an AR application at least once. Seven participants had never attempted to verify a hologram before. In the mental rotation test, the majority of participants scored reasonably ($M = 0.8, SD = 0.14$). With 19 participants assessing 4 holograms with 3 interfaces, we obtained 228 samples.

5.4.5.5 Hypotheses

Based on our observation and the insights gained during pre-studies, we had the following hypotheses: *H1:* The hybrid UI will be the fastest among all interfaces. *H2:* The alignment UI will be the most accurate one, but slow. *H3:* The constrained navigation UI will be the easiest to use.

The hybrid interface combines desirable elements from alignment (accurate end position) and constrained navigation (marked interaction space). With a small number of reference views, checking should be very fast (*H1*). The revised alignment interface should assure the most accurate capture positions and consequently has the best prospects for accurate matching and verification (*H2*). This might come at the cost of increased capture time. The constrained navigation approach gives most freedom to the user. The pie slice layout could be familiar to users, although accuracy w.r.t. single reference views might not be as good and by design, a bigger space needs to be sampled (*H3*).

5.4.5.6 Findings

We performed an analysis of TCT, user and system performance and user experience aspects for hologram verification.

Task Completion Time For capture time (the time from start of the task until the presentation of system results), a two-way within-subjects analysis of variance showed a significant main effect for interface, $F(2, 36) = 3.60$, $p = .038$, $partial\ \eta^2 = .17$ and a significant main effect of hologram, $F(3, 54) = 4.04$, $p = .012$, $partial\ \eta^2 = .18$. The interaction between interface and hologram was not significant.

Multiple pairwise post-hoc comparisons with Bonferroni correction for interface revealed that the mean score for capture time (in seconds) for the hybrid interface ($M = 37.22, SD = 38.20$) was significantly different compared to alignment ($M = 57.01, SD = 55.77$) ($t(75) = 3.44, p = .001$), but not compared to constrained navigation ($M = 44.43, SD = 20.70$). Also, there was no significant difference between constrained navigation and alignment.

Multiple pairwise post-hoc comparisons with Bonferroni correction for hologram revealed that the mean score for capture time (in seconds) for hologram 2 ($M = 39.61, SD = 29.39$) was significantly different compared to hologram 4 ($M = 55.19, SD = 53.51$), $t(56) = -3.23, p = .002$, but not compared to hologram 1 ($M = 45.58, SD = 40.97$) or hologram 3. Also there were no other significant differences between holograms. Furthermore, there where no learning effects for either interface or hologram as indicated by a within-subjects analysis of variance.

The decision time (the time spent in the summary screens) over all interfaces was on average 18.45 seconds ($SD = 15.32$). A two-way within-subjects analysis of variance showed no significant main effect for interface but for hologram $F(3, 54) = 3.233, p = .029, partial\ \eta^2 = .152$. However, multiple pairwise post-hoc comparisons with Bonferroni correction for hologram did not indicate any significant pairwise differences (hologram 1 $M = 17.7, SD = 12.72$, hologram 2 $M = 23.7, SD = 19.54$, hologram 3 $M = 15.53, SD = 8.85$, hologram 4 $M = 16.98, SD = 17.54$). The interaction between interface and hologram was not significant.

To summarize, the capture time using the hybrid interface was significantly faster than the alignment interface and for hologram 2 compared to hologram 4. For decision time, no pairwise significant differences could be found. There were no learning effects for interface or hologram.

User and system performance Over all participants and holograms, 79.6% of the users' decisions were correct (treating both items 'I am totally sure that the hologram is [in]valid' and 'I am sure that the hologram is [in]valid' as correct answers). For 12.5% of the decisions, the users where unsure if the hologram was valid or fake. An investigation of the effects of the predictors interface and hologram on the dichotomous dependent variable 'correctness of user decision' using logistic regression was statistically not significant. Note that we had to exclude one participant from this sub-evaluation due to incomplete data.

73.1% of the system decision were correct. The system was unsure if the hologram is valid or fake in 11% of all cases. As for user decision, we used logistic regression to investigate the effects of interface and hologram on the dichotomous dependent variable 'correctness of system decision'. The logistic regression model was statistically significant $X^2(5) = 58.83, p < .0001$, explained 37.5% (Nagelkerke's R^2) of the variance in system decision and correctly classified 81.5% of the cases. The Wald criterion demonstrated that hologram made a significant contribution to prediction ($Wald\ X^2(3) = 20.80, p < .0001$), but interface did not. The system only made correct decisions in 50.0% for hologram 1 (neutral: 27.8%, hologram 2 correct: 100%, 3 correct: 94.4%, 3 neutral: 0.04%), 4 correct: 74.1%, 4 neutral: 13.0%).

To summarize, users were able to correctly validate (decide if the hologram is valid or false) in 80% of the cases, but the system only in 73%. Hologram was a significant predictor for system decision with a validation performance for hologram 1 of only 50.0%.

User experience We investigated ease of use and satisfaction with task duration with the ASQ, cognitive load with the NASA TLX, and hedonic and motivational aspects with AttrakDiff and Intrinsic Motivation Inventory questionnaires, after each participant had finished using a single interface.

A one-way Friedmann ANOVA by ranks did not indicate a significant effect of interface on ease-of-use. Similarly, for satisfaction with task duration (over all 4 holograms per interface), there was no significant effect of interface. Note that we had to exclude one participant from this sub-evaluation due to missing data.

For cognitive workload as measured by NASA TLX, one-way Friedmann ANOVAs by ranks did not indicate significant effects of interface on the subscales (mental demand, physical demand, temporal demand, performance, effort, frustration) or the overall measure. Due to space reasons and the non-significance of the omnibus tests, we will not report further statistics here.

Similar, for *PQ*, *HQ-I* and *HQ-S*, as measured by AttrakDiff, and for value-usefulness and interest-enjoyment as measured by the Intrinsic Motivation Inventory, one-way Friedmann ANOVAs by ranks did not indicate significant effects of interface.

In the final questionnaire, 47% of the participants indicated that CON was easiest to use (ALI: 21%, HYB: 32%), 42% indicated that CON was fastest to use (ALI: 16 % HYB: 42%) and 47% favored CON overall (ALI: 26.5%, HYB: 26.5%).

In summary, the statistical analysis could not indicate significant effects of the interfaces on usability, workload, hedonic qualities or intrinsic motivation. Still, about half of the participants preferred CON overall and indicated that it was easiest to use.

5.4.6 Discussion

Our analysis did not fully confirm hypothesis H1. The hybrid interface was the fastest one, taking roughly 40 s for image capture, being significantly faster than the alignment interface (which took around one minute for verification). However, the hybrid interface was not significantly faster than the constrained navigation interface (ca. 45 s).

While this is a significant improvement over related work ([103], but using up to six views), this is still a long time span and probably not feasible for a quick check in a real-world situation. However, as most checked documents will be originals, an early exit for such samples could further decrease checking time. As decision time did not vary significantly between interfaces, they are all suited to recording data for verification.

Around 73% of the system decisions were correct, which may seem rather low. As there was no significant effect of any interface, hypothesis H2 does not hold in this regard. If we only neglect wrong decisions (i.e., combine positive and neutral decisions), the system performance would still be below the combined rate for user decisions (system: 84% correct vs. user: 92% correct). It seems that users either came up with their own (more invariant) similarity measure during the study, or they used additional appearance information gathered through the sampling process for their decisions, which was not available to our system (e.g., due to non-matching viewing direction). However, most of the neutral system decisions (around 63%) were caused by hologram 1 (50 EUR banknote, see Figure 5.19). This hologram shows rainbow colors, which is a very difficult case for our matching approach. Together with the rather conservative parametrization of our system (avoiding false positives), and the encouraging results of hologram 4 (around 90% combined rate), we speculate that the type of hologram has considerable influence on its verifiability with the proposed approach.

While the statistical analysis did not indicate significant effects of interface and user experience measures, we obtained a large number of comments in the post-hoc interviews throughout the study. The HYB UI, being the fastest one, was described four times as 'intuitive', 'good to use' or 'easy' (CON: 7, ALI: 3). However, four participants reported that the movements required were initially not clear (CON: 4, ALI: 5). With the CON UI, four users recognized the freedom in movement. For the slow ALI UI three users expressed their interest in that UI ('interesting', 'cool idea', 'visually best'). One user stated that it was 'easy to spot, what to do, but difficult to accomplish'. Two users also gave positive comments about the usefulness of the summary.

For the CON and HYB interface, one user suggested to always display the pointing rubber band, even when the widget is perfectly at the screen center. For CON, one user suggested an additional completion indicator for pie slices involving the pie region itself instead of the border. The same user also suggested to use additional indicators for viewing ray alignment in the ALI UI.

Despite being the fastest one (around 40 s), the hybrid user interface did not receive the same degree of user consent as the CON interface when taking into account the comments. Users explicitly criticized the final alignment stage involved. As a take away, it seems that the most efficient interface does not necessarily reflect the general preference of the user. Such awareness should be considered for real-world deployments of mobile *AR* user interfaces requiring fine-grained maneuvering.

5.5 Summary

Within this chapter we investigated the utility of *AR* user interfaces for small information surfaces. As an instance of such user interfaces we focused on security documents.

First, we investigated if capturing and checking view-dependent elements using off-the-shelf mobile devices is possible at all. We demonstrated that, given a mobile phone has a built-in flashlight, we can create a system that guides users to reference poses on a security document and subsequently allows to compare a hologram picture taken from this view with a pre-recorded reference image. A user study indicated that the initial alignment-based approach resulted in comparable verification performance as a baseline printed guide. However, this alignment-based approach resulted in a significantly longer verification time and in significantly larger mental strain compared to the printed guide.

Hence, we iteratively explored the design of several further user interfaces for checking holograms, with the goal to considerably reduce verification time for the user. Specifically, we re-designed the alignment interface to have both a coarse and a fine alignment phase, introduced a constrained navigation and a hybrid approach. All three approaches could be combined with either manual checking (as in the initial prototype) or with automatic matching. An additional user study indicated that, although the hybrid interface had the fastest completion time, users preferred the constrained navigation interface over the other two. As each of the interfaces served equally in the capture of verification data, the choice of the user interface (constrained navigation or hybrid) might depend on user preference or previous training. Still, verification performance of the system should be improved, which this left for future work.

6 Interaction with Public and Personal Electronic Displays

Contents

6.1 Facilitating AR at public displays 150

6.2 Multi-fidelity interaction with displays on and around the body154

6.3 Summary 170

So far, we have investigated printed information surfaces. In this chapter, we will investigate interaction with digital surfaces instead.

First, we will look at how augmentations of large information surfaces in public space such as digital signage systems can be facilitated. Previous systems for interaction with large public displays have been limited to only basic interaction techniques on self-contained mobile devices, or they have required considerable infrastructure for large screen interaction, making them impractical outside the lab. In the first part of this chapter, we introduce an interactive system called HeadLens, which addresses several limitations for interaction between mobile devices and situated displays. Firstly, HeadLens demonstrates how only access to a screencast of the situated display is sufficient to create personal augmentations. This can be easily provided through common streaming platforms. Secondly, our system provides *AR* interaction using the *ML* metaphor between situated displays and mobile devices with geometrically correct rendering from the user's point of view. Our system is self contained and performs all computations on the mobile device. Hence, it easily scales to multiple users.

In the second part of this chapter, we turn our focus from large public to small private displays. These display devices on and around the body such as smartwatches, head-mounted displays or tablets enable users to interact on the go. However, diverging input and output fidelities of these devices can lead to interaction seams that can inhibit efficient mobile interaction, when users employ multiple devices at once. We introduce MultiFi, an interactive system that combines the strengths of multiple displays and overcomes the seams of mobile interaction with widgets distributed over multiple devices. A comparative user study indicates that combined *HMD* and smartwatch interfaces can outperform interaction with single wearable devices.

6.1 Facilitating AR at public displays

Today's handheld devices allow for a decoupling of public and private interaction with digital displays and for interaction at a distance, which can be of great value when interacting with public displays. However, while *ML* interaction with situated information, such as printed posters or digital displays, has been explored for over 20 years [245], to date, it failed to move out of the lab. With the proliferation of large format screens and handheld devices, "second screen" apps for handheld devices providing background information for live TV programming are becoming increasingly popular. Spatial interaction between handheld and situated displays should be the obvious next step.

We believe that the major obstacle preventing spatial interaction between mobile and situated displays is the need for additional infrastructure. Previous attempts at showing perspectively correct overlays from the user's point of view have required stationary outside-in 3D tracking, often in combination with projectors. Such proof-of-concept implementations do not allow mobile operation outside the lab.

We contribute the first system that allows user-perspective magic lenses for situated displays, showing dynamically changing content at interactive frame rates without the need for additional infrastructure. This system, HeadLens, only needs access to a remote streamcast of the display content and is otherwise self-contained. It utilizes *NFT* of the screen content and 3D face tracking on mobile devices with dual camera access. Consequently, HeadLens brings *ML* interaction to arbitrary situated displays such as public information screens and easily scales to multiple user (Figure 6.1).

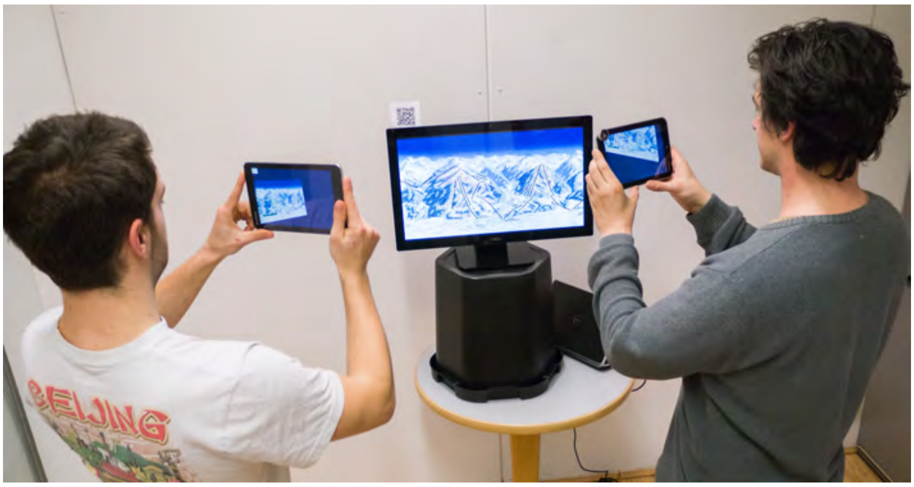


Figure 6.1: HeadLens requires only access to a remote screencast and is otherwise self-contained, making it suitable for multiple users

6.1.1 User perspective ML approach

User-perspective magic lenses create two challenges. First, the pose of the mobile device relative to the situated screen content has to be tracked with six *DOF*. This can be challenging due to the dynamic nature of content on the situated screen. Second, the

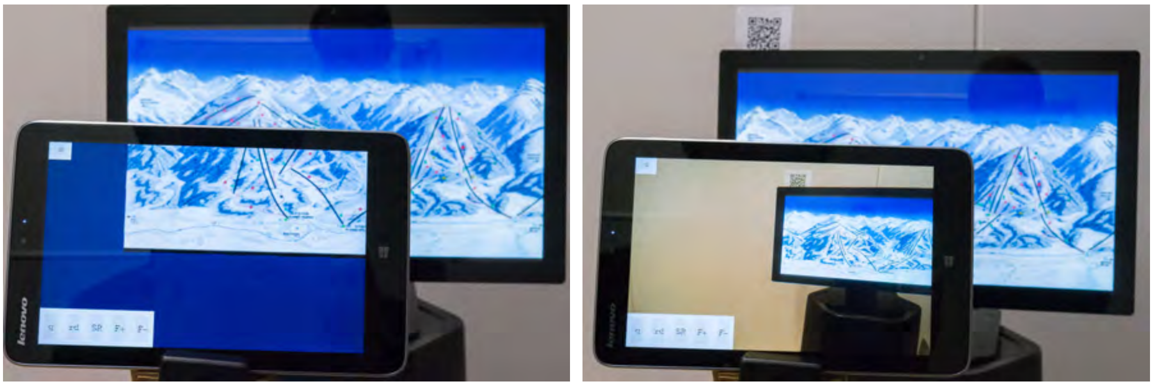


Figure 6.2: (left) Simultaneous 3D head tracking and *NFT* enable user perspective magic lenses for situated display content. (right) Device perspective rendering usually found in handheld Augmented Reality devices.

pose of the user's head relative to the mobile device needs to be known to enable the rendering of the scene from the user's viewpoint. Tracking with six *DOF* on handheld devices has been an important research topic in *AR*. *NFT* [252] and even *SLAM* are now becoming commonplace on handhelds [135]. Since the tracking of dynamic screen content precludes the use of fiducials or static *NFT*/*SLAM* models, a dynamic *NFT* approach is required. The *NFT* model used by the handheld must be re-initialized if changes on the screen exceed a threshold, which can be determined by image differencing from the screencast. The *NFT* model is then used to track the situated display in the back-facing camera stream of the handheld. For tracking the user's head, we can exploit the fact that the user needs to face the handheld in order to look at the screen. Consequently, we can use the input from the front-facing camera for non-rigid face tracking [244]. Figure 6.2, left, shows user-perspective rendering achievable with our approach, compared to device perspective rendering (Figure 6.2, right).

6.1.2 Implementation

We implemented the HeadLens algorithm as depicted in Figure 6.3. It runs self-contained on mobile devices with dual camera access, with OpenCV for tracking, OpenSceneGraph for 3D rendering and FFMPEG for screencasting. Initializing Situated Display Interaction To enable interaction between a mobile device and a situated screen, HeadLens needs access to a screencast and a measurement of the physical extent of the screen. In our prototypical implementation we employed FFMPEG to capture and wirelessly screencast content at FullHD resolution (1920x1080) via MPEG-TS. Knowledge of the screen extent is necessary to allow metric estimation of the camera pose. The meta information (URL for screencast via RTSP, display extent) is communicated to the client via a QR code attached to the bezel of the situated display or temporarily displayed on the display itself.

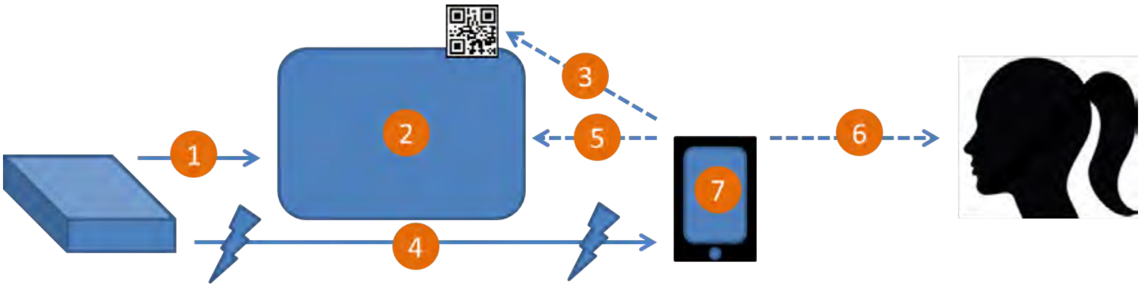


Figure 6.3: Overview of the HeadLens algorithm. (1) A content source such as a PC sends a video signal to a situated display. (2) The situated display shows the corresponding image. (3) The handheld device decodes a QR code to determine the screencast channel. (4) A screencast hardware or software multicasts the video signal to a wireless network. (5) The handheld device tracks the location of the situated display with the back-facing camera (6) and the location of the user's face with the front-facing camera. (7) Finally, the handheld device displays user-perspective *AR* content.

6.1.2.1 Tracking screen content

For tracking screen content we employ *NFT* from FAST keypoints, ORB descriptors for detecting keypoint correspondences and a patch tracker based on normalized cross-correlation for incremental tracking. Depending on the computational capabilities of the mobile device, there are several choices for re-initialization when the display content changes. The simplest approach re-initializes the tracking system for every frame. However, this is wasteful as it ignores frame-to-frame coherence. A more efficient approach uses fast image differencing and updates only parts of the *NFT* model that have significantly changed. Re-initialization may also be indicated if the *NCC* score of the patch tracker falls below a threshold. Generation of a new or updated *NFT* model is computationally expensive and introduces unwanted latency, potentially resulting in jerky motion or intermittent loss of tracking. The generation process is therefore masked by interleaving it with tracking the current *NFT* model, and amortizing the generation of the new *NFT* model over several frames.

6.1.2.2 Head tracking

For tracking the user's face, we combine a 2D deformable face tracker with a solver for the perspective-n-point problem. In a first step, 2D image points of facial landmarks are estimated using deformable model fitting [213]. For the second step, we use a rigid 3D model which is mapped to selected image points of the 2D model (eyes, nostrils, temples). The origin of the model is situated between the user's eyes. The estimation of the pose of the front facing camera given these 3D-to-2D point correspondences is achieved via EPnP [147] with a pose prior (front facing head at 40 cm distance).

6.1.2.3 Rendering

Rendering the screen content from the user's perspective is done in two stages. In the first stage, the screencast image delivered via MPEG-TS is written to an OpenGL texture buffer and texture mapped to a quad, which is positioned according to the tracking data of the back-facing camera. In the second stage, this scene is rendered to texture from a camera position estimated by the 3D head tracker, while only considering the translation and assuming an orientation towards the screen center. In the main rendering pass, we use projective texture mapping from the same viewpoint to map the previously captured scene onto a screen-aligned quad, which has the extent of the mobile device screen. Calibration Approximated user-perspective rendering can be achieved given the approach described above. However, to ensure the best overlay quality, further adjustments should be considered. First, the intrinsics (i.e., focal length, principal point and skew) and the relative poses of the back and front facing cameras should be estimated. Intrinsics can easily be computed by standard camera calibration methods. The relative poses of the cameras towards each other and the screen center can either be estimated manually by measuring the offsets between the camera centers and the screen center or more accurately by multi-camera calibration [42]. Finally, the landmark points on the user's face should be calibrated, so that the distance from device to head can be correctly measured, and the focal length of the virtual camera used in rendering can be scaled accordingly. In practice, we use the interpupillary distance as a scale parameter. If the interpupillary distance is unknown, for example, when multiple users share a device, the focal length of the virtual camera must be set manually.

6.1.2.4 Sample application

We implemented a prototype application, which resembles public display content typically found in modern skiing resorts. A situated display shows the status of ski slopes as well as weather information. Passers-by can initialize interaction by pointing their mobile device towards the QR code attached at the situated display. The display visualizes their personal lift rides and the total number of kilometers they have skied. Screencast via multicast networking allows an arbitrary number of users to interact concurrently with the situated display without performance degradation.

6.1.3 Performance

Our prototypical implementation runs at approximately 6fps on a Lenovo Miix 2 tablet with Intel Atom Z3740 processor, 2GB RAM, Intel HD4000 GPU, 1280x800 px resolution and 640x480 px front + back camera resolution. On a Dell XPS12 tablet with Intel i7-3667U processor and 8GB RAM, Intel HD4000 GPU at 1920x1080 px resolution, the implementation runs with 13 fps. The average runtime performance of individual modules are as follows for the Lenovo tablet (Dell in parentheses): *NFT* detection: 50 ms (10 ms) for initialization, *NFT* patch tracking: 15 ms (4 ms), 2D face tracking: 40 ms (17 ms), 3D head pose estimation: 0.4 ms (0.12 ms), all of which happen in the scenegraph event traversal. The render traversals take 100 ms (50 ms).

6.1.4 Limitations and improvements

Our system can be improved in several ways. To increase the robustness of the *NFT* system, alternative keypoint descriptors or more iterative predictive tracking approaches

[252] can be incorporated. Also, we currently do not employ all facial feature points provisioned by the 2D face tracker. By learning new face models targets at runtime, we could drastically reduce the number of tracked 2D points, thus increasing performance. Another limitation concerns the physical setup of the front facing cameras in some handheld devices. If the *FoV* of the camera is too narrow or the camera is placed at a corner of the bezel, the face tracker might not be able to track faces centered in front of screen in typical interaction distances (30-50 cm). In this case, an adhesive wide-angle lens could be used.

In addition to performance-related constraints, we also plan to improve the interaction capabilities of HeadLens. A significant limitation of any *NFT* tracking system is its reliance on observing a sufficient number of discriminative feature points. When the handheld is brought close to or even rests on the situated display surface, the handheld's camera observes only a very small area (20x20 pixels for direct contact in our test setup). This is not sufficient for tracking from arbitrary screen content. However, if we allow a tightly synchronized feedback channel, the situated display could show imperceptible fiducial patterns [259]that allow tracking even at minimal distances. Moreover, user-perspective rendering of digital screen content could be combined with real-world content acquired through *SLAM* or image-based warping. This would allow to extent the spatial scope of interaction beyond the boundaries of the situated displays.

6.1.5 Conclusion

With HeadLens, we demonstrated that bringing rich *ML* interaction with situated displays into the real world is feasible and allows interactive frame rates. Our key observation is that HeadLens is solely dependent on lightweight screencasting and "second screen" channels. All time-critical computations are performed locally on the handheld. Consequently, HeadLens overcomes the need for special infrastructure and naturally scales to multiple users. In the future, we want to extend HeadLens for ad-hoc multi-display interaction, enabling rich interaction spaces beyond a single display.

6.2 Multi-fidelity interaction with displays on and around the body

In this part we turn away from large public screen towards small personal displays.

Personal, public and ambient displays form a pervasive infrastructure around us. However, displays are typically unaware of each other and make little attempt to coordinate what is shown across them. The emergence of second-screen applications, screen mirroring and remote desktop access demonstrates the benefits of suitably designed coordination. In particular, when users carry multiple displays on and around their body, these displays form a space that can be leveraged for seamless interaction across display boundaries.

In this work, we introduce MultiFi, a platform for implementing user interface widgets across multiple displays with different fidelities for input and output. Widgets such as toolbars or sliders are usually specific to a single display platform, and widgets that can be used between and across displays are largely unexplored. This may come from the problems introduced by the variations in fidelity of input and output across devices. For input,

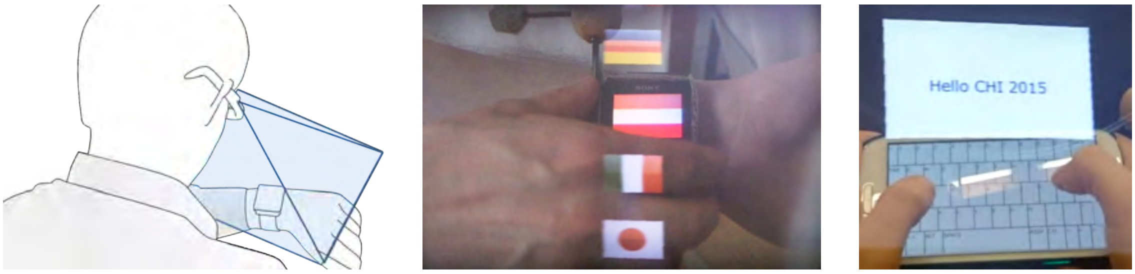


Figure 6.4: MultiFi widgets crossing device boundaries based on proxemics dimensions (left), e.g., middle: *ring menu* on a smartwatch with *HMD* or right: *soft keyboard* with full-screen input area on a handheld device and *HMD*.

we must accommodate different modes and DOF. For output, we must accommodate for variations in resolution and FoV. Both input and output affect the exactness of the user experience. Moving across devices can make the differences in fidelity apparent and introduce seams affecting the interaction.

MultiFi aims to reduce such seams and combine the individual strengths of each display into a joint interactive system for mobile interaction. For example, consider continuous navigation support, regardless of where a person is looking. Such navigation may employ a range of worn, handheld or embedded displays. Even if the navigation system is capable of switching among displays in a context-aware manner, the user will still need to contend with varying and uncoordinated fidelities of interaction.

MultiFi addresses the design problem of "interaction on the go" across multiple mobile displays with the following contributions: 1) We explore the design space of multiple displays on and around the body and identify key concepts for seamless interactions across devices. 2) We introduce a set of cross-display interaction techniques. 3) We present empirical evidence that combined interaction techniques can outperform individual devices such as smartwatches or head-mounted displays for information browsing and selection tasks.

Unlike prior work, we focus on the dynamic alignment of multiple body-worn displays, using body motion for spatial interaction.

6.2.1 Interaction by dynamic alignment

MultiFi aims to reduce the access cost of involving multiple devices in micro-interactions by dynamically leveraging complementary input and output fidelities. We propose *dynamic alignment* of both devices and widgets shown on these devices as an interaction technique.

Dynamic alignment can be seen as an application of proxemics [82]: Computers can react to users and other devices based on factors such as distance, orientation, or movement. In MultiFi, dynamic alignment changes the interaction mode of devices based on a combination of proxemic dimensions. We focus on *distance* and *orientation* between devices. However, different alignment styles can be explored, which are location-aware, vary between personal and public displays or consider movement patterns.

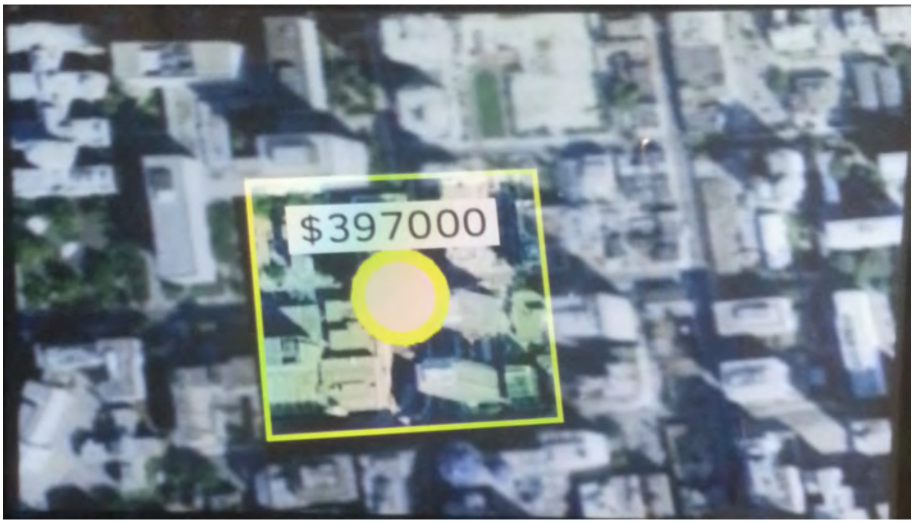


Figure 6.5: The extended screen space metaphor for showing a high resolution inlay of a map on smartwatch inside a low resolution representation on a *HMD*.

6.2.1.1 Design factors

To better understand the design implications of dynamic alignment, we begin with a characterization of the most relevant design factors determined throughout the iterative development process of MultiFi.

Spatial reference frames encompass where in space information can be placed, if this information is fixed or movable (with respect to the user) and if the information has a tangible physical representation (i.e., if the virtual screen space coincides with a physical screen space) [65].

Direct vs. indirect input. We use the term direct input, if input and output space are spatially registered, and indirect input, if they are separated. As a consequence of allowing various spatial reference frames, both direct and indirect input must be supported.

Fidelity of individual devices concerns the quality of output and input channels such as spatial resolution, color contrast of displays, focus distance, or achievable input precision. We also understand the display size as a fidelity factor, as it governs the amount and hence quality of information that can be perceived from a single screen.

Continuity. The ease of integrating information across several displays not only depends on the individual display fidelities, but also on the quality difference or gap between those displays, in particular, if interaction moves across display boundaries. We call this *continuity of fidelity.* In addition, *continuity of the spatial reference frame* describes if the information space is continuous, as with virtual desktops, or discrete, e.g., when virtual display areas are bound to specific body parts [46]. Continuity factors pose potential challenges when combining multiple on and around the body displays. For example, combining touch screen and *HMD* extends the output beyond a physical screen of a smartwatch, but not the input. This leads to potential interaction challenges, when users associate the extension of the output space with an extension of the input space.

Social acceptability of interactions with mobile, on and around body devices have been extensively studied [196], revealing the personal and subjective nature of what is deemed acceptable. This varies due to many factors including the technology, social situation or location. Dynamic alignment allows for some degree of interaction customization, allowing people to tailor their interactions in a way which best suits their current context, rather than having to rely on default device patterns which may be wholly unsuited to the context of use.

6.2.1.2 Alignment modes

For the combination of *HMD* and touch device, we distinguish three possible alignment modes (see Figure 6.6):

In *body-aligned mode*, the devices share a common information space, which is spatially registered to the user's body (Figure 6.6, left). While wearable information displays could be placed anywhere in the 3D space around the body, we focus on widgets in planar spaces, as suggested by Ens et al. [65]. The *HMD* acts as a low fidelity viewing device into a body-referenced information space, allowing one to obtain a fast overview. The touchscreen provides a high fidelity inset, delivering detail-on-demand, when the user points to a particular location in the body-referenced space. Also, in contrast to common spatial pointing techniques, the touchscreen provides haptic input into the otherwise intangible information space.

In *device-aligned mode* , the information space is spatially registered to the touchscreen device and moves with it (Figure 6.6, middle). The *HMD* adds additional, peripheral information at lower fidelity, thus *extending the screen space* of the touch screen, yielding a focus+context display.

In *side-by-side mode*, interaction is redirected from one device to the other without requiring a spatial relationship among devices (Figure 6.6, right). For example, if the *HMD* shows a body-referenced information space, a touch device can provide indirect interaction. The touch device can display related information, and input on the touch device can affect the body-referenced display. If the touch device is outside the user's *FoV*, the touch screen can still be operated blindly.

6.2.1.3 Navigation

The principal input capabilities available to the user are spatial pointing with the touch device, or using the touch screen. Spatial pointing with the touch device is a natural navigation method in body-aligned mode. Once the alignment is recognized (the user's viewpoint, the handheld and the chosen item are aligned on a ray), the *HMD* clears the area around the element to let the handheld display a high resolution inset. This navigation method can be used for selection or even drag-and-drop in the body-referenced information space. However, extended use can lead to fatigue.

Spatial pointing in device-aligned mode can be seen as a more indirect form of navigation, which allows one to obtain a convenient viewpoint on the device-aligned information space. Navigation of the focus area will naturally be done by scrolling on the touch screen, but this can be inefficient, if the touch screen is small. Hence, users may mitigate the limitation of input to the physical screen with a clutch gesture that temporarily switches to

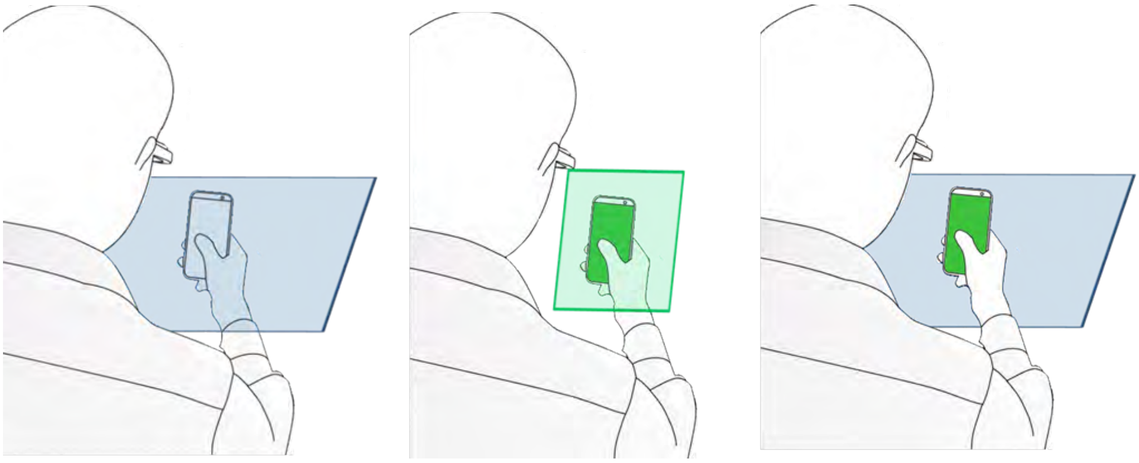


Figure 6.6: In body-aligned mode (left) devices are spatially registered in a shared information space relative to the user's body. In device-aligned mode (middle) the screen space of the touchscreen is extended. In side-by-side mode (right) devices have separated information spaces and do not require a spatial relationship.

body-aligned mode. At the press of a button (or dwell gesture), the information space can be fixed in air at the current position. Then users can physically select a new area of the information space by physical pointing, making it tangible again.

6.2.1.4 Focus representation and manipulation

An additional design decision is the representation shown on the higher fidelity display: The first option is to solely display a higher *visual level of detail*. For example, the user could align a touch screen over a label to improve the readability of text (Figure 6.5). The second option presents *semantic level of detail* [182], revealing additional information through a *ML* metaphor [24]. Here, the widget changes appearance to show additional information. For example, in Figure 6.7, the "Bedrooms" label turns into a scrollable list, once the borders of the handheld and the label represenation in the *HMD* are spatially aligned. Similarly, in Figure 6.8 (bottom row), a handheld shows a richer variation of a widget group including photos and detailed text, once it is aligned with the low fidelity representation on the user's arm.

An interactive focus representation on the touch device can naturally be operated with standard touch widgets. In body-aligned mode, this leads to a continuous coarse-to-fine *cascaded interaction*: The user spatially points to an item with a low fidelity representation and selects it with dwelling or a button press. A high fidelity representation of the item appears on the touch screen and can be manipulated by the user through direct touch (Figures 6.5, 6.7, 6.8).

For simple operations, this can be done directly in body-aligned mode. For example, widgets such as checkbox groups may be larger than the screen of a smartwatch, but individual checkboxes can be conveniently targeted by spatial pointing and flipped with a tap. However, holding the touch device still at arm's length or at awkward angles may be demanding for more complex operations. In this case, it may be more suitable to *tear*

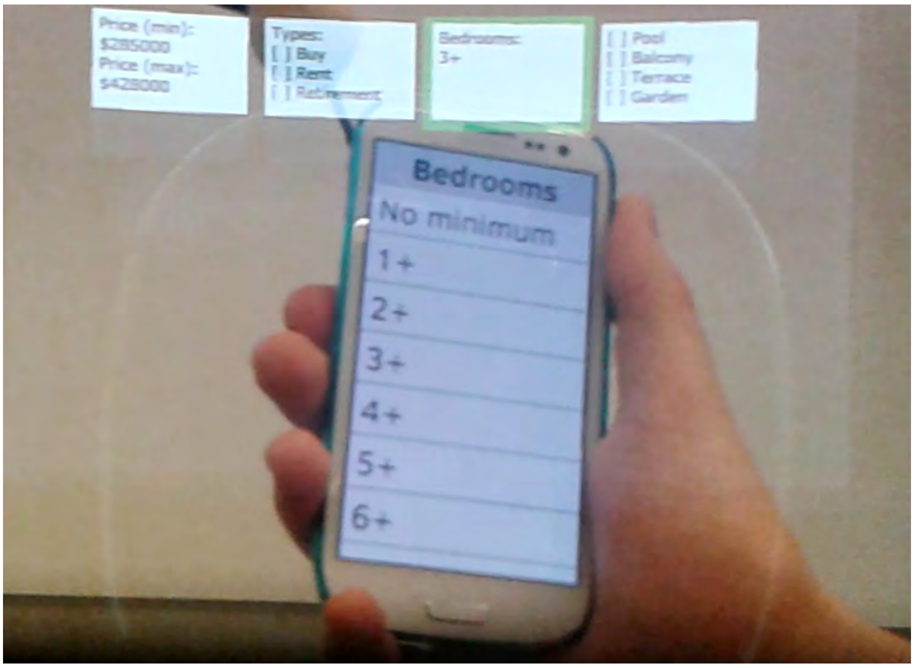


Figure 6.7: Spatial pointing via a handheld triggers a low fidelity widget on the *HMD* to appear in high fidelity on the handheld.

off the focus representation from the body-aligned information space by automatically switching to side-by-side mode. A rubberband effect snaps the widget back into alignment, once the user is done interacting with it. This approach overcomes limitations of previous work, which required users to either focus on the physical object or on a separate display for selection [52].

6.2.1.5 Widgets and applications

MultiFi widgets adapt their behavior to the current alignment of devices. For example, widgets can relocate from one device to the other, if a certain interaction fidelity is required. We have identified a number of ways how existing widgets can be adapted across displays. Here we discuss several widget designs and applications employing such widgets to exemplify our concepts.

Menus and lists: On a smartwatch, menu and list widgets can only show a few items at once due to limited screen space. We use an *HMD* to extend the screen space of the smartwatch, so users get a quick preview of nearby items in a *ring menu*, Figure 6.4, middle. Similarly, list widgets on an *HMD* can adapt their appearance to show more information once a handheld device is aligned, Figure 6.7.

Interactive map: Navigation of large maps is often constrained by screen space. We introduce two map widgets that combine *HMD* and touch screen. The first map widget works similar to the list widget, but extends the screen space of a touch display in both directions. Interaction is achieved via the touch display.

The second variant makes use of a body-referenced information space. The map is displayed in the *HMD* relative to the upper body, either horizontally, vertically or tilted (Figure 6.5). If the map size is larger than the virtual display space, the touchpad on the smartwatch provides additional pan and zoom operations.

Arm clipboard: Existing body-centric widgets for handhelds [46, 151] rely on proprioceptive or kinesthetic memorization, because the *FoV* of the handheld is small. With an additional *HMD*, users can see where on their body they store through head pointing and subsequently retrieve information with a handheld device. If a list widget displays additional information on one side of the smartwatch (overview+detail), we can let users store selected items on their lower arm (Figure 6.8). Aligning the handheld with one of the items stored on the arm automatically moves the item to the higher fidelity handheld. For prolonged interaction, the item can now be manipulated with two hands on the handheld. Through the combination of *HMD* for overview and touch enabled displays for selection and manipulation, body-referenced information spaces could become more accessible compared to previous approaches solely relying on proprioceptive memory [46, 151].

Text input: Using MultiFi text widgets, we have implemented a full-screen soft keyboard application for a handheld used with a *HMD*. The additional screen real estate on the handheld allows MultiFi to enlarge the soft keys significantly, while the text output is redirected to the *HMD*. As soon as a *HMD* is aligned, the text output area can relocate from one device to the other (see Figure 6.4, right). This results in two potential benefits. First, the larger input area could help speed up the writing process. Second, the written text is not publicly visible, hence supporting privacy.

6.2.2 Implementation

6.2.2.1 Software

The MultiFi prototype is based on HTML5, JavaScript, WebSockets for communication, three.js for rendering and hammer.js for touch gesture recognition. All client devices open a website in a local browser and connect to the Java-based application server. JSON is used to encode the distributed messages, and tracking data is received via VRPN.

Widgets have potentially multiple graphical representations in replicated and synchronized scenegraphs and a common state which is shared via the central application server. For widgets that do not change their appearance and simply span multiple devices, multiple camera views on the same 3D scene are used (e.g., ring menu, map). Widgets that adapt their appearance (such as list items) use multiple synchronized representations. Interaction across devices relies on the known 3D poses of individual devices, shared via the central application server. For example, selection of an item in the *HMD* via a touch screen is realized through intersection from the touch point with the virtual *HMD* image plane.

As our system relies on the accurate registration between devices, calibration of individual components is required. Foremost, the *HMD* is calibrated via *OST* calibration methods (using single or multiple point active alignment methods [91]). In addition, the image masks for the touch screen devices (i.e., the area that should not be rendered on the *HMD*) and thus their positions relative to their tracking markers have to be determined. For this the user manually aligns the touch screen with a pre-rendered rectangle displayed

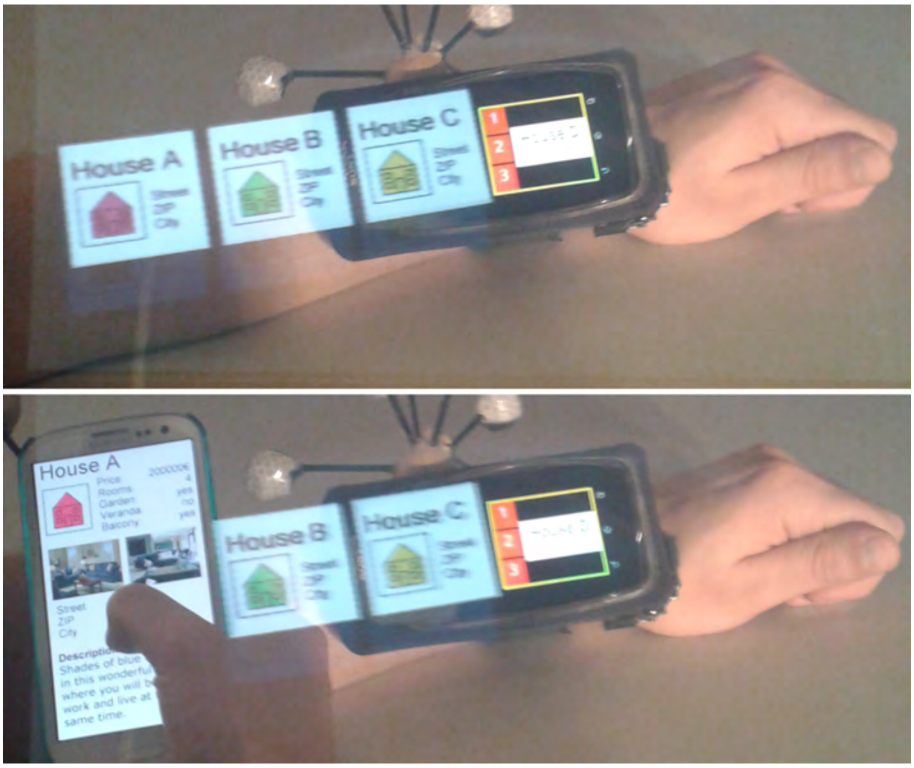


Figure 6.8: Arm clipboard with extended screen space for low fidelity widgets (top). Spatial pointing enables switching to high fidelity on a handheld (bottom).

on the *HMD* (having the same size as the touch screen) which allows MutliFi to determine the transformation between the touch screen and the attached tracking target. Please note that these calibration steps typically have to be carried out only once for each user and device respectively.

6.2.2.2 Devices

We implemented a MultiFi prototype using a Samsung Galaxy SIII (resolution: 1280x720 px, 306 ppi, screen size: 107x61 mm) as smartphone, a Vuzix STAR 1200 XL *HMD* (resolution: 852x480 px, horizontal FoV: 30.5° vertical FoV: 17.15°, focus plane distance: 3 m, resolution: 13 ppi at 3 m, weight with tracking markers: 120 g) and another smartphone (Sony Xperia Z1 compact) as smartwatch substitute (resolution: 1280x720 px, cropped extent: 550x480 px, 342 ppi, weight with tracking markers: 200 g). We chose this approach to simulate next generation smartwatches with higher display resolution and more processing power. To this aim, we limited the screen extent to 40x35 mm to emulate the screen extent of a typical smartwatch. The *HMD* viewing parameters were matched with virtual cameras which rendered the test scenes used in the smartphone, *HMD* and smartwatch.

6.2.2.3 Tracking

We used an A.R.T. outside-in tracking system to determine the 3D positions of all devices. This currently limits our prototype to stationary use in laboratory environments. Still, mobile scenarios could be supported by relying on local sensors only. For example HMDs with in-built (depth) cameras could be used to determine the 3D position of touch screens relative to the *HMD* [186]. Alternatively, in-built orientation sensors could track the touch screen and *HMD* positions relative to a body-worn base station (such as an additional smartphone in the user's chest pocket). Please note that the later approach would likely result in less accuracy and drift over time. This would need to be considered in the adaptation rules for widgets when spanning multiple devices.

6.2.3 User study

We conducted a laboratory user study to investigate if combined device interaction can be a viable alternative to established single device interaction for mobile tasks. For the study we concentrated on two atomic tasks: information search and selection. Those tasks were chosen as they can be executed on the go and underpin a variety of more complex tasks.

6.2.3.1 Experimental design

We designed a within-subjects study to compare performance and user experience aspects of MultiFi interaction to single device interaction for two low level tasks. We complemented the focus on these atomic tasks with user inquiries about the potential and challenges of joint on and around the body interaction. For both tasks, we report on the following dependent variables: *TCT*, *errors, subjective workload* as measured by NASA TLX [100] as well as user experience measures (after scenario questionnaire [149], *hedonic and usability aspects* as measured by AttrakDiff [106]) and overall *preference* (ranking). The independent variable for both tasks was interface with five conditions:

Handheld: The Samsung Galaxy SIII was used as only input and output device. This serves as the baseline for a handheld device with high input and output fidelity.

Smartwatch (SW): The wrist-worn Sony Xperia Z1 compact was used as only input and output device. The input and output area was 40x35 mm and highlighted by a yellow border, as shown in Figure 6.5. Participants were notified by vibration if they touched outside the input area. This condition serves as baseline for a wearable device with low input and output fidelity (high resolution, but small display space).

HMD: The Vuzix STAR 1200XL was used as an output device. We employed indirect input as in the smartwatch condition using a control-display ratio of 1 with the touch area limited to the central screen area of the *HMD*. This condition serves as the baseline for a *HMD* with low input and output fidelity, which can be operated with an arm-mounted controller.

Body-referenced interaction (BodyRef): The content was displayed in front of the participant in body-aligned mode with additional touch scrolling. Selection was achieved by aligning the smartwatch with the target visible in front of the user and touching the target rendered on the smartwatch.

Smartwatch referenced (SWRef): The information space was displayed in device-aligned mode (Figure 6.12). All other aspects were as in BodyRef.

6.2.3.2 Apparatus and data collection

The study was conducted in a controlled laboratory environment. The devices employed were the ones described in the implementation section. The translation of virtual cameras for panning via touch in all conditions parallel to the screen was set to ensure a control-display ratio of 1. Pinch to zoom was implemented by the formula $s = s_0 \cdot s_g$, with s being the new scale factor, s_0 the map's scale factor at gesture begin and s_g the relation between the finger distances at gesture begin and end. While the system is intended for mobile use, here participants conducted the tasks while seated at a table (120x90 cm, height 73 cm, height adjustable chair) due to the strenuous nature of the repetitive tasks in the study. Null hypothesis significance tests were carried out at a .05 significance level, and no data was excluded, if not otherwise noted. For ANOVA (repeated measures ANOVA or Friedman ANOVA), Mauchly's test was conducted. If the sphericity assumption had been violated, *DOF* were corrected using Greenhouse-Geisser estimates of sphericity. For post-hoc tests (pairwise t-test or Wilcoxon signed rank) Bonferroni correction was applied.

6.2.3.3 Procedure

After an introduction and a demographic questionnaire, participants were introduced to the first task (counterbalanced) and the first condition (randomized). For each condition, a training phase was conducted. For each task, participants completed a number of trials (as described in the individual experiment sections) in five blocks, each block for a different condition. Between each block, participants filled out the questionnaires. At the end of the study, a semi-structured interview was conducted and participants filled out a separate preference questionnaire. Finally, the participants received a book voucher worth 10 Euros as compensation. Participants were free to take a break between individual blocks and tasks. Overall, the study lasted ca. 100 minutes per participant.

6.2.3.4 Participants

Twenty-six participants volunteered in the study. We had to exclude three participants due to technical errors (failed tracking or logging). In total, we analyzed data from twenty-three participants (1 f, average age: 26.75 y, σ=5.3, average height: 179 cm, σ= 6, 7 users wore glasses, three contact lenses, 2 left-handed users). All but one user were smartphone owners (one less than a year). Nobody was a user of smartwatches or head-mounted displays. Twenty users had a high interest in technology and strong computer skills (three medium).

6.2.3.5 Hypotheses

One of our main interests was to investigate if combined display interaction could outperform interaction with individual wearable devices. We included Handheld interaction as a baseline and did not expect the combined interfaces to outperform it. Hence, we had

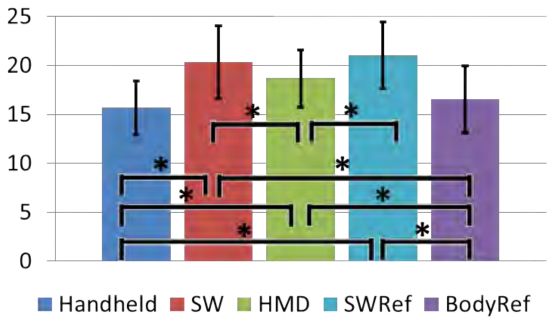


Figure 6.9: *TCT* (s) for the locator task.

the following hypotheses: *H1:* Handheld will be fastest for all tasks. *H2:* BodyRef will be faster than *HMD* and smartwatch (ideally close to Handheld). *H3:* BodyRef will result in fewer errors than *HMD* and smartwatch. *H4:* SWRef will be faster than *HMD* and smartwatch (ideally close to Handheld). *H5:* SWRef will result in fewer errors than *HMD* and smartwatch.

6.2.4 Experiment 1: Locator task on map

A common task on mobile mapping applications is to search for an object with certain target attributes [191]. We employed a locator task similar to previous studies involving handheld devices and multi-display environments [88, 187]. Participants had to find the lowest price label (text size 12 pt) among five labels on a workspace size of 400x225 mm. We determined the workspace size empirically, to still allow direct spatial pointing for the BodyRef condition. While finding the lowest price could easily be solved with other widgets (such as a sortable list view), our task is only an instance of general locator tasks, which can encompass non-quantifiable attributes such as textual opinions of users, which cannot be sorted automatically. Users conducted ten trials per condition. With 23 participants, five interface levels and 10 trials, there was a total of 23x5x10=1150 trials.

6.2.4.1 Task completion time and errors

The *TCTs* (in seconds), for the individual conditions can be seen in Figure 6.9 and were as follows: Handheld (M=15.67, σ=5.45), SW (M=20.60, σ=7.62), *HMD* (M=18.68, σ=6.45), BodyRef (M=16.57, σ=6.16), SWRef (M=21.05, σ=10.28). A repeated measures ANOVA indicated that there was a significant effect of interface on *TCT*, F(3.10, 709.65)=42.21, p<.001. Post-hoc tests indicated that both Handheld and BodyRef were significantly faster than all remaining interfaces with medium to large effect sizes (see also Figure 6.9). *HMD* was significantly faster than both smartwatch and SWRef. There were no significant differences between Handheld-BodyRef and SW-SWRef.

From 230 selections, eight false selections were made in the Handheld, *HMD* and BodyRef conditions. In the SW condition, 13 errors have been made, in SWRef five errors. No significant differences were found.

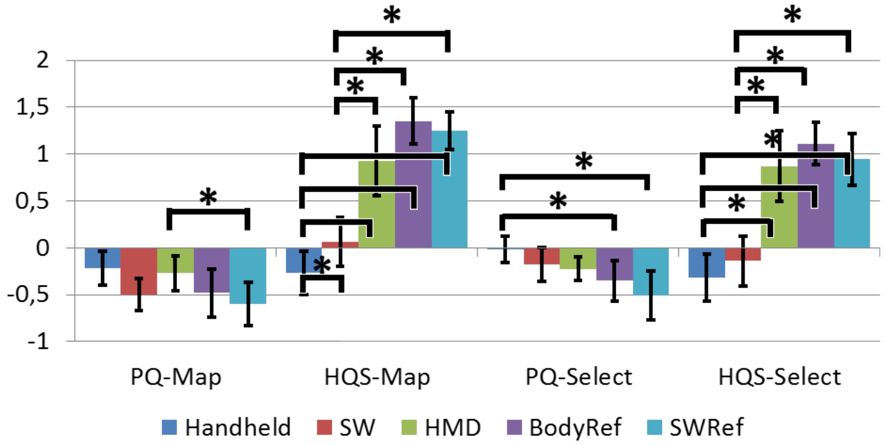


Figure 6.10: *PQ* and *HQ-S* measures (normalized range -2..2) for the locator task (left) and the select task (right).

6.2.4.2 Subjective workload and user experience

Repeated measures ANOVAs indicated that there were significant effects of interface on all dimensions. Post-hoc tests indicated that BodyRef resulted in a higher mental demand than smartwatch (albeit with a small effect size). The handheld condition resulted in significantly lower subjective workload for all other dimension compared to most other interfaces. The analysis of the after scenario questionnaire (repeated measures ANOVA and post-hoc tests) indicated that for Handheld ease of task was significantly higher than for SWRef. Analysis of AttrakDiff showed that all interfaces scored slightly below average for *PQ*, see Figure 6.10, and only a significant difference between *HMD*-SWRef could be found (but with a small effect size). For *HQ-S*, the Handheld and SW interface were rated significantly lower than the other three conditions. Preference analysis showed that Handheld (MD=2, M=1.13, σ=1.13) was significantly more preferred than SW (MD=4, M=3.87, σ=1.10), Z=-4.25, p<.001.

6.2.5 Experiment 2: 1D target acquisition

We employed a discrete 1D pointing task similar to the one used by Zhao et al. [266] (Figure 6.11). Participants navigated to a target (green stripe) in each trial using touch input (for Handheld, SW, *HMD*, SWRef) or spatial pointing (BodyRef). Final target selection was confirmed by a touch on the target region in all conditions. The participants were asked to use their index finger to interact with the touch surfaces. For each trial, the task was to scroll the background (Handheld, SW, *HMD*, SWRef) or to move the smartwatch towards the target (BodyRef) until it appeared on the *selection area*. Prior to each trial, participants hit a start button at the center of the screen to ensure a consistent start position and to prevent unintended gestures before scrolling. The target was only revealed after the start button was hit. After successful selection, the target disappeared. For BodyRef, participants returned to a neutral start position centered in front of them before the next trial. In the experiment design we fixed target width to 20 mm (0.5*width of the

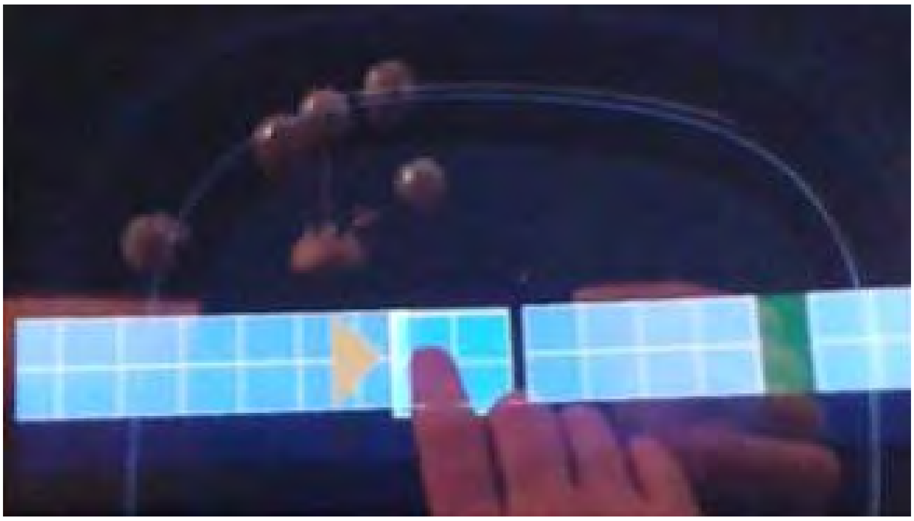


Figure 6.11: The selection task for SWRef.

smartwatch), use the control window and display window sizes of the individual displays and use two target distances (short: 15 cm, long: 30 cm)[1]. The conditions were blocked by interface. Per condition, each participant conducted eight trials (plus two training trials). With twenty three participants, five interface levels, two target distances, two directions and eight trials per condition a total of 23x5x2x2x8=3680 trials were conducted.

Please note that the focus of this experiment is not to derive a new target acquisition model but rather to get an initial insight into the potential for combined wearable device interaction compared to individual devices only. Hence, in the experiment design, we do not vary all parameters as one would need for deriving a robust model. Specifically, we fix target width to 20 mm (0.5*width of the smartwatch), use the control window and display window sizes of the individual displays and use two target distances (short: 15 cm, long: 30 cm). In addition to interface and target distance, we also introduced target direction (same side as hand carrying the smartwatch and opposite side), as independent variable as we expected performance differences in the BodyRef condition.

6.2.5.1 Task completion time and errors

TCTs are depicted in Figure 6.12. Repeated measures ANOVAs indicated that for both distances (15 cm, 30 cm) and smartwatch sides (towards and away from dominant hand) interface had a significant effect on TCT. The pairwise significant differences are depicted in Figure 6.12. Handheld was the fastest interface for both directions and distances. BodyRef was significantly faster than all remaining interfaces. No other significant effects of interface on TCT were found.

Selection errors occurred when participants tapped outside the target region. The total number of errors for individual interfaces were as follows: Handheld: 53 (M=.07, σ=.28), SW: 34 (M=.05, σ=.23), *HMD*: 223 (M=.30, σ=.77), BodyRef: 258 (M=.35,

[1] We fixed those parameters as the focus of the experiment was not on generating a new target aquisition model.

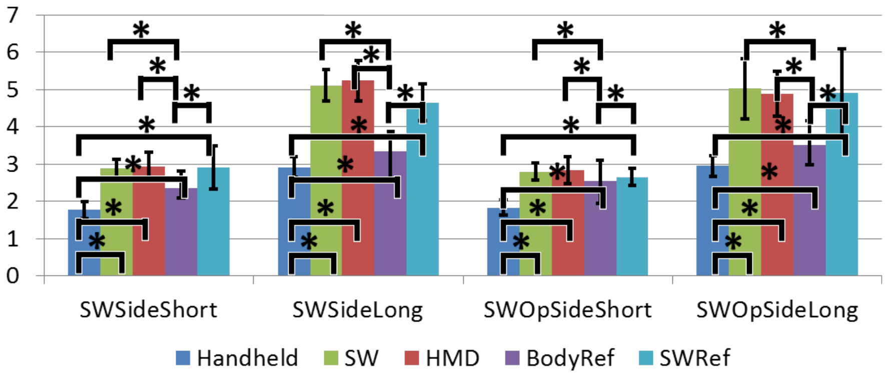


Figure 6.12: *TCT* (s) for the select task. SWSide: side on which smartwatch was worn, SWOpSide: opposite side.

σ=.78), SWRef: 37 (M=.05, σ=.24). A Friedman ANOVA indicated that there was a significant effect of interface on error count ($\chi^2(4)$=231.68, p<.001). Post-hoc tests indicated significant differences between BodyRef and all interfaces except *HMD*, as well as between *HMD* and all interfaces (except BodyRef).

6.2.5.2 Subjective workload and user experience

A repeated measures ANOVA indicated that there were significant effects of interface on all dimensions but temporal demand and performance. Post-hoc tests indicated that Handheld resulted in a significantly lower mental demand than most other conditions (except SW) and in a significantly lower overall demand than all conditions. BodyRef and SWRef resulted in significantly higher physical demands compared to Handheld and *HMD* (but not SW). Frustration was significantly higher for SW and SWRef compared to Handheld. Analysis of results of the after scenario questionnaire indicated a significant difference between Handheld and SWRef for ease of task (Z= -3.36, p=.01). As in the locator task, all interfaces scored below average for *PQ*-S (see Figure 6.12). BodyRef and SWRef scored significantly lower than Handheld (indicated by repeated measures ANOVA and post-hoc t-tests). For *HQ-S*, the Handheld and SW interface were rated significantly lower than the other three conditions as in the locator task.

6.2.5.3 Qualitative feedback

In semi-structured interviews participants commented on potentials and limitations of the prototypical MultiFi implementation. Most participants (21) commented on the benefits of having an *extended view space* compared to individual touch screens with one participant saying "*Getting an overview with simple head movements is intuitive and natural*". Those participants also valued the fact that precise selection was enabled through the smartwatch with one typical comment being "*The HMD gives you the overview, and the smartwatch lets you be precise in your selection*". Three participants highlighted the potentially lower *access costs* of MultiFi over smartphones, with one comment being "*I don't have to constantly monitor my smartphone*". In line participants felt that BodyRef interaction was fastest (even though this is not confirmed by the objective measurements). Five participants commented on the benefits of MultiFi over *HMD* only interaction highlighting the direct interaction or that they could "*take advantage of proprioception and motion control*".

Many participants (15) commented on the limitations of the hardware, specifically the quality of the employed *HMD* with a typical comment being "*The combined interfaces [SWRef, BodyRef] gave me trouble because of display quality*". Specifically, the employed *HMD* obscured parts of the users' *FoV* "*preventing the ability of glancing down (on the smartwatch) without moving your head*". Another issue highlighted by 6 participants was the *cost of focus switching* which refers to the accommodation to different focus depths of the touch screen and the virtual *HMD* screen with a typical comment being: "*I have to focus on three layers, which is overwhelming: smartwatch, HMD and real world*". This also led to *coordination problems across devices* as mentioned by 9 participants. Hence, some participants suggested not to concurrently use *HMD* and smartwatch as output: "*Pairing the two devices is good, but use one as input, the other as output, not both as output, it's confusing*". Also, social concerns of spatial pointing were raised, "*I could not imagine this in a packed bus*".

6.2.6 Discussion

The study results indicate that combined smartwatch and *HMD* interaction in body-referenced information spaces can outperform individual wearable devices in terms of *TCT* (*H2* holds) and that handheld interaction is not always fastest (H1 does not hold). However, this currently comes at the expense of higher workload and lower usability ratings. We see two major sources for this. First, compared to commercially available wearable devices, we used relatively heavy laboratory equipment (smartphone and *HMD* with separate retro-reflective markers). Participants mentioned that they would prefer the combined interaction more, if it were lighter. Second, we compared novel interaction techniques involving continuous spatial pointing with established touch screen interaction. Hence, we assume that both lighter devices and more training could mitigate these workload effects.

In the selection task, BodyRef and *HMD* resulted in a significantly higher error number than the other interfaces (H3 does not hold). Also, SWRef did not result in significantly less errors (H5 does not hold). For *HMD*, this could be explained by the indirect touch input combined with a smaller control window (smartwatch area) compared to the larger display window. For BodyRef, it turned out that the outside-in tracking system for spatial pointing and our system architecture introduced an average end-to-end delay from user motion to display update of 154 ms (σ=36). A further video analysis revealed that users were tapping the smartwatch repeatedly when they have reached the target area, even though they were informed to select as precisely as possible. While this is clearly a limitation of our current experimental system setup, we believe that future tracking systems will minimize delay, allowing more precise physical pointing.

Semi-structured interviews revealed that users generally preferred Handheld, as it was the most familiar device, had the largest touch input area and was most comfortable to use. BodyRef was preferred as it felt fast and separated target search via head pointing and selection via spatial pointing with the smartwatch. User comments included "*Moving* *your head to get an overview is very intuitive*" and "*knowing where to move before you move makes it easier than other conditions*". Still, confirming spatial selection with the touchpad was not welcomed by all, "*I would prefer to just point with my fingers or eyes*".

SWRef performed (for both tasks) not better than individual devices, even though they are based on the extended screen space metaphor as the body-referenced condition (H4 does not hold). Subjective feedback in the semi-structured interviews indicated that participants could not efficiently use the SWRef condition due to the need for refocusing between the smartwatch display (˜40 cm distance) and the focus plane of the *HMD* (˜300 cm). In addition, the *HMD* had a lower visual fidelity, which likely increased the effort for reading the labels. Some participants still favored the SWRef condition, specifically for the selection task. They indicated that the *HMD* gave them "*a peripheral awareness when the target approaches the smartwatch*". This hints that smartwatch referenced display space extension could be beneficial, if the visual fidelity of the *HMD* and costs of display switching is considered in the design process. For example, instead of rendering a map continuously across displays without adjustments, individual map regions could be adjusted to be more readable across displays (or to avoid the need for actually reading the text on the *HMD* at all).

Smartwatch alone was least preferred due to cumbersome interaction with a small input and output area. Specifically, swiping motions were deemed inefficient. For example, in the select task, participants mentioned a lack of overview, "*I did not know when I passed the target*". *HMD* was preferred by some users, because they could keep their head and arm in comfortable positions and have a "lean back" experience. They mention it is "better than Google Glass as I can use the smartwatch as touchpad". One participant said: "I could imagine using this for presentations were I can see the slides in the *HMD* and keep eye contact with the audience when controlling the app with my smartwatch".

Revisiting MultiFi we see that the spectrum of *dynamic alignment* ranging from uncoupled individual devices to closely coupled spatially registered interaction is a key concept for supporting a broad range of mobile scenarios. It facilitates the idea that over time, users can develop individual preferences for multi-display interaction styles just like current touch interfaces offer multiple ways of interaction. The qualitative feedback in the study indicated that users could see benefits of MultiFi over individual device interaction in terms of *access costs* and *direct interaction.* Being able to directly interact within this view space through a touch screen distinguishes MultiFi from other approaches like mid-air interaction via depth-sensors, which lack the haptic feedback of touch screens and through this potentially result in a lower selection precision.

However, such benefits may come at an increased *coordination cost* across displays. Specifically, while we presented a first set of possible widgets, our study revealed that those widgets have to be designed carefully to be able to efficiently lower interaction gaps introduced by individual devices (such as focus distance and resolution differences). Simply extending the display space for widgets across displays without adapting their appearance and operation (as done with the smartwatch referenced map) seems not to be enough to overcome interaction seams. This indicates the need for more research to further investigate the particulars of efficient cross display widgets for interaction on the go. For example, for the map widget we could imagine to further reduce the visual complexity on the low fidelity *HMD* by simply indicating the location of *POIs* with details only appearing on the high fidelity display (as in the arm clipboard).

6.3 Summary

Within this chapter, we presented several prototypes for interaction with electronic displays. Our first prototype, HeadLens, demonstrated how the infrastructure effort for realizing augmentations on public displays can be minimized. At the same time, the prototype supported user-perspective rendering of the public display content for devices that support dual camera access.

In the second part of this chapter, we have presented MultiFi, an interactive system that combines the strengths of multiple displays on and around the body. We explored how to minimize seams in interaction with multiple devices by dynamic alignment between interfaces. Furthermore, we discussed the implications for user interface widgets and demonstrated the feasibility of our concept through a working prototype system. Finally, we demonstrated that combined *HMD* and smartwatch interaction can outperform interaction with single wearable devices in terms of *TCT*, albeit with higher workload.

7 Conclusion

Contents

7.1 Summary of the thesis results 171

7.2 Limitations 172

7.3 Directions for future work 173

7.4 Summary 175

7.1 Summary of the thesis results

This thesis investigated the potential of *AR* user interfaces for increasing utilitarian and hedonic user experience aspects when mobile consumers interact with information surfaces. A central assumption of this thesis was that this potential is heavily dependent on contextual factors.

After framing this thesis in related work (Chapter 2), we started with surveys about the current state of context-aware *AR* systems, the usage of *AR* browsers and about information access at public information surfaces in Chapter 3. The findings of these surveys informed both the detailed evaluation of specific context factors in a series of user studies and the design and implementation of *AR* interfaces that aim at supporting interaction with information surfaces.

The user studies presented in Chapter 4 specifically concentrated on large printed information surfaces in public space. Two application types were investigated. First, gaming at public posters was investigated. A series of studies showed that the social context and spatial setting of interaction can influence the use of *AR* user interfaces. Then, we investigated utility driven information browsing at public maps in a touristic setting. A series of studies indicated that for DIN A0 sized posters *AR* did not deliver benefits for users over *SP* interfaces, but as the poster size grew *AR* resulted in a better usability. Based on the insights of these results and the findings of Chapter 3 we then proposed hybrid user interfaces, which combine *AR* with *SP* elements, for interaction with large information surfaces in mobile contexts.

In Chapter 5, we turned our focus on small printed information surfaces. We investigated the feasibility of supporting the checking of security documents and banknotes through handheld *AR* user interfaces for hologram verification. Our first prototype indicated this feasibility, but it also resulted in long verification times and a high workload. Hence, we designed and evaluated a number of further prototypes in an iterative design process and could show that they resulted in lower verification times and workload. The key idea behind these new interfaces was to give users freedom in navigating a small information space instead of forcing them to precisely align a six *DOF* pose. However, our results also indicated that the current interfaces are likely still too slow to be operated in real world usage situations.

Finally, in Chapter 6, we studied both *AR* user interfaces for large and for small electronic information surfaces. First, we proposed a workflow for augmenting public displays, without the need for extensive infrastructure. Within this workflow, we also demonstrated how to achieve user perspective rendering for public display content. Then, we turned our focus on improving the interaction across multiple personal wearable displays such as smartphones, smartwatches and head-mounted displays. We did so by exploring the design space of cross device interaction with wearables and showcased several prototypical applications. Finally, through a user study we could show that interaction tasks such as selection or information browsing can be conducted more efficient with the combination of multiple devices compared to interaction with single devices.

7.2 Limitations

This thesis concentrated on mobile *AR* user interfaces for information surfaces. It explicitly did not explore in-depth related research areas such as surface computing (e.g., tabletop interaction, interaction above surfaces or touch-based interaction at public displays), spatial *AR* or *AR* interaction with complex physical 3D objects.

Furthermore, while this thesis tried to sample several application areas relevant for interaction with information surfaces, ranging from gaming over information browsing to security inspection, other potentially interesting application domains, such as medical support or industrial maintenance, are left for future research.

Also, not all characteristics of surfaces, such as shape or spatial configurations, were examined in detail. Instead, we focused on selected relevant configurations for individual application domains. i.e., typical poster sizes for print media that are vertically mounted or typical sizes of security documents.

Finally, this thesis also heavily relied on evaluation methods that go beyond quantitative measures, often found in laboratory-based completely randomized experiments. Instead triangulation of qualitative and quantitative methods, both in the field and in the laboratory, was used. Still, individual studies reported in this thesis have varying weights on quantitative and qualitative measures as suggested as best practice for studies in the wild [206].

7.3 Directions for future work

We see several research directions that could complement the results of this thesis.

In Chapter 4 we employed laboratory studies and quasi-experiments in the field to study the usage of *AR* for interacting with large printed information surfaces. While these types of studies are suited for identifying usability issues, performance measures (such as studied in chapter 5) or to identify potentially relevant context factors, they are not without challenges when studying the real-world use of user interfaces. Specifically, the experimenter was always present during the studies, which could lead to participant bias. Furthermore, the intrinsic motivation for using an application could not be verified. Instead, users were likely externally motivated to participate in these studies. Specifically, it remains unclear if they would show similar usage behavior outside of the study setting. Hence, it is advisable to complement the findings of these studies with large scale deployments of near to product apps incorporating remote logging. While these remote studies can not easily deliver as rich data as in-situ observations, they are likely to result in a larger ecological validity. In fact, we already started to work on such a product-like app for information access at large hiking maps and are looking forward to deploy this app in the upcoming hiking season in an Austrian hiking resort. Further, the current studies concentrated on use by individuals. However, large information surfaces potentially lend themselves for collaborative interaction between multiple users. In fact, tourists often travel in groups and we see potential in supporting their collaborative interactions with mobile *AR* user interfaces.

Furthermore, it would be interesting how to support interaction beyond the boundaries of single information surfaces. Specifically, it would be interesting to explore the creation and interaction with ad-hoc mobile multi-display environments. While our MultiFi prototype presented in Chapter 6 allowed interaction across multiple wearable displays, it did so in a prepared laboratory environment. In a first step, we want to build a fully mobile prototype using only mobile sensors. Further, the investigation of interaction techniques for multiple wearable displays could be expanded. Specifically, it should be explored in more detail which tasks would benefit specifically from bi-manual interaction (e.g., with a smartwatch on each arm), in addition to having a *HMD* device available.

Looking beyond wearable displays, it would also be interesting to extend the input and output space of handheld displays in mobile contexts (see Figure 7.1, left). While a number of concepts exists both for around device interaction and mobile multi-display environments, most prototypes require modifications of handheld displays and have only been demonstrated in laboratory settings. Instead, we would aim at only using unmodified off-the-shelf devices. In a first step, we would like to explore the use of sunglasses, which, in conjunction with the front-facing camera of smartphones, provide a projection of the surrounding of the phone. Using this information, one could utilize the space around the smartphone as input space. Similarly, a large tabletop display could be assembled from several tablets in and ad-hoc fashion by registering their relative positions through the sunglass reflections. In a further step, we would like to study how to loosen the constraint of wearing separate sunglasses and instead rely on reflections of the eye. While likely not achieving the same input resolution as with sunglasses, coarse pointing might be achievable with corneal reflective imaging techniques [173].

Also, in Chapter 6 we proposed a workflow for augmentations of public displays. However, the demonstrated worklow, as most other computer vision-based tracking approaches for electronic displays, suffers from a limited interaction range. Specifically, it is not possible to register the handheld device to the large electronic display on its surface. However, a continuous interaction range from on surface to above the surface could open up new interaction opportunities. As an example, we envision simply putting a handheld display onto the surface of a larger electronic display in a map navigation scenario to reveal additional information layers (see Figure 7.1, middle). To enable such interaction, we would like to explore the embedding of non-perceivable codes into arbitrary display content (in contrast time-multiplexed codes that only work on homogeneous regions like [259]). Also, the provision of AR content without the need for explicit modelling before deployment could be further investigated (c.f. [174]). For facilitating the deployment further we also envision to explore web frameworks for AR (c.f. [175]).

Similarly, we want to further explore the interaction with printed information surfaces above and on them. In future research, we want to explore how to combine the benefits of digital and printed media into one unifying user experience. For example, we envision being able to annotate or search through a printed book almost as easily as through a digital PDF. To enable this interaction, we envision to utilize the digital nature of publishing processes,i.e., that for most printed books digital PDFs are already available. We want to enable information extraction and subsequent recommendation of content for unknown paper documents but also enable annotations through handheld devices directly on the book surface (see Figure 7.1, right). In contrast to electronic displays, in the foreseeable future it seems not possible to embed time-varying imperceptible codes into printed materials. Instead, we would like to explore the use of dual-camera tracking. As long as the information surface is still recognizable with the default back-facing camera, tracking is done through it. Once a critical distance to the surface is reached, tracking switches to the front camera. The challenge lies in still being able to robustly and precisely track, at least with two *DOF* on the surface, using most likely sparse visual information from the front camera.

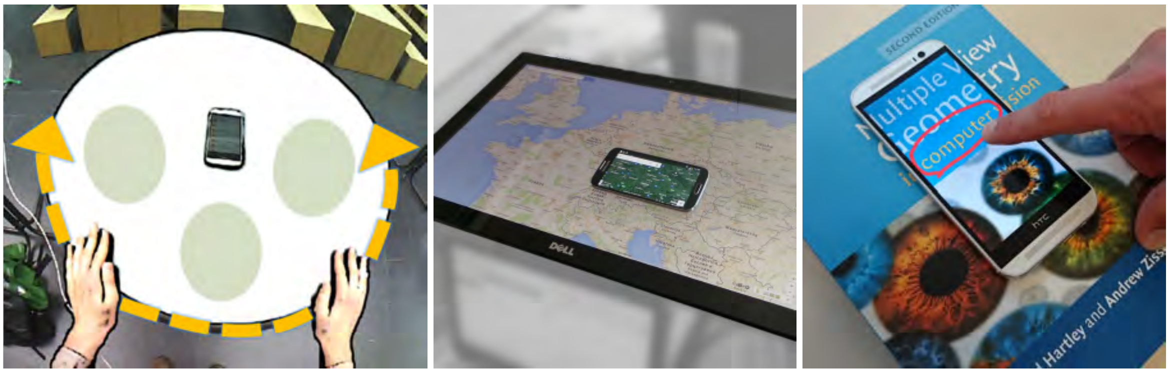


Figure 7.1: How to support ad-hoc around device interaction (left), continuous interaction with handhelds on and above large electronic information surfaces (middle) and cross-media interaction (right)?

Finally, we see great potential in further exploring context-sources presented in Chapter

3. For example, we envision to estimate physiological states of users on the go in order to adapt the user interface accordingly. To this end, front facing camera of handheld devices or in-built eye-tracking cameras of head-mounted displays could be used to extract stress and visual attention measures. For social awareness, we envision to employ information about places, i.e., the meanings which humans assign to physical locations, and social networks, i.e., sets of people or organizations and their paired relationships, into the design of mobile *AR* user interfaces. For example, this could be applied in an *AR* game which suggests players to switch to alternative interfaces based on visual scene analysis which measures the crowdedness of a street.

7.4 Summary

This chapter summarized the results of this dissertation in the light of discoveries made throughout the chapters. This thesis aimed at investigating the potential of *AR* user interfaces for increasing utilitarian and hedonic user experience for interaction with information surfaces. While we presented our investigations and results, we also identified limitations of this work. Finally, we presented future research directions that were induced through the work on this thesis. A thank you to the reader for paying attention to this thesis, in hope of inducing motivation for related research.

A

List of Acronyms

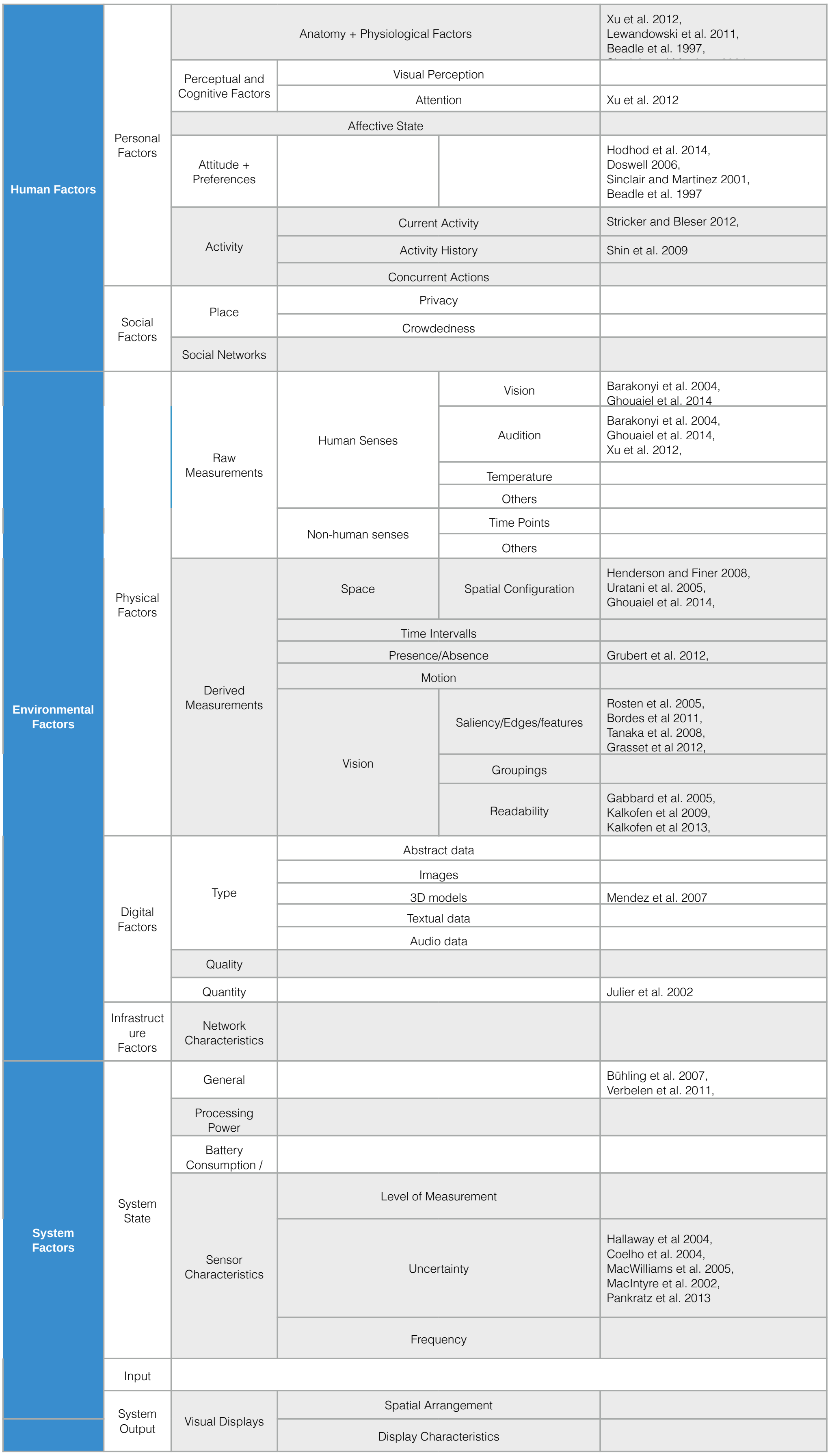
AR	Augmented Reality
BRDF	Bidirectional Reflectance Distribution Function
DM	Digital Manual
DOF	Degrees of Freedom
DP	Dynamic Peephole
FoV	Field Of View
GPS	Global Positioning System
HMD	Head-Mounted Display
HQ-I	Hedonic Quality - Identity
HQ-S	Hedonic Quality - Stimulation
IE	Interest/Enjoyment
MIRW	Mobile Interaction with the Real World
ML	Magic Lens
NCC	Normalized Cross Correlation
NFT	Natural Feature Tracking
NHST	Null Hypothesis Significance Testing
OST	Optical See-Through
PDA	Personal Digital Assistant
POI	Point of Interest
PQ	Pragmatic Quality
PTAM	Parallel Tracking and Mapping
SDK	Software Development Kit
SLAM	Simultaneous Tracking and Mapping
SP	Static Peephole
SVBRDF	Spatially Varying Bidirectional Reflectance Distribution Function
TCT	Task Completion Time
VST	Video See-Through
VU	Value/Usefulness

B

Overview Context-Aware AR Systems

System Input			Hallaway et al. 2004, Stricker et al. 2012, Xu et al. 2012, {Lewandowski et al. 2011, Henderson et al 2008, Grubert et al 2012
System Configuration	Sensors + Registration Techniques		MacWilliams et al 2005, Verbelen et al. 2011, Hallaway et al 2004
System Output	Content		Sinclair et al 2001, Beadle et al 1997, Bühling et al 2012, Shin et al 2009, Dünser et al 2011, Hodhod et al 2014, Coelho et al. 2004, Barakonyi et al 2004
	Information Presentation	Spatial Arrangement	Rosten et al 2005, Grasset et al 2012, Bordes et al 2011, Tanaka et al 2008
		Appearance Adaptation	Kalkofen et al. 2013, MacIntyre et al 2002, Bordes et al. 2011, Uratani et al 2005, Sinclair et al. 1997, Bühling et al 2012, Hallaway et al 2004, Kalkofen et al. 2009, Mendez et al 2007 Ghouaiel et al 2014 Grubert et al 2012, Bordes et al 2011, Gabbard et al. 2005,, Pankratz et al 2013

Figure B.1: Context targets relevant for *AR* interaction and the associated papers.



Human Factors	Personal Factors	Anatomy + Physiological Factors			Xu et al. 2012, Lewandowski et al. 2011, Beadle et al. 1997,
		Perceptual and Cognitive Factors	Visual Perception		
			Attention		Xu et al. 2012
		Affective State			
		Attitude + Preferences			Hodhod et al. 2014, Doswell 2006, Sinclair and Martinez 2001, Beadle et al. 1997
		Activity	Current Activity		Stricker and Bleser 2012,
			Activity History		Shin et al. 2009
			Concurrent Actions		
	Social Factors	Place	Privacy		
			Crowdedness		
		Social Networks			
Environmental Factors	Physical Factors	Raw Measurements	Human Senses	Vision	Barakonyi et al. 2004, Ghouaiel et al. 2014
				Audition	Barakonyi et al. 2004, Ghouaiel et al. 2014, Xu et al. 2012,
				Temperature	
				Others	
			Non-human senses	Time Points	
				Others	
		Derived Measurements	Space	Spatial Configuration	Henderson and Finer 2008, Uratani et al. 2005, Ghouaiel et al. 2014,
			Time Intervalls		
			Presence/Absence		Grubert et al. 2012,
			Motion		
			Vision	Saliency/Edges/features	Rosten et al. 2005, Bordes et al 2011, Tanaka et al. 2008, Grasset et al 2012,
				Groupings	
				Readability	Gabbard et al. 2005, Kalkofen et al 2009, Kalkofen et al 2013,
	Digital Factors	Type	Abstract data		
			Images		
			3D models		Mendez et al. 2007
			Textual data		
			Audio data		
		Quality			
		Quantity			Julier et al. 2002
	Infrastructure Factors	Network Characteristics			
System Factors	System State	General			Bühling et al. 2007, Verbelen et al. 2011,
		Processing Power			
		Battery Consumption /			
		Sensor Characteristics	Level of Measurement		
			Uncertainty		Hallaway et al 2004, Coelho et al. 2004, MacWilliams et al. 2005, MacIntyre et al. 2002, Pankratz et al. 2013
			Frequency		
	Input				
	System Output	Visual Displays	Spatial Arrangement		
			Display Characteristics		

Figure B.2: Context sources relevant for *AR* interaction and the associated papers.

Bibliography

[1] Achanta, R., Hemami, S., Francisco, E., and Suessstrunk, S. (2009). Frequency-tuned salient region detection. In *IEEE International Conference on Computer Vision and Pattern Recognition*, CVPR '09. (page 36)

[2] Akpan, I., Marshall, P., Bird, J., and Harrison, D. (2013). Exploring the effects of space and place on engagement with an interactive installation. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '13, pages 2213–2222. ACM. (page 10, 28, 30, 89)

[3] Alt, F., Shirazi, A. S., Kubitza, T., and Schmidt, A. (2013). Interaction techniques for creating and exchanging content with public displays. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '13, pages 1709–1718. ACM. (page 108)

[4] Ashbrook, D., Baudisch, P., and White, S. (2011). Nenya: subtle and eyes-free mobile input with a magnetically-tracked finger ring. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '11, pages 2043–2046. ACM. (page 18)

[5] Azuma, R. T. (1997). A survey of augmented reality. *Presence: Teleoperators and Virtual Environments*, 6(4):355–385. (page 15)

[6] Baldauf, M., Dustdar, S., and Rosenberg, F. (2007). A survey on context-aware systems. *International Journal of Ad Hoc and Ubiquitous Computing*, 2(4):263–277. (page 23)

[7] Baldauf, M. and Fröhlich, P. (2013). The augmented video wall: multi-user ar interaction with public displays. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI EA '13, pages 3015–3018. ACM. (page 19)

[8] Baldauf, M., Fröhlich, P., Buchta, J., and Stürmer, T. (2013). From touchpad to smart lens: A comparative study on smartphone interaction with public displays. *International Journal of Mobile Human Computer Interaction*, 5(2):1–20. (page 22)

[9] Baldauf, M., Lasinger, K., and Fröhlich, P. (2012a). Private public screens: detached multi-user interaction with large displays through mobile augmented reality. In *Proceedings of the 11th International Conference on Mobile and Ubiquitous Multimedia*, MUM '12, page 27. ACM. (page 19)

[10] Baldauf, M., Lasinger, K., and Fröhlich, P. (2012b). Real-world drag'n'drop - bidirectional camera-based media transfer between smartphones and large displays. In *Adjunct Proceedings of the 10th International Conference on Pervasive Computing.* (page 19)

[11] Ballagas, R., Borchers, J., Rohs, M., and Sheridan, J. G. (2006). The smart phone: a ubiquitous input device. *Pervasive Computing, IEEE*, 5(1):70–77. (page 2, 18)

[12] Ballagas, R., Rohs, M., and Sheridan, J. G. (2005). Sweep and point and shoot: phonecam-based interactions for large public displays. In *Proceedings of the SIGCHI*

Conference on Human Factors in Computing Systems, CHI EA '05, pages 1200–1203. ACM. (page 2, 19)

[13] Barakonyi, I., Psik, T., and Schmalstieg, D. (2004). Agents that talk and hit back: Animated agents in augmented reality. In *Proceedings of the IEEE International Symposium on Mixed and Augmented Reality*, ISMAR '04, pages 141–150. IEEE. (page 35)

[14] Baricevic, D., Lee, C., Turk, M., Hollerer, T., and Bowman, D. A. (2012). A handheld ar magic lens with user-perspective rendering. In *Proceedings of the IEEE International Symposium on Mixed and Augmented Reality*, ISMAR '12, pages 197–206. IEEE. (page 17)

[15] Batra, R. and Ahtola, O. T. (1991). Measuring the hedonic and utilitarian sources of consumer attitudes. *Marketing letters*, 2(2):159–170. (page 3, 9)

[16] Baum, L. F. (1901). *The Master Key: An Electrical Fairy Tale, Founded Upon the Mysteries of Electricity and the Optimism of Its Devotees. It Was Written For Boys, But Others May Read It.* Bowen-Merrill Company. (page 1, 15)

[17] Baur, D., Boring, S., and Feiner, S. (2012). Virtual projection: exploring optical projection as a metaphor for multi-device interaction. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '12, pages 1693–1702. ACM. (page 17)

[18] Bay, H., Ess, A., Tuytelaars, T., and Gool, L. V. (2008). Speeded-up robust features (surf). *Computer Vision and Image Understanding*, 110(3):346–359. (page 123)

[19] Beaudouin-Lafon, M., Huot, S., Nancel, M., Mackay, W., Pietriga, E., Primet, R., Wagner, J., Chapuis, O., Pillias, C., Eagan, J., Gjerlufsen, T., and Klokmose, C. (2012). Multisurface interaction in the wild room. *Computer*, 45(4):48–56. (page 21)

[20] Bederson, B. B. (1995). Audio augmented reality: a prototype automated tour guide. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '95, pages 210–211. ACM. (page 16)

[21] Benko, H., Ishak, E. W., and Feiner, S. (2005). Cross-dimensional gestural interaction techniques for hybrid immersive environments. In *Proceedings of IEEE Virtual Reality*, VR '05, pages 209–216. IEEE. (page 21)

[22] Berlyne, D. E. (1970). Novelty, complexity, and hedonic value. *Perception & Psychophysics*, 8(5):279–286. (page 9)

[23] Bettini, C., Brdiczka, O., Henricksen, K., Indulska, J., Nicklas, D., Ranganathan, A., and Riboni, D. (2010). A survey of context modelling and reasoning techniques. *Pervasive and Mobile Computing*, 6:161–180. (page 23)

[24] Bier, E. A., Stone, M. C., Pier, K., Buxton, W., and DeRose, T. D. (1993). Toolglass and magic lenses: the see-through interface. In *Proceedings of the ACM SIG Conference on Graphics*, SIGGRAPH '93, pages 73–80. ACM. (page 158)

[25] Billinghurst, M., Bowskill, J., Dyer, N., and Morphett, J. (1998). An evaluation of wearable information spaces. In *Proceedings of the Virtual Reality Annual International Symposium*, VRAIST '98, pages 20–27. IEEE. (page 20)

[26] Billinghurst, M., Kato, H., and Poupyrev, I. (2001). The magicbook-moving seamlessly between reality and virtuality. *Computer Graphics and Applications, IEEE*, 21(3):6–8. (page 20)

[27] Billinghurst, M. and Starner, T. (1999). Wearable devices: new ways to manage information. *Computer*, 32(1):57–64. (page 20)

[28] Böhmer, M., Hecht, B., Schöning, J., Krüger, A., and Bauer, G. (2011). Falling asleep with angry birds, facebook and kindle: a large scale study on mobile application usage. In *Proceedings of the 13th International Conference on Human Computer Interaction with Mobile Devices and Services*, MobileHCI '11, pages 47–56. ACM. (page 59)

[29] Bordes, L., Breckon, T., Katramados, I., and Kheyrollahi, A. (2011). Adaptive object placement for augmented reality use in driver assistance systems. In *Proc. 8th European Conference on Visual Media Production.* (page 36)

[30] Boring, S., Altendorfer, M., Broll, G., Hilliges, O., and Butz, A. (2007). Shoot & copy: phonecam-based information transfer from public displays onto mobile phones. In *Proceedings of the 4th international conference on mobile technology, applications, and systems and the 1st international symposium on Computer human interaction in mobile technology*, MobileHCI '14, pages 24–31. ACM. (page 19)

[31] Boring, S., Baur, D., Butz, A., Gustafson, S., and Baudisch, P. (2010). Touch projector: mobile interaction through video. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '10, pages 2287–2296. ACM. (page 19, 21)

[32] Boring, S., Gehring, S., Wiethoff, A., Blöckner, A. M., Schöning, J., and Butz, A. (2011). Multi-user interaction on media facades through live video on mobile devices. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '11, pages 2721–2724. ACM. (page 19)

[33] Boring, S., Jurmu, M., and Butz, A. (2009). Scroll, tilt or move it: using mobile phones to continuously control pointers on large public displays. In *Proceedings of the 21st Annual Conference of the Australian Computer-Human Interaction Special Interest Group: Design: Open 24/7*, pages 161–168. ACM. (page 19)

[34] Broll, G., Haarländer, M., Paolucci, M., Wagner, M., Rukzio, E., and Schmidt, A. (2008). Collect&drop: A technique for multi-tag interaction with real world objects and information. In *Ambient Intelligence*, pages 175–191. Springer. (page 18)

[35] Broll, G. and Hausen, D. (2010). Mobile and physical user interfaces for nfc-based mobile interaction with multiple tags. In *Proceedings of the 12th international conference on Human computer interaction with mobile devices and services*, MobileHCI '10, pages 133–142. ACM. (page 18)

[36] Broll, G., Siorpaes, S., Rukzio, E., Paolucci, M., Hamard, J., Wagner, M., and Schmidt, A. (2007). Comparing techniques for mobile interaction with objects from the real world. workshop permid 2007 in conjunction with pervasive 2007. (page 19)

[37] Brown, B., Reeves, S., and Sherwood, S. (2011). Into the wild: challenges and opportunities for field trial methods. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '11, pages 1657–1666. ACM. (page 91)

[38] Budhiraja, R., Lee, G. A., and Billinghurst, M. (2013). Using a hhd with a hmd for mobile ar interaction. In *Proceedings of the IEEE International Symposium on Mixed and Augmented Reality*, ISMAR '13, pages 1–6. IEEE. (page 21)

[39] Butler, A., Izadi, S., and Hodges, S. (2008). Sidesight: multi-touch interaction around small devices. In *Proceedings of the 21st annual ACM symposium on User interface software and technology*, UIST '08, pages 201–204. ACM. (page 18)

[40] Cao, X., Li, J. J., and Balakrishnan, R. (2008). Peephole pointing: modeling acquisition of dynamically revealed targets. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '08, pages 1699–1708. ACM. (page 21)

[41] Cappiello, I., Puglia, S., and Vitaletti, A. (2009). Design and initial evaluation of a ubiquitous touch-based remote grocery shopping process. In *First International Workshop on Near Field Communication, 2009*, NFC '09, pages 9–14. IEEE. (page 19)

[42] Carrera, G., Angeli, A., and Davison, A. J. (2011). Slam-based automatic extrinsic calibration of a multi-camera rig. In *Proceedings of the IEEE International Conference on Robotics and Automation*, ICRA '11, pages 2652–2659. IEEE. (page 153)

[43] Carter, S., Liao, C., Denoue, L., Golovchinsky, G., and Liu, Q. (2010). Linking digital media to physical documents: Comparing content-and marker-based tags. *IEEE Pervasive Computing*, 9(2):46–55. (page 19)

[44] Caudell, T. P. and Mizell, D. W. (1992). Augmented reality: An application of heads-up display technology to manual manufacturing processes. In *Proceedings of the Twenty-Fifth Hawaii International Conference on System Sciences*, volume 2 of *ICSS '92*, pages 659–669. IEEE. (page 1, 15)

[45] Chen, X., Grossman, T., Wigdor, D. J., and Fitzmaurice, G. (2014). Duet: exploring joint interactions on a smart phone and a smart watch. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '14, pages 159–168. ACM. (page 21)

[46] Chen, X., Marquardt, N., Tang, A., Boring, S., and Greenberg, S. (2012). Extending a mobile device's interaction space through body-centric interaction. In *Proceedings of the 14th international conference on Human-computer interaction with mobile devices and services*, MobileHCI '12, pages 151–160. ACM. (page 156, 160)

[47] Coelho, E. M., MacIntyre, B., and Julier, S. J. (2004). Osgar: A scene graph with uncertain transformations. In *Proceedings of the IEEE International Symposium on Mixed and Augmented Reality*, ISMAR '04, pages 6–15. (page 38)

[48] Copic Pucihar, K., Coulton, P., and Alexander, J. (2013). Evaluating dual-view perceptual issues in handheld augmented reality: device vs. user perspective rendering. In *Proceedings of the 15th ACM on International conference on multimodal interaction*, ICMI '13, pages 381–388. ACM. (page 17)

[49] Dachselt, R., Häkkilä, J., Jones, M., Löchtefeld, M., Rohs, M., and Rukzio, E. (2012). Pico projectors: firefly or bright future? *interactions*, 19(2):24–29. (page 18)

[50] Darroch, I., Goodman, J., Brewster, S., and Gray, P. (2005). The effect of age and font size on reading text on handheld computers. In *Proceedings of the 2005 IFIP TC13 international conference on Human-Computer Interaction*, INTERACT '05, pages 253–266. Springer. (page 97)

[51] Deci, E. L. and Ryan, R. M. (2000). The" what" and" why" of goal pursuits: Human needs and the self-determination of behavior. *Psychological inquiry*, 11(4):227–268. (page 73)

[52] Delamare, W., Coutrix, C., and Nigay, L. (2013). Designing disambiguation techniques for pointing in the physical world. In *Proceedings of the 5th ACM SIGCHI symposium on engineering interactive computing systems*, EICS '13, pages 197–206. ACM. (page 159)

[53] Dey, A. K. and Abowd, G. D. (1999). Towards a better understanding of context and context-awareness. *Computing Systems*, 40:304–307. (page 23, 28)

[54] DiVerdi, S., Höllerer, T., and Schreyer, R. (2004). Level of detail interfaces. In *Proceedings of the IEEE International Symposium on Mixed and Augmented Reality*, ISMAR '04, pages 300–301. (page 33, 37)

[55] Dix, A., Rodden, T., Davies, N., Trevor, J., Friday, A., and Palfreyman, K. (2000). Exploiting space and location as a design framework for interactive mobile systems. *ACM Transactions on Computer-Human Interaction*, 7(3):285–321. (page 23, 24)

[56] Doswell, J. T. (2006). Augmented learning: context-aware mobile augmented reality architecture for learning. In *Proceedings of the Sixth International Conference on Advanced Learning Technologies*, ICALT '06, pages 1182–1183. IEEE. (page 35)

[57] Dourish, P. (2001). Seeking a foundation for context-aware computing. *Human Computer Interaction*, 16:229–241. (page 23)

[58] Dourish, P. (2004). What we talk about when we talk about context. *Personal and Ubiquitous Computing*, 8:19–30. (page 23)

[59] Dubuisson, M.-P. and Jain, A. (1994). A modified hausdorff distance for object matching. In *Proceedings of the 12th International Conference on Computer Vision, Image Processing and Pattern Recognition*, volume 1 of *IAPR '94*, pages 566–568 vol.1. (page 123)

[60] Dünser, A. and Billinghurst, M. (2011). Evaluating augmented reality systems. In *Handbook of Augmented Reality*, pages 289–307. Springer. (page 2)

[61] DüNser, A., Billinghurst, M., Wen, J., Lehtinen, V., and Nurminen, A. (2012). Exploring the use of handheld ar for outdoor navigation. *Computers & Graphics*, 36(8):1084–1095. (page 22)

[62] Dünser, A., Grasset, R., and Billinghurst, M. (2008). *A survey of evaluation techniques used in augmented reality studies.* Human Interface Technology Laboratory New Zealand. (page 2)

[63] Dünser, A., Grasset, R., and Farrant, H. (2011). Towards immersive and adaptive augmented reality exposure treatment. *Studies in health technology and informatics*, 167:37–41. (page 34, 39, 41)

[64] Ens, B., Finnegan, R., and Irani, P. (2014a). The personal cockpit: a spatial interface for effective task switching on head-worn displays. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '14, pages 3171–3180. ACM. (page 20)

[65] Ens, B., Hincapié-Ramos, J. D., and Irani, P. (2014b). Ethereal planes: a design framework for 2d information space in 3d mixed reality environments. In *Proceedings of the 2nd ACM symposium on spatial user interaction*, SUI '14, pages 2–12. ACM. (page 156, 157)

[66] Feiner, S., MacIntyre, B., Haupt, M., and Solomon, E. (1993a). Windows on the world: 2d windows for 3d augmented reality. In *Proceedings of the 6th annual ACM symposium on User interface and software technology*, UIST '93, pages 145–155. ACM. (page 20)

[67] Feiner, S., MacIntyre, B., Höllerer, T., and Webster, A. (1997). A touring machine: Prototyping 3d mobile augmented reality systems for exploring the urban environment. *Personal Technologies*, 1(4):208–217. (page 1, 16, 42)

[68] Feiner, S., Macintyre, B., and Seligmann, D. (1993b). Knowledge-based augmented reality. *Communications of the ACM*, 36(7):53–62. (page 15)

[69] Feiner, S. and Shamash, A. (1991). Hybrid user interfaces: Breeding virtually bigger interfaces for physically smaller computers. In *Proceedings of the 4th annual ACM symposium on User interface software and technology*, UIST '91, pages 9–17. ACM. (page 20)

[70] Fitzmaurice, G. W. (1993). Situated information spaces and spatially aware palmtop computers. *Communications of the ACM*, 36(7):39–49. (page 1, 15, 20)

[71] Frank, A. U. (1998). Different types of" times" in gls. *Spatial and temporal reasoning in geographic information systems*, page 40. (page 31)

[72] Gabbard, J., Swan, J.E., I., Hix, D., Schulman, R., Lucas, J., and Gupta, D. (2005). An empirical user-based study of text drawing styles and outdoor background textures for augmented reality. *Proceedings of IEEE Virtual Reality.* (page 36)

[73] Garner, P., Rashid, O., Coulton, P., and Edwards, R. (2006). The mobile phone as a digital spraycan. In *Proceedings of the SIGCHI international conference on Advances in computer entertainment technology*, ACE '06, page 12. ACM. (page 18)

[74] Georgel, P. F., Schroeder, P., and Navab, N. (2009). Navigation tools for viewing augmented cad models. *Computer Graphics and Applications, IEEE*, 29(6):65–73. (page 110)

[75] Ghouaiel, N., Cieutat, J.-M., and Jessel, J.-P. (2014). Adaptive augmented reality: plasticity of augmentations. In *Proceedings of the 2014 Virtual Reality International Conference*, VRIC '14, page 10. ACM. (page 37)

[76] Goh, D. H.-L., Lee, C. S., and Razikin, K. (2011). Comparative evaluation of interfaces for presenting location-based information on mobile devices. In *Digital Libraries: For Cultural Heritage, Knowledge Dissemination, and Future Creation*, pages 237–246. Springer. (page 22, 90)

[77] Grasset, R., Langlotz, T., Kalkofen, D., Tatzgern, M., and Schmalstieg, D. (2012). Image-driven view management for augmented reality browsers. In *Proceedings of the IEEE International Symposium on Mixed and Augmented Reality*, ISMAR '12, pages 177–186. IEEE. (page 33, 36, 40)

[78] Grasset, R., Looser, J., and Billinghurst, M. (2006). Transitional interface: concept, issues and framework. In *Proceedings of the IEEE International Symposium on Mixed and Augmented Reality*, ISMAR '06, pages 231–232. IEEE. (page 20, 113)

[79] Grasset, R., Mulloni, A., Billinghurst, M., and Schmalstieg, D. (2011). Navigation techniques in augmented and mixed reality: Crossing the virtuality continuum. In *Handbook of Augmented Reality*, pages 379–407. Springer. (page 20)

[80] Grauman, K. and Leibe, B. (2011). *Visual object recognition*. Number 11. Morgan & Claypool Publishers. (page 18)

[81] Greenberg, S. (2001). Context as a dynamic construct. (page 23)

[82] Greenberg, S., Marquardt, N., Ballendat, T., Diaz-Marino, R., and Wang, M. (2011). Proxemic interactions: the new ubicomp? *Interactions*, 18(1):42–50. (page 155)

[83] Grubert, J., Grasset, R., and Reitmayr, G. (2012a). Exploring the design of hybrid interfaces for augmented posters in public spaces. In *Proceedings of the 7th Nordic Conference on Human-Computer Interaction: Making Sense Through Design*, NordiCHI '12, pages 238–246. ACM. (page 4, 7, 13, 36, 39, 90, 108)

[84] Grubert, J., Gruendler, R., Nixon, L., and Reitmayr, G. (2011a). Annotate that: Preparing event posters for augmentation. In *ISMAR 2011 Workshop on Authoring Solutions for Augmented Reality*, ISMAR '11. (page 114)

[85] Grubert, J., Heinisch, M., Quigley, A., and Schmalstieg, D. (2015a). Multifi: Multi fidelity interaction with displays on and around the body. In *Proceedings of the SIGCHI*

Conference on Human Factors in Computing Systems, CHI '15, pages 3933–3942. ACM. (page 4, 9, 14)

[86] Grubert, J., Langlotz, T., and Grasset, R. (2011b). Augmented reality browser survey. Technical Report 1101, Institute for Computer Graphics and Vision. (page 3, 5, 11, 28, 83, 99, 111)

[87] Grubert, J., Morrison, A., Munz, H., and Reitmayr, G. (2012b). Playing it real: magic lens and static peephole interfaces for games in a public space. In *Proceedings of the 14th international conference on Human-computer interaction with mobile devices and services*, MobileHCI '12, pages 231–240. ACM. (page 3, 6, 12, 83, 87, 92, 93, 98, 99)

[88] Grubert, J., Pahud, M., Grasset, R., Schmalstieg, D., and Seichter, H. (2015b). The utility of magic lens interfaces on handheld devices for touristic map navigation. *Pervasive and Mobile Computing*, 18(0):88 – 103. (page 3, 7, 12, 36, 164)

[89] Grubert, J. and Schmalstieg, D. (2013). Playing it real again: a repeated evaluation of magic lens and static peephole interfaces in public space. In *Proceedings of the 15th international conference on Human-computer interaction with mobile devices and services*, MobileHCI '13, pages 99–102. ACM. (page 3, 6, 12, 30, 36, 92)

[90] Grubert, J., Seichter, H., and Schmalstieg, D. (2014). Towards user perspective augmented reality for public displays. In *Proceedings of the IEEE International Symposium on Mixed and Augmented Reality*, ISMAR '14, pages 339–340. IEEE. (page 4, 8, 14)

[91] Grubert, J., Tuemler, J., Mecke, R., and Schenk, M. (2010). Comparative user study of two see-through calibration methods. In *Proceedings of the IEEE Virtual Reality Conference*, VR '10, pages 269–270. (page 160)

[92] Grubert, J., Zollmann, S., and Langlotz, T. (2015c). Context-aware augmented reality: Trends and opportunities. *Transactions of Visualization and Computer Gaphics*, (submitted). (page 3, 5, 10, 11)

[93] Guven, S., Feiner, S., and Oda, O. (2006). Mobile augmented reality interaction techniques for authoring situated media on-site. In *Proceedings of the IEEE International Symposium on Mixed and Augmented Reality*, ISMAR '06, pages 235–236. IEEE. (page 113)

[94] Haindl, M. and Filip, J. (2013). *Visual Texture*. Advances in Computer Vision and Pattern Recognition. Springer Verlag. (page 121)

[95] Hall, E. T. and Hall, E. T. (1969). *The hidden dimension*, volume 1990. Anchor Books New York. (page 31, 86, 92)

[96] Hallaway, D., Feiner, S., and Höllerer, T. (2004). Bridging the gaps: Hybrid tracking for adaptive mobile augmented reality. *Applied Artificial Intelligence, Special Edition on Artificial Intelligence in Mobile Systems*. (page 31, 38)

[97] Hancock, M., Ten Cate, T., and Carpendale, S. (2009). Sticky tools: full 6dof force-based interaction for multi-touch tables. In *Proceedings of the ACM International Conference on Interactive Tabletops and Surfaces*, ITS '09, pages 133–140. ACM. (page 111)

[98] Hang, A., Rukzio, E., and Greaves, A. (2008). Projector phone: a study of using mobile phones with integrated projector for interaction with maps. In *Proceedings of the 10th international conference on Human computer interaction with mobile devices and services*, MobileHCI '08, pages 207–216. ACM. (page 18)

[99] Harrison, C. and Hudson, S. E. (2009). Abracadabra: wireless, high-precision, and unpowered finger input for very small mobile devices. In *Proceedings of the 22nd annual ACM symposium on User interface software and technology*, UIST '09, pages 121–124. ACM. (page 18)

[100] Hart, S. G. and Staveland, L. E. (1988a). Development of nasa-tlx (task load index): Results of empirical and theoretical research. *Advances in psychology*, 52:139–183. (page 162)

[101] Hart, S. G. and Staveland, L. E. (1988b). *Human Mental Workload*, chapter Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. North Holland Press, Amsterdam. (page 129, 143)

[102] Hartl, A., Grubert, J., Reinbacher, C., Arth, C., and Schmalstieg, D. (2015a). Mobile user interfaces for efficient verification of holograms. In *Proceedings of IEEE Virtual Reality 2015.* (to appear). (page 4, 8, 13)

[103] Hartl, A., Grubert, J., Schmalstieg, D., and Reitmayr, G. (2013). Mobile interactive hologram verification. In *Proceedings of the IEEE International Symposium on Mixed and Augmented Reality*, ISMAR '13, pages 75–82. IEEE. (page 4, 8, 13, 136, 140, 145)

[104] Hartl, A., Grubert, J., Schmalstieg, D., Reitmayr, G., and Dressel, O. (2014). Verfahren zur ausrichung an einer beliebigen pose mit 6 freitheitsgraden fuer ar anwendungen (procedure for view-alignment to an arbitrary six degrees of freedom for augmented reality applications). (page 13)

[105] Hartl, A., Grubert, J., Schmalstieg, D., Reitmayr, G., and Dressel, O. (2015b). Aufnahme der svbrdf von blickwinkelabhaengigen elementen mit mobilen geraeten (svbrdf capture of view-dependent elements with mobile devices). (page 13)

[106] Hassenzahl, M., Burmester, M., and Koller, F. (2003). Attrakdiff: Ein fragebogen zur messung wahrgenommener hedonischer und pragmatischer qualität. In *Mensch und Computer*, M&C '03, pages 187–196. Springer. (page 93, 98, 104, 129, 143, 162)

[107] Heger, S., Portheine, F., Ohnsorge, J. A. K., Schkommodau, E., and Radermacher, K. (2005). User-interactive registration of bone with a-mode ultrasound. *Engineering in Medicine and Biology Magazine, IEEE*, 24(2):85–95. (page 137)

[108] Henderson, J. M. and Hollingworth, A. (1999). High-level scene perception. *Annual review of psychology*, 50:243–271. (page 24)

[109] Henderson, S. J. and Feiner, S. (2008). Opportunistic controls: leveraging natural affordances as tangible user interfaces for augmented reality. In *Proceedings of the 2008 ACM symposium on Virtual reality software and technology*, VRST '08, pages 211–218. ACM. (page 36, 39)

[110] Henricksen, K. and Indulska, J. (2004). A software engineering framework for context-aware pervasive computing. In *Proceedings of the second IEEE Annual Conference on Pervasive Computing and Communications*, PerCom '04, pages 77–86. (page 23)

[111] Henze, N. and Boll, S. (2010). Evaluation of an off-screen visualization for magic lens and dynamic peephole interfaces. In *Proceedings of the 12th international conference on Human computer interaction with mobile devices and services*, MobileHCI '10, pages 191–194. ACM. (page 69, 71, 81)

[112] Hewett, T., Baecker, R., Card, S., and Carey, T. (1992). Acm sigchi curricula for human-computer interaction. (page 29)

[113] Hill, A., Schiefer, J., Wilson, J., Davidson, B., Gandy, M., and MacIntyre, B. (2011). Virtual transparency: Introducing parallax view into video see-through ar. In *Proceedings of the IEEE International Symposium on Mixed and Augmented Reality*, ISMAR '11, pages 239–240. IEEE. (page 17)

[114] Hillier, B. (2007). Space is the machine: a configurational theory of architecture. (page 31)

[115] Hinckley, K., Pierce, J., Sinclair, M., and Horvitz, E. (2000). Sensing techniques for mobile interaction. In *Proceedings of the 13th annual ACM symposium on User interface software and technology*, UIST '00, pages 91–100. ACM. (page 18)

[116] Hinckley, K., Ramos, G., Guimbretiere, F., Baudisch, P., and Smith, M. (2004). Stitching: pen gestures that span multiple displays. In *Proceedings of the working conference on Advanced visual interfaces*, AVI '04, pages 23–31. ACM. (page 21)

[117] Hodhod, R., Fleenor, H., and Nabi, S. (2014). Adaptive augmented reality serious game to foster problem solving skills. In *Proceedings of the 3rd International Workshop on the Reliability of Intelligent Environments*, WoRIE '14, pages 273–284. (page 35, 39)

[118] Hohl, F., Kubach, U., Leonhardi, A., Rothermel, K., and Schwehm, M. (1999). Nexus - an open global infrastructure for spatial-aware applications. Technical report, Universitaetsbibliothek der Universitaet Stuttgart, Holzgartenstr. 16, 70174 Stuttgart. (page 42)

[119] Holbrook, M. B. and Hirschman, E. C. (1982). The experiential aspects of consumption: consumer fantasies, feelings, and fun. *Journal of consumer research*, pages 132–140. (page 9)

[120] Höllerer, T., Feiner, S., Terauchi, T., Rashid, G., and Hallaway, D. (1999). Exploring mars: developing indoor and outdoor user interfaces to a mobile augmented reality system. *Computers & Graphics*, 23(6):779–785. (page 1, 16)

[121] Hong, D., Schmidtke, H., and Woo, W. (2007). Linking context modelling and contextual reasoning. *4th International Workshop on Modeling and Reasoning in Context (MRC)*, pages 37–48. (page 24, 32)

[122] Hürst, W. and Helder, M. (2011). Mobile 3d graphics and virtual reality interaction. In *Proceedings of the 8th International Conference on Advances in Computer Entertainment Technology*, ACE '08, page 28. ACM. (page 111, 117)

[123] Itti, L., Koch, C., and Niebur, E. (1998). A model of saliency-based visual attention for rapid scene analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 20:1254–1259. (page 34)

[124] Jachnik, J., Newcombe, R. A., and Davison, A. J. (2012). Real-time surface lightfield capture for augmentation of planar specular surfaces. In *Proceedings of the IEEE International Symposium on Mixed and Augmented Reality*, ISMAR '12, pages 91–97. IEEE. (page 121)

[125] Julier, S., Lanzagorta, M., Baillot, Y., Rosenblum, L., Feiner, S., Hollerer, T., and Sestito, S. (2000). Information filtering for mobile augmented reality. In *Proceedings of the IEEE and ACM International Symposium on Augmented Reality*, ISAR '00, pages 3–11. IEEE. (page 33, 37, 40)

[126] Kaehaeri, M. and Murphy, D. J. (2006). Mara - sensor based augmented reality system for mobile imaging device. In *Proceedings of theIEEE International Symposium on Mixed and Augmented Reality*, ISMAR '06. (page 17)

[127] Kahneman, D. (1973). *Attention and effort*. Citeseer. (page 2)

[128] Kalkofen, D., Veas, E., Zollmann, S., Steinberger, M., and Schmalstieg, D. (2013). Adaptive ghosted views for augmented reality. In *Proceedings of the IEEE International Symposium on Mixed and Augmented Reality*, ISMAR '13. (page 33, 37, 39)

[129] Kalkofen, D., Zollman, S., Schall, G., Reitmayr, G., and Schmalstieg, D. (2009). Adaptive visualization in outdoor ar displays. In *Proceedings of the IEEE International Symposium on Mixed and Augmented Reality*, ISMAR '09. (page 37)

[130] Kallio, T., Kaikkonen, A., et al. (2005). Usability testing of mobile applications: A comparison between laboratory and field testing. *Journal of Usability studies*, 1(4-16):23–28. (page 115)

[131] Kato, H. and Billinghurst, M. (1999). Marker tracking and hmd calibration for a video-based augmented reality conferencing system. In *Proceedings of the IEEE International Workshop on Augmented Reality*, IWAR '99, pages 85–94. IEEE. (page 16)

[132] Keil, J., Zollner, M., Becker, M., Wientapper, F., Engelke, T., and Wuest, H. (2011). The house of olbrich - an augmented reality tour through architectural history - arts, media, and humanities. In *Proceedings of the IEEE International Symposium on Mixed and Augmented Reality - Arts, Media, and Humanities*, ISMAR AMH '11, pages 15–18. IEEE. (page 114)

[133] Kindberg, T. (2002). Implementing physical hyperlinks using ubiquitous identifier resolution. In *Proceedings of the 11th international conference on World Wide Web*, pages 191–199. ACM. (page 19)

[134] Klein, G. and Murray, D. (2007). Parallel tracking and mapping for small ar workspaces. In *Symposium on Mixed and Augmented Reality*, pages 1–10. (page 16)

[135] Klein, G. and Murray, D. (2009). Parallel tracking and mapping on a camera phone. In *Proceedings of the IEEE International Symposium on Mixed and Augmented Reality*, ISMAR '09, pages 83–86. IEEE. (page 16, 151)

[136] Kooper, R. and MacIntyre, B. (2003). Browsing the real-world wide web: Maintaining awareness of virtual information in an ar information space. *International Journal of Human-Computer Interaction*, 16(3):425–446. (page 16, 42)

[137] Kratz, S. and Rohs, M. (2009). Hoverflow: expanding the design space of around-device interaction. In *Proceedings of the 11th International Conference on Human-Computer Interaction with Mobile Devices and Services*, MobileHCI '09, page 4. ACM. (page 18)

[138] Kruijff, E., Swan II, J. E., and Feiner, S. (2010). Perceptual issues in augmented reality revisited. In *Proceedings of the IEEE International Symposium on Mixed and Augmented Reality*, volume 9 of *ISMAR '10*, pages 3–12. (page 2)

[139] Lacoche, J., Duval, T., Arnaldi, B., Maisel, E., and Royan, J. (2014). A survey of plasticity in 3d user interfaces. In *7th Workshop on Software Engineering and Architectures for Realtime Interactive Systems*. (page 24)

[140] Langlotz, T., Grubert, J., and Grasset, R. (2013a). Augmented reality browsers: essential products or only gadgets? *Communications of the ACM*, 56(11):34–36. (page 2, 3, 5, 11, 90)

[141] Langlotz, T., Grubert, J., and Grasset, R. (2013b). Augmented reality in the real world: Ar browsers - essential products or only gadgets? *Communications of the ACM*. (page 28)

[142] Langlotz, T., Nguyen, T., Schmalstieg, D., and Grasset, R. (2014). Next-generation augmented reality browsers: Rich, seamless, and adaptive. *Proceedings of the IEEE*, 102:155–169. (page 28)

[143] Lee, G. A., Yang, U., Kim, Y., Jo, D., Kim, K.-H., Kim, J. H., and Choi, J. S. (2009). Freeze-set-go interaction method for handheld mobile augmented reality environments. In *Proceedings of the 16th ACM Symposium on Virtual Reality Software and Technology*, pages 143–146. ACM. (page 113)

[144] Lee, S. and Zhai, S. (2009). The performance of touch screen soft buttons. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '09, pages 309–318. ACM. (page 97)

[145] Lehtinen, V., Nurminen, A., and Oulasvirta, A. (2012). Integrating spatial sensing to an interactive mobile 3d map. In *3D User Interfaces (3DUI), 2012 IEEE Symposium on*, pages 11–14. IEEE. (page 112)

[146] Lemmelä, S., Vetek, A., Mäkelä, K., and Trendafilov, D. (2008). Designing and evaluating multimodal interaction for mobile contexts. In *Proceedings of the 10th international conference on Multimodal interfaces*, pages 265–272. ACM. (page 109)

[147] Lepetit, V., Moreno-Noguer, F., and Fua, P. (2009). Epnp: An accurate o (n) solution to the pnp problem. *International journal of computer vision*, 81(2):155–166. (page 152)

[148] Lewandowski, J., Arochena, H. E., Naguib, R. N. G., and Chao, K.-M. (2011). A portable framework design to support user context aware augmented reality applications. In *Third International Conference on Games and Virtual Worlds for Serious Applications*, pages 144–147. IEEE. (page 34, 39)

[149] Lewis, J. R. (1991). Psychometric evaluation of an after-scenario questionnaire for computer usability studies: The asq. *SIGCHI Bull.*, 23(1):78–81. (page 143, 162)

[150] Leykin, A. and Tuceryan, M. (2004). Automatic determination of text readability over textured backgrounds for augmented reality systems. In *Proceedings of the IEEE International Symposium on Mixed and Augmented Reality*, ISMAR '04, pages 224–230. IEEE. (page 59)

[151] Li, F. C. Y., Dearman, D., and Truong, K. N. (2009). Virtual shelves: interactions with orientation aware devices. In *Proceedings of the 22th annual ACM symposium on User interface software and technology*, UIST '09, pages 125–128. ACM. (page 22, 160)

[152] Lindt, I. (2009). *Adaptive 3D-User-Interfaces.* PhD thesis. (page 29)

[153] Ljungstrand, P., Redström, J., and Holmquist, L. E. (2000). Webstickers: using physical tokens to access, manage and share bookmarks to the web. In *Proceedings of DARE 2000 on Designing augmented reality environments*, pages 23–31. ACM. (page 19)

[154] Lowe, D. (1984). Perceptual organization and visual recognition. (page 30)

[155] MacIntyre, B., Coelho, E., and Julier, S. (2002). Estimating and adapting to registration errors in augmented reality systems. In *Proceedings of the IEEE Virtual Reality Conference*, pages 73–80. IEEE. (page 38, 40)

[156] MacIntyre, B., Hill, A., Rouzati, H., Gandy, M., and Davidson, B. (2011). The argon ar web browser and standards-based ar application environment. In *Proceedings of the IEEE International Symposium on Mixed and Augmented Reality*, ISMAR '11, pages 65 –74. (page 17)

[157] Macwilliams, A. (2005). *A Decentralized Adaptive Architecture for Ubiquitous Augmented Reality Systems.* PhD thesis, Technische Universitaet Muenchen. (page 28, 38)

[158] Marr, D. (1982). Vision: A computational investigation into the human representation and processing of visual information. *Phenomenology and the Cognitive Sciences*, 8(4):397. (page 30)

[159] McAuley, E., Duncan, T., and Tammen, V. V. (1989). Psychometric properties of the intrinsic motivation inventory in a competitive sport setting: A confirmatory factor analysis. *Research quarterly for exercise and sport*, 60(1):48–58. (page 93, 98, 129, 132, 143)

[160] Mehra, S., Werkhoven, P., and Worring, M. (2006). Navigating on handheld displays: Dynamic versus static peephole navigation. *ACM Transactions on Computer-Human Interaction (TOCHI)*, 13(4):448–457. (page 21)

[161] Mendez, E. and Schmalstieg, D. (2007). Adaptive augmented reality using context markup and style maps. ISMAR '07, pages 1–2. Ieee. (page 38)

[162] Möller, A., Diewald, S., Roalter, L., and Kranz, M. (2012a). Mobimed: comparing object identification techniques on smartphones. In *Proceedings of the 7th Nordic Conference on Human-Computer Interaction: Making Sense Through Design*, NordiCHI '12, pages 31–40. ACM. (page 18, 19)

[163] Möller, A., Kranz, M., Diewald, S., Roalter, L., Huitl, R., Stockinger, T., Koelle, M., and Lindemann, P. A. (2014). Experimental evaluation of user interfaces for visual indoor navigation. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, CHI '14, pages 3607–3616. ACM. (page 20)

[164] Möller, A., Kranz, M., Huitl, R., Diewald, S., and Roalter, L. (2012b). A mobile indoor navigation system interface adapted to vision-based localization. In *Proceedings of the 11th International Conference on Mobile and Ubiquitous Multimedia*, page 4. ACM. (page 20)

[165] Morrison, A., Mulloni, A., Lemmelä, S., Oulasvirta, A., Jacucci, G., Peltonen, P., Schmalstieg, D., and Regenbrecht, H. (2011). Collaborative use of mobile augmented reality with paper maps. *Computers & Graphics*, 35(4):789–799. (page 23, 82)

[166] Morrison, A., Oulasvirta, A., Peltonen, P., Lemmela, S., Jacucci, G., Reitmayr, G., Näsänen, J., and Juustila, A. (2009). Like bees around the hive: a comparative study of a mobile augmented reality map. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '09, pages 1889–1898. ACM. (page 23, 83, 90, 108)

[167] Mulloni, A., Dünser, A., and Schmalstieg, D. (2010). Zooming interfaces for augmented reality browsers. In *Proceedings of the 12th international conference on Human computer interaction with mobile devices and services*, MobileHCI '10, page 161. ACM. (page 20, 113, 116)

[168] Mulloni, A., Grubert, J., Seichter, H., Langlotz, T., Grasset, R., Reitmayr, G., and Schmalstieg, D. (2012a). Experiences with the impact of tracking technology in mobile augmented reality evaluations. In *MobileHCI 2012 Workshop MobiVis*. (page 7, 14, 22)

[169] Mulloni, A., Seichter, H., and Schmalstieg, D. (2012b). Indoor navigation with mixed reality world-in-miniature views and sparse localization on mobile devices. In *Proceedings of the International Working Conference on Advanced Visual Interfaces*, AVI '12, pages 212–215. (page 33)

[170] Mustonen, T., Olkkonen, M., and Hakkinen, J. (2004). Examining mobile phone text legibility while walking. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI EA '04, pages 1243–1246. ACM. (page 59)

[171] Newman, J., Ingram, D., and Hopper, A. (2001). Augmented reality in a wide area sentient environment. In *Proceedings of the IEEE International Symposium on Augmented Reality*, ISMAR '01, pages 77–86. IEEE. (page 16)

[172] Nister, D. and Stewenius, H. (2006). Scalable recognition with a vocabulary tree. In *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, CVPR '06, pages 2161–2168. (page 123)

[173] Nitschke, C., Nakazawa, A., and Takemura, H. (2013). Corneal imaging revisited: An overview of corneal reflection analysis and applications. *Information and Media Technologies*, 8(2):389–406. (page 173)

[174] Nixon, L. J., Grubert, J., Reitmayr, G., and Scicluna, J. (2012). Smartreality: Integrating the web into augmented reality. In *I-SEMANTICS (Posters & Demos)*, pages 48–54. (page 174)

[175] Oberhofer, C., Grubert, J., and Reitmayr, G. (2012). Natural feature tracking in javascript. In *Proceedings of the IEEE Virtual Reality Conference*, VR '12, pages 113–114. (page 174)

[176] Olsson, T. and Salo, M. (2011). Online user survey on current mobile augmented reality applications. In *Proceedings of the IEEE International Symposium on Mixed and Augmented Reality*, ISMAR '11, pages 75–84. IEEE. (page 83)

[177] Olsson, T. and Salo, M. (2012). Narratives of satisfying and unsatisfying experiences of current mobile augmented reality applications. In *Proceedings of the SIGCHI conference on human factors in computing systems*, CHI '12, pages 2779–2788. ACM. (page 22)

[178] Olwal, A. and Feiner, S. (2009). Spatially aware handhelds for high-precision tangible interaction with large displays. In *Proceedings of the 3rd International Conference on Tangible and Embedded Interaction*, pages 181–188. ACM. (page 17)

[179] Oulasvirta, A., Estlander, S., and Nurminen, A. (2009). Embodied interaction with a 3d versus 2d mobile map. *Personal and Ubiquitous Computing*, 13(4):303–320. (page 111, 112)

[180] Pahud, M., Hinckley, K., Iqbal, S., Sellen, A., and Buxton, B. (2013). Toward compound navigation tasks on mobiles via spatial manipulation. In *Proceedings of the 15th international conference on Human-computer interaction with mobile devices and services*, MobileHCI '13, pages 113–122. ACM. (page 22, 68)

[181] Pankratz, F., Dippon, A., Coskun, T., and Klinker, G. (2013). User awareness of tracking uncertainties in ar navigation scenarios. In *Proceedings of the IEEE International Symposium on Mixed and Augmented Reality*, ISMAR '13, pages 285–286. IEEE. (page 33, 39, 40)

[182] Perlin, K. and Fox, D. (1993). Pad: an alternative approach to the computer interface. In *Proceedings of the ACM SIGGRAPH1993 conference*, SIGGRAPH '93, pages 57–64. ACM. (page 158)

[183] Picard, R. (2000). *Affective computing*. (page 40)

[184] Piekarski, W. P. and Thomas, B. H. (2001). Tinmith-evo5–an architecture for supporting mobile augmented reality environments. In *Proceedings of the IEEE and ACM International Symposium on Augmented Reality*, ISAR '01, pages 177–177. IEEE. (page 1, 16)

[185] Pock, T., Urschler, M., Zach, C., Beichel, R., and Bischof, H. (2007). A duality based algorithm for tv-l1-optical-flow image registration. In *In Proccedings of the Conference on Medical Image Computing and Computer-Assisted Intervention*, MICCAI '07, pages 511–518. (page 130)

[186] Rädle, R., Jetter, H.-C., Marquardt, N., Reiterer, H., and Rogers, Y. (2014). Huddlelamp: Spatially-aware mobile displays for ad-hoc around-the-table collaboration. In *Proceedings of the ACM International Conference on Interactive Tabletops and Surfaces*, ITS '14, pages 45–54. ACM. (page 162)

[187] Rashid, U., Nacenta, M. A., and Quigley, A. (2012). The cost of display switching: a comparison of mobile, large display and hybrid ui configurations. In *Proceedings of the working conference on Advanced visual interfaces*, AVI '12, pages 99–106. ACM. (page 164)

[188] Raskar, R., Van Baar, J., Beardsley, P., Willwacher, T., Rao, S., and Forlines, C. (2006). ilamps: geometrically aware and self-configuring projectors. In *ACM SIGGRAPH 2006 Courses*, SIGGRAPH '06, page 7. ACM. (page 16)

[189] Reetz, A. and Gutwin, C. (2014). Big gestures?: Factors that influence gesture visibility. (page 6)

[190] Reeves, S., Benford, S., O'Malley, C., and Fraser, M. (2005). Designing the spectator experience. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, CHI '05, pages 741–750. ACM. (page 6, 12, 68)

[191] Reichenbacher, T. (2001). Adaptive concepts for a mobile cartography. *Journal of Geographical Sciences*, 11(1):43–53. (page 91, 164)

[192] Reichlen, B. A. (1993). Sparcchair: A one hundred million pixel display. In *Proceedings of the Virtual Reality Annual International Symposium*, VRAIS '93, pages 300–307. IEEE. (page 20)

[193] Reitmayr, G. and Drummond, T. (2006). Going out: robust model-based tracking for outdoor augmented reality. In *Proceedings of the IEEE International Symposium on Mixed and Augmented Reality*, ISMAR '06, pages 109–118. IEEE. (page 16)

[194] Rekimoto, J. (1997). Navicam: A magnifying glass approach augmented reality system. *Teleoperators and Virtual Environment*, 6(4):399–412. (page 2, 15)

[195] Rekimoto, J., Ayatsuka, Y., and Hayashi, K. (1998). Augment-able reality: Situated communication through physical and digital spaces. In *Proceedings of the 2nd IEEE International Symposium on Wearable Computer*, page 68. (page 16)

[196] Rico, J. and Brewster, S. (2009). Gestures all around us: User differences in social acceptability perceptions of gesture based interfaces. In *Proceedings of the 11th International Conference on Human-Computer Interaction with Mobile Devices and Services*, MobileHCI '09, pages 64:1–64:2. ACM. (page 68, 157)

[197] Rico, J. and Brewster, S. (2010). Usable gestures for mobile interfaces: evaluating social acceptability. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '10, pages 887–896. ACM. (page 6, 73, 81, 87)

[198] Rohs, M. (2005). Real-world interaction with camera phones. In *Ubiquitous Computing Systems*, pages 74–89. Springer. (page 2)

[199] Rohs, M. and Oulasvirta, A. (2008). Target acquisition with camera phones when used as magic lenses. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '08, pages 1409–1418. ACM. (page 21)

[200] Rohs, M., Oulasvirta, A., and Suomalainen, T. (2011). Interaction with magic lenses: real-world validation of a fitts' law model. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '11, pages 2725–2728. ACM. (page 21)

[201] Rohs, M., Schleicher, R., Schöning, J., Essl, G., Naumann, A., and Krüger, A. (2009). Impact of item density on the utility of visual context in magic lens interactions. *Personal and Ubiquitous Computing*, 13(8):633–646. (page 21, 91, 92, 93, 99, 100, 106, 107)

[202] Rohs, M., Schöning, J., Raubal, M., Essl, G., and Krüger, A. (2007). Map navigation with mobile devices: virtual versus physical movement with and without visual context. In *Proceedings of the 9th international conference on Multimodal interfaces*, pages 146–153. ACM. (page 2, 7)

[203] Ronkainen, S., Koskinen, E., Liu, Y., and Korhonen, P. (2010). Environment analysis as a basis for designing multimodal and multidevice user interfaces. *Human–Computer Interaction*, 25(2):148–193. (page 109)

[204] Rosten, E., Reitmayr, G., and Drummond, T. (2005). Real-time video annotations for augmented reality. In Bebis, G., Boyle, R., Koracin, D., and Parvin, B., editors, *Advances in Visual Computing*, volume 3804 of *Lecture Notes in Computer Science*, pages 294–302. Springer. (page 36, 39, 40)

[205] Roth, V. and Turner, T. (2009). Bezel swipe: conflict-free scrolling and multiple selection on mobile touch screen devices. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '09, pages 1523–1526. ACM. (page 115)

[206] Roto, V., Väätäjä, H., Jumisko-Pyykkö, S., and Väänänen-Vainio-Mattila, K. (2011). Best practices for capturing context in user experience studies in the wild. In *Proceedings of the 15th International Academic MindTrek Conference: Envisioning Future Media Environments*, pages 91–98. ACM. (page 172)

[207] Rukzio, E. (2006). *Physical mobile interactions: Mobile devices as pervasive mediators for interactions with the real world*. PhD thesis, University of Munich. (page 18)

[208] Rukzio, E., Broll, G., Leichtenstern, K., and Schmidt, A. (2007). Mobile interaction with the real world: An evaluation and comparison of physical mobile interaction techniques. In *Ambient Intelligence*, pages 1–18. Springer. (page 18, 19)

[209] Rukzio, E., Leichtenstern, K., Callaghan, V., Holleis, P., Schmidt, A., and Chin, J. (2006). An experimental comparison of physical mobile interaction techniques: Touching, pointing and scanning. In *In Proceedings of the 8th international conference on Ubiquitous Computing*, UBICOMP '06, pages 87–104. Springer. (page 2, 18, 19)

[210] Rukzio, E., Schmidt, A., and Hussmann, H. (2004). Physical posters as gateways to context-aware services for mobile devices. In *Proceedings of the Sixth IEEE Workshop on Mobile Computing Systems and Applications*, WMCSA '04, pages 10–19. IEEE. (page 65)

[211] Salo, M., Baldauf, M., Fröhlich, P., and Suette, S. (2013). Peak moments of physical mobile interaction techniques. (page 19)

[212] Sanchez, I., Riekki, J., and Pyykkoenen, M. (2008). Touch & control: interacting with services by touching rfid tags. In *Proceedings of the Second International Workshop on RFID Technology -Concepts, Applications, Challenges*, IWRT '08, pages 359–364. (page 18)

[213] Saragih, J. M., Lucey, S., and Cohn, J. F. (2009). Face alignment through subspace constrained mean-shifts. In *Proceedings of the IEEE International Conference on Computer Vision*, ICCV '09, pages 1034–1041. IEEE. (page 152)

[214] Scarr, J., Cockburn, A., and Gutwin, C. (2012). Supporting and exploiting spatial memory in user interfaces. *Interaction*, 6(1):1–84. (page 21)

[215] Schall, G., Schöning, J., Paelke, V., and Gartner, G. (2011). A survey on augmented maps and environments: Approaches, interactions and applications. *Advances in web-based GIS, mapping services and applications. Taylor & Francis Group, UK*. (page 6)

[216] Schildbach, B. and Rukzio, E. (2010). Investigating selection and reading performance on a mobile phone while walking. In *Proceedings of the 12th international conference on Human computer interaction with mobile devices and services*, MobileHCI '10, pages 93–102. ACM. (page 59)

[217] Schmalstieg, D. and Wagner, D. (2008). Mobile phones as a platform for augmented reality. In *Proceedings of the IEEE Virtual Reality Workshop on Software Engineering and Architectures for Realtime Interactive Systems*, pages 43–44. (page 2)

[218] Schmidt, A., Beigl, M., and Gellersen, H.-W. (1999). There is more to context than location. *Computers & Graphics*, 23(6):893–901. (page 23, 28, 29)

[219] Schöning, J., Krüger, A., and Müller, H. J. (2006). *Interaction of mobile camera devices with physical maps.* na. (page 2)

[220] Shin, C., Lee, W., Suh, Y., Yoon, H., Lee, Y., and Woo, W. (2009). Camar 2.0: Future direction of context-aware mobile augmented reality. In *2009 International Symposium on Ubiquitous Virtual Reality*, pages 21–24. IEEE. (page 39)

[221] Short, J., Williams, E., and Christie, B. (1976). The social psychology of telecommunications. (page 73)

[222] Sinclair, P. and Martinez, K. (2001). Adaptive hypermedia in augmented reality. In *Proceedings of the 3rd workshop on adaptive hypertext and hypermedia systems.* (page 34)

[223] Sodhi, R. S., Jones, B. R., Forsyth, D., Bailey, B. P., and Maciocci, G. (2013). Bethere: 3d mobile collaboration with spatial input. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '13, pages 179–188. ACM. (page 18)

[224] Song, J., Sörös, G., Pece, F., Fanello, S. R., Izadi, S., Keskin, C., and Hilliges, O. (2014). In-air gestures around unmodified mobile devices. In *Proceedings of the 27th annual ACM symposium on User interface software and technology*, UIST '14, pages 319–329. ACM. (page 18)

[225] Speiginer, G. and MacIntyre, B. (2014). Ethereal. In *Proceedings of the adjunct publication of the 27th annual ACM symposium on User interface software and technology*, UIST '14 Adjunct, pages 113–114. ACM. (page 33, 37)

[226] Spindler, M. and Dachselt, R. (2009). Paperlens: advanced magic lens interaction above the tabletop. In *Proceedings of the ACM International Conference on Interactive Tabletops and Surfaces*, ITS '09, page 7. ACM. (page 17)

[227] Spindler, M., Schuessler, M., Martsch, M., and Dachselt, R. (2014). Pinch-drag-flick vs. spatial input: rethinking zoom & pan on mobile displays. In *Proceedings of the 32nd annual ACM conference on Human factors in computing systems*, pages 1113–1122. ACM. (page 22, 68)

[228] Spohrer, J. C. (1999). Information in places. *IBM Systems Journal*, 38(4):602–628. (page 16)

[229] Starner, T., Mann, S., Rhodes, B., Levine, J., Healey, J., Kirsch, D., Picard, R. W., and Pentland, A. (1997). Augmented reality through wearable computing. *Presence: Teleoperators and Virtual Environments*, 6(4):386–398. (page 16)

[230] Strang, T. and Linnhoff-Popien, C. (2004). A context modeling survey. *Graphical Models*, Workshop o:1–8. (page 23)

[231] Strauss, A. and Corbin, J. M. (1990). *Basics of qualitative research: Grounded theory procedures and techniques.* Sage Publications, Inc. (page 29, 56)

[232] Stricker, D. and Bleser, G. (2012). From interactive to adaptive augmented reality. In *International Symposium on Ubiquitous Virtual Reality*, ISUVR '12, pages 18–21. IEEE. (page 35, 39)

[233] Stuerzlinger, W. and Wingrave, C. A. (2011). *The value of constraints for 3D user interfaces.* Springer. (page 111, 117)

[234] Sutherland, I. E. (1968). A head-mounted three dimensional display. In *Proceedings of the December 9-11, 1968, fall joint computer conference, part I*, pages 757–764. ACM. (page 1)

[235] Svanaes, D. (2001). Context-aware technology: a phenomenological perspective. *Human–Computer Interaction*, 16(2-4):379–400. (page 23)

[236] Sweetser, P. and Wyeth, P. (2005). Gameflow: a model for evaluating player enjoyment in games. *Computers in Entertainment*, 3(3):3–3. (page 73)

[237] Tacca, M. C. (2011). Commonalities between perception and cognition. *Frontiers in psychology*, 2. (page 30)

[238] Tamminen, S., Oulasvirta, A., Toiskallio, K., and Kankainen, A. (2004). Understanding mobile contexts. *Personal and ubiquitous computing*, 8(2):135–143. (page 31, 65, 109)

[239] Tanaka, K., Kishino, F., and Miyamae, M. (2008). An information layout method for an optical see-through head mounted display focusing on the viewability. In *Proceedings of the IEEE International Symposium on Mixed and Augmented Reality*, ISMAR '08, pages 139–142. Ieee. (page 36)

[240] Thevenin, D. and Coutaz, J. (1999). Plasticity of user interfaces: Framework and research agenda. In *Proceedings of the Proceedings of the IFIP TC13 International Conference on Human-Computer Interaction*, volume 99 of *INTERACT '99*, pages 110–117. (page 24, 29, 33)

[241] Thomas, B. H. (2012). A survey of visual, mixed, and augmented reality gaming. *Computers in Entertainment (CIE)*, 10(3):3. (page 90)

[242] Tomasello, M. (2008). *Origins of human communication.* MIT press Cambridge. (page 6, 68)

[243] Tomioka, M., Ikeda, S., and Sato, K. (2013). Approximated user-perspective rendering in tablet-based augmented reality. In *Proceedings of the IEEE International Symposium on Mixed and Augmented Reality*, ISMAR '13, pages 21–28. IEEE. (page 17)

[244] Tresadern, P. A., Ionita, M. C., and Cootes, T. F. (2012). Real-time facial feature tracking on a mobile device. *International Journal of Computer Vision*, 96(3):280–289. (page 151)

[245] Ullmer, B. and Ishii, H. (1997). The metadesk: models and prototypes for tangible user interfaces. In *Proceedings of the 10th annual ACM symposium on User interface software and technology*, UIST '97, pages 223–232. ACM. (page 17, 150)

[246] Uratani, K., Machida, T., Kiyokawa, K., and Takemura, H. (2005). A study of depth visualization techniques for virtual annotations in augmented reality. In *Proceedings of IEEE Virtual Reality*, VR '05, pages 295–296. IEEE. (page 37)

[247] van Renesse, R. L. (2005). *Optical Document Security.* Artech House, third edition. (page 120)

[248] Vandenberg, S. G. and Kuse, A. R. (1978). Mental rotations, a group test of three-dimensional spatial visualization. *Perceptual and Motor Skills*, 47(2):599–604. (page 142)

[249] Veas, E., Mulloni, A., Kruijff, E., Regenbrecht, H., and Schmalstieg, D. (2010). Techniques for view transition in multi-camera outdoor environments. In *Proceedings of Graphics Interface*, GI '10, pages 193–200. Canadian Information Processing Society. (page 20, 113)

[250] Verbelen, T., Stevens, T., Simoens, P., De Turck, F., and Dhoedt, B. (2011). Dynamic deployment and quality adaptation for mobile augmented reality applications. *Journal of Systems and Software*, 84(11):1871–1882. (page 39)

[251] Vlahakis, V., Karigiannis, J., Tsotros, M., Gounaris, M., Almeida, L., Stricker, D., Gleue, T., Christou, I. T., Carlucci, R., and Ioannidis, N. (2001). Archeoguide: first results of an augmented reality, mobile computing system in cultural heritage sites. In *Proceedings of the 2001 Conference on Virtual Reality, Archeology, and Cultural Heritage, Glyfada, Greece, November 28-30, 2001*, pages 131–140. (page 16)

[252] Wagner, D., Reitmayr, G., Mulloni, A., Drummond, T., and Schmalstieg, D. (2008). Pose tracking from natural features on mobile phones. In *Proceedings of the IEEE International Symposium on Mixed and Augmented Reality*, ISMAR '08, pages 125–134. Ieee. (page 16, 151, 154)

[253] Wagner, D. and Schmalstieg, D. (2003). First steps towards handheld augmented reality. In *Proceedings of the 7th IEEE International Symposium on Wearable Computers*, ISWC '03, page 127. (page 16)

[254] Wang, J., Zhai, S., and Canny, J. (2006). Camera phone based motion sensing: interaction techniques, applications and performance study. In *Proceedings of the 19th annual ACM symposium on User interface software and technology*, UIST '06, pages 101–110. ACM. (page 18)

[255] Want, R., Fishkin, K. P., Gujar, A., and Harrison, B. L. (1999). Bridging physical and virtual worlds with electronic tags. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '99, pages 370–377. (page 18)

[256] Wasserman, S. and Faust, K. (1994). *Social Network Analysis: Methods and Applications*, volume 8. (page 30)

[257] Weiser, M. (1991). The computer for the 21st century. *Scientific american*, 265(3):94–104. (page 18)

[258] Weiser, M. (1993). Hot topics-ubiquitous computing. *Computer*, 26(10):71–72. (page 18)

[259] Woo, G., Lippman, A., and Raskar, R. (2012). Vrcodes: Unobtrusive and active visual codes for interaction by exploiting rolling shutter. In *Proceedings of the IEEE International Symposium on Mixed and Augmented Reality*, ISMAR '12, pages 59–64. IEEE. (page 18, 154, 174)

[260] Xiao, R., Lew, G., Marsanico, J., Hariharan, D., Hudson, S., and Harrison, C. (2014). Toffee: enabling ad hoc, around-device interaction with acoustic time-of-arrival correlation. In *Proceedings of the 16th international conference on Human-computer interaction with mobile devices and services*, MobileHCI '14, pages 67–76. ACM. (page 18)

[261] Xu, Y., Barba, E., Radu, I., Gandy, M., Shemaka, R., Schrank, B., MacIntyre, B., and Tseng, T. (2011). Pre-patterns for designing embodied interactions in handheld augmented reality games. In *Proceedings of the IEEE International Symposium on Mixed and Augmented Reality - Arts, Media, and Humanities*, ISMAR AMH '11, pages 19–28. IEEE. (page 113)

[262] Xu, Y., Stojanovic, N., Stojanovic, L., Cabrera, A., and Schuchert, T. (2012). An approach for using complex event processing for adaptive augmented reality in cultural heritage domain. In *Proceedings of the 6th ACM International Conference on Distributed Event-Based Systems, DEBS'12*, pages 139–148. (page 34, 36, 39)

[263] Yang, J. and Wigdor, D. (2014). Panelrama: enabling easy specification of cross-device web applications. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '14, pages 2783–2792. ACM. (page 21)

[264] Yee, K.-P. (2003). Peephole displays: pen interaction on spatially aware handheld computers. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, CHI '03, pages 1–8. ACM. (page 19, 22, 113)

[265] Yu, L., Svetachov, P., Isenberg, P., Everts, M. H., and Isenberg, T. (2010). Fi3d: Direct-touch interaction for the exploration of 3d scientific visualization spaces. *Visualization and Computer Graphics, IEEE Transactions on*, 16(6):1613–1622. (page 111)

[266] Zhao, J., Soukoreff, R. W., Ren, X., and Balakrishnan, R. (2014). A model of scrolling on touch-sensitive displays. *International Journal of Human-Computer Studies*, 72(12):805–821. (page 165)

[267] Zimmermann, A., Lorenz, A., Oppermann, R., and Augustin, S. (2007). An operational definition of context. In *Proceedings of the 6th international and interdisciplinary conference on Modeling and using context*, CONTEXT '07, pages 558–571. (page 23)

[268] Zollmann, S., Kalkofen, D., Mendez, E., and Reitmayr, G. (2010). Image-based ghostings for single layer occlusions in augmented reality. In *Proceedings of the IEEE International Symposium on Mixed and Augmented Reality*, ISMAR '10, pages 19–26. IEEE. (page 37)